Foundations of Rule Learning

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Johannes Fürnkranz
Dragan Gamberger
Nada Lavrač

Foundations of Rule Learning

Authors

Johannes Fürnkranz
TU Darmstadt
Hochschulstraße 10
D-64289 Darmstadt
Germany
Email: `juffi@ke.tu-darmstadt.de`

Dragan Gamberger
Rudjer Bošković Institute
Bijenićka 54
CRO-10000 Zagreb
Croatia
Email: `dragan.gamberger@irb.hr`

Nada Lavrač
Jožef Stefan Institute
Jamova 39
SI-1000 Ljubljana
Slovenia
Email: `Nada.Lavrac@ijs.si`

Vienna	*Johannes Fürnkranz*
Zagreb	*Dragan Gamberger*
Ljubljana	*Nada Lavrač*

September 2011

Table of Contents

1. Machine Learning and Data Mining

Machine learning and data mining are research areas of computer science whose quick development is due to the advances in data analysis research, growth in the database industry and the resulting market needs for methods that are capable of extracting valuable knowledge from large data stores. This chapter gives an informal introduction to machine learning and data mining, and describes selected machine learning and data mining methods illustrated by examples. After a brief general introduction, Section 1.2 briefly sketches the historical background of the research area, followed by an outline of the knowledge discovery process and the emerging standards in Section 1.3. Section 1.4 establishes the basic terminology and provides a categorization of different learning tasks. Predictive and descriptive data mining techniques are illustrated by means of simplified examples of data mining tasks in Sections 1.5 and 1.6, respectively. In Section 1.7, we highlight the importance of relational data mining techniques. The chapter concludes with some speculations about future developments in data mining.

1.1 Introduction

Machine learning (Mitchell, 1997) is a mature and well-recognized research area of computer science, mainly concerned with the discovery of models, patterns, and other regularities in data. Machine learning approaches can be roughly categorized into two different groups:

Symbolic approaches. Inductive learning of symbolic descriptions, such as rules (Michalski et al., 1986b; Clark & Niblett, 1989), decision trees (Quinlan, 1986) or logical representations (Muggleton, 1992; Lavrač & Džeroski, 1994a; De Raedt, 2008). Textbooks that focus on this line of research include (Mitchell, 1997; Langley, 1996; Witten & Frank, 2005).

Statistical approaches. Statistical or pattern-recognition methods, including k-nearest neighbor or instance-based learning (Dasarathy, 1991; Aha et al., 1991), Bayesian classifiers (Pearl, 1988), neural network learning (Rumelhart & McClelland, 1986), and support vector machines (Vapnik, 1995; Schölkopf & Smola, 2001). Textbooks in this area include (Bishop, 1995; Ripley, 1996; Duda et al., 2000; Hastie et al., 2001).

Although the approaches taken in these fields are often quite different, their effectiveness in learning is often comparable (Michie et al., 1994b). Also, there are many approaches that cross the boundaries between the two approaches. For example, there are decision tree (Breiman et al., 1984) and rule

† This chapter is partly based on (Lavrač & Grobelnik, 2003).

learning (Friedman & Fisher, 1999) algorithms that are firmly based in statistics. Similarly, ensemble techniques such as boosting (Freund & Schapire, 1997), bagging (Breiman, 1996) or random forests (Breiman, 2001a) may combine the predictions of multiple logical models on a sound statistical basis (Schapire et al., 1998; Mease & Wyner, 2008; Bennett et al., 2008). This book is concerned only with the first group of methods, which result in symbolic, human-understandable patterns and models.

Due to the growth in the database industry and the resulting market needs for methods that are capable of extracting valuable knowledge from large data stores, *data mining* (DM) and *knowledge discovery in databases* (KDD) (Piatetsky-Shapiro & Frawley, 1991; Fayyad et al., 1995; Han & Kamber, 2001) have recently emerged as a new scientific and engineering discipline, with separate workshops, conferences and journals. According to Witten & Frank (2005), data mining means "solving problems by analyzing data that already exists in databases". In addition to the mining of structured data stored in data warehouses—e.g., in the form of relational data tables—there has recently also been increased interest in the mining of unstructured data such as text and web.

Research areas related to machine learning and data mining include database technology and data warehouses, pattern recognition and soft computing, text and web mining, visualization, and statistics.

- *Database technology and data warehouses* are concerned with the efficient storage, access and manipulation of data.
- *Pattern recognition and soft computing* typically provide techniques for classifying data items.
- *Text and web mining* are used for web page analysis, text categorization, as well as filtering and structuring of text documents; natural language processing can provide useful tools for improving the quality of text mining results.
- *Visualization* concerns the visualization of data as well as the visualization of data mining results.
- *Statistics* is a classical data analysis discipline, mainly concerned with the analysis of large collections of numerical data.

As statistics already provides numerous data analysis tools (Friedman, 1998; Breiman, 2001b), a relevant question is whether machine learning and data mining are needed at all. There are several possible answers. First, as industry needs solutions for real-life problems, one of the most important issues is the problem solving speed: many data mining methods are able to deal with very large datasets in a very efficient way, while the algorithmic complexity of statistical methods may turn out to be prohibitive for their use on very large databases. Next, the results of the analysis need to be represented in an appropriate, usually human understandable way; apart from the analytical languages used in statistics, data mining methods also use other forms of formalisms, the most popular being decision trees and rule sets. The next important issue in a real-life setting concerns the assumptions about the data. In general one may claim that data mining deals with all sorts of structured tabular data (e.g., non-numeric, highly unbalanced, unclean data) as well as with non-structured data (e.g., text documents, images, multimedia), and does not make assumptions about the distribution of the data. Finally, while one of the main goals of statistics is hypothesis testing, one of the main goals of data mining is the construction of hypotheses.

1.2 Historical background

Machine learning is a well-established research area of computer science. Early machine learning algorithms were perceptrons (later called neural networks, (Rumelhart & McClelland, 1986)), decision tree learners like ID3 (Quinlan, 1979, 1986) and CART (Breiman et al., 1984), and rule learners like AQ (Michalski, 1969; Michalski et al., 1986b) and INDUCE (Michalski, 1980). These early algorithms were typically used to induce classifiers from a relatively small set of training examples (up to a thousand) described by a small set of attributes (up to a hundred). An overview of early work in machine learning can be found in (Michalski et al., 1983, 1986a).

Data mining and knowledge discovery in databases appeared as a recognizable research discipline in the early 1990s (Piatetsky-Shapiro & Frawley, 1991), with the advent of a series of data mining workshops. The birth of this area was triggered by a need in the database industry to deliver solutions enhancing the traditional database management systems and technologies. At that time, these systems were able to solve the basic data management issues like how to deal with the data in transactional processing systems. In online transactional processing (OLTP) most of the processing scenarios were predefined. The main emphasis was on the stability and safety of solutions.

As the business emphasis changed from automation to decision support, limitations of OLTP systems in business support led to the development of the next-generation data management technology known as data warehousing. The motivation for data warehousing was to provide tools for supporting analytical operations for decision support that were not easily provided by the existing database query languages. Online analytical processing (OLAP) was introduced to enable inexpensive data access and insights which did not need to be defined in advance. However, the typical operations on data warehouses were similar to the ones from the traditional OLTP databases in that the user issued a query and received a data table as a result. The major difference between OLTP and OLAP is the average number of records accessed per typical operation. While a typical operation in OLTP affects only up to tens or hundreds of records in predefined scenarios, a typical operation in OLAP affects up to millions of records (sometimes all records) in the database in a non-predefined way.

The role of data mining in the above framework can be explained as follows. While typical questions in OLTP and OLAP are of the form: *'What is the answer to the given query?'*, data mining—in a somewhat simplified and provocative formulation—addresses the question *'What is the right question to ask about this data?'*. The following explanation can be given. Data warehousing/OLAP provides analytical tools enabling only user-guided analysis of the data, where the user needs to have enough advance knowledge about the data to be able to raise the right questions in order to get the appropriate answers. The problem arises in situations when the data is too complex to be appropriately understood and analyzed by a human. In such cases data mining can be used, providing completely different types of operations for handling the data, aimed at hypothesis construction, and providing answers to questions which—in most cases—cannot be formulated precisely.

1.3 Knowledge discovery process and standardization

Data mining is the core stage of the knowledge discovery process that is aimed at the extraction of interesting—nontrivial, implicit, previously unknown and potentially useful—information from data in large databases (Fayyad et al., 1996). Data mining projects were initially carried out in many different ways with each data analyst finding their own way of approaching the problem, often through trial-and-error. As the data mining techniques and businesses evolved, there was a need for data analysts to better understand and standardize the knowledge discovery process, which would—as a side effect—demonstrate to prospective customers that data mining was sufficiently mature to be adopted as a key element of their business. This led to the development of the *cross-industry standard process for data mining* (CRISP-DM; Chapman et al., 2000), which is intended to be independent of the choice of data mining tools, industry segment, and the application/problem to be solved.

The CRISP-DM methodology defines the crucial steps of the knowledge discovery process. Although in most data mining projects, several iterations of individual steps or step sequences need to be performed, these basic guidelines are very useful both for the data analyst and the client. Below is a list of CRISP-DM steps.

1. *Business understanding:* understanding and defining of business goals and the actual goals of data mining.
2. *Data understanding:* familiarization with the data and the application domain, by exploring and defining the relevant prior knowledge.
3. *Data preparation through data cleaning and preprocessing*: creating the relevant data subset through data selection, as well as finding of useful properties/attributes, generating new attributes, defining appropriate attribute values and/or value discretization.

4. *Data mining*: the most important step of this process, which is concerned with choosing the most appropriate data mining tools—from the available tools for summarization, classification, regression, association, clustering—and searching for patterns or models of interest.
5. *Evaluation and interpretation of results*: aided by pattern/model visualization, transformation, and removal of redundant patterns.
6. *Deployment:* the use of the discovered knowledge.

A terminological note needs to be made at this point. While data mining is considered to be the core step of the knowledge discovery process, in this book—as with most industrial applications—we use the term data mining interchangeably with knowledge discovery.

In addition to the CRISP-DM standardized methodology for building data mining applications, standards covering specific phases of the process are also emerging. These standards include:

– the XML-based Predictive Modeling Markup Language (PMML) (Pechter, 2009) standard for storing and sharing data mining results,
– a standard extending the Microsoft analysis server with new data mining functionality (OLE DB for data mining, using a customized SQL language),
– part of the ISO effort to define multimedia and application-specific SQL types and their methods, including support for data mining functionality (SQL/MM), and
– the emerging Java API for data mining (JDM).

The standardization efforts and numerous tools available (IBM Intelligent Miner, SAS Enterprise Miner, SPSS Clementine, and many others), including the publicly available academic data mining platforms Weka (Witten & Frank, 2005; Hall et al., 2009), Rapid-I (formerly YALE; Mierswa et al. 2006), the Konstanz Information Miner KNIME (Berthold et al., 2009), Orange (Demšar et al., 2004), and the statistical data analysis package R (Everitt & Hothorn, 2006; Torgo, 2010) demonstrate that data mining has made progress towards becoming a mature and widely used technology for analytical practices.

Most of the available tools are capable of mining data in tabular format, describing a dataset in terms of a fixed collection of attributes (properties), as is the case with transactional databases. More sophisticated tools are available for data mining from relational databases, data warehouses and stores of text documents. Methods and tools for the mining of advanced database systems and information repositories (object-oriented and object-relational databases, spatial databases, time-series data and temporal data, multimedia data, heterogeneous and legacy data, World-Wide Web) still lack commercial deployment.

No.	Education	Marital Status	Sex	Has Children	Approved
1	primary	single	male	no	no
2	primary	single	male	yes	no
3	primary	married	male	no	yes
4	university	divorced	female	no	yes
5	university	married	female	yes	yes
6	secondary	single	male	no	no
7	university	single	female	no	yes
8	secondary	divorced	female	no	yes
9	secondary	single	female	yes	yes
10	secondary	married	male	yes	yes
11	primary	married	female	no	yes
12	secondary	divorced	male	yes	no
13	university	divorced	female	yes	no
14	secondary	divorced	male	no	yes

Table 1.1: A sample database.

1.4 Terminology and categorization of learning tasks

In the simplest case, data mining techniques operate on a single data table. Rows in the data table correspond to objects (*training examples*) to be analyzed in terms of their properties (*attributes*) and the concept (*class*) to which they belong. There are two main approaches:

Supervised learning. Supervised learning assumes that training examples are classified (labeled by class labels)

Unsupervised learning. Unsupervised learning concerns the analysis of unclassified examples.

In both cases, the goal is to induce a *model* for the entire dataset, or to discover one or more *patterns* that hold for some part of the dataset.

In supervised learning, data is usually formed from examples (records of given attribute values) which are labeled by the class to which they belong (Kotsiantis et al., 2006). The task is to find a model (a classifier) that will enable a newly encountered instance to be classified. Examples of discrete classification tasks are classification of countries based on climate, classification of cars based on gas consumption, or prediction of a diagnosis based on patient's medical condition.

Let us formulate a classification/prediction task, and illustrate it by a simplified example. As described above, we are given a database of observations described with a fixed number of attributes $\mathbf{A}_i$, and a designated class attribute $\mathbf{C}$. The learning task is to find a mapping f that is able to compute the class value $\mathbf{C} = f(\mathbf{A}_1, \ldots, \mathbf{A}_n)$ from the attribute values of new, previously unseen observations.

Table 1.1 shows a very small, artificial sample database.[1] The database contains the results of a survey on 14 individuals, concerning the approval or disapproval of a certain issue. Each individual is characterized by four attributes—Education (with possible values primary school, secondary school, or university), MaritalStatus (with possible values single, married, or divorced), Sex (male or female), and HasChildren (yes or no)—that encode rudimentary information about their sociodemographic background. The last column, Approved, is the class attribute, encoding whether the individual approved or disapproved of the issue.

The task is to use the information in this *training set* to derive a model that is able to predict whether a person is likely to approve or disapprove the issue, based on the four demographic characteristics. While there are statistical techniques that are able to solve particular instances of this problem, mainly focusing on the analysis of numeric data, machine learning and data mining techniques focus on the analysis of categorical, non-numeric data, and on the interpretability of the result.

Typical data mining approaches find patterns or models in a single data table, while some, like most of the relational data mining approaches, (Lavrač & Džeroski, 1994a; Džeroski & Lavrač, 2001) find patterns/models from data stored in multiple tables, e.g., in a given relational database.

Propositional learning. Data mining approaches that find patterns/models in a given single table are referred to as *attribute-value* or *propositional* learning approaches, as the patterns/models they find can be expressed in propositional logic.

Relational learning. *First-order learning* approaches are also referred to as *relational data mining* (RDM) (Džeroski & Lavrač, 2001), *relational learning* (RL) (Quinlan, 1990) or *inductive logic programming* (ILP) (Muggleton, 1992; Lavrač & Džeroski, 1994a), as the patterns/models they find are expressed in relational formalisms of first-order logic.

We further distinguish between predictive and descriptive data mining. In the example above, a predictive data mining approach will aim at building a predictive classification model for classifying new instances into one of the two class values (yes or no). On the other hand, in descriptive data mining the input data table will typically not contain a designated class attribute and will aim at finding patterns describing the relationships between other attribute values.

Predictive data mining. Predictive data mining methods are supervised. They are used to induce models or theories (such as decision trees or rule sets) from class-labeled data. The induced models can be used for prediction and classification.

[1] The dataset is adapted from the well-known dataset Quinlan (1986).

Fig. 1.1: A decision tree describing the dataset shown in Table 1.1.

Descriptive data mining. Descriptive data mining methods are typically unsupervised. They are used to induce interesting patterns (such as association rules) from unlabeled data. The induced patterns are useful in exploratory data analysis.

While there is no clear distinction in the literature, we will generally use the term *pattern* for results of a descriptive data mining process, whereas we will use the terms *model*, *theory*, or *hypothesis* for results of a predictive data mining task.

The next two sections briefly introduce the two main learning approaches, predictive and descriptive induction.

1.5 Predictive data mining: Induction of models

This data analysis task is concerned with the induction of models for classification and prediction purposes, and is referred to as *predictive induction*. Two symbolic data mining methods that result in classification/prediction models are outlined in this section: decision tree induction and rule set induction.

1.5.1 Decision tree induction

A *decision tree* is a classification model whose structure consists of a number of *nodes* and *arcs*. In general, a node is labeled by an attribute name, and an arc by a valid value of the attribute associated with the node from which the arc originates. The top-most node is called the *root* of the tree, and the bottom nodes are called the *leaves*. Each leaf is labeled by a class (value of the class attribute). When used for classification, a decision tree is traversed in a top-down manner, following the arcs with attribute values satisfying the instance that is to be classified. The traversal of the tree leads to a leaf node and the instance is assigned the class label of the leaf. Figure 1.1 shows a decision tree induced from the training set shown in Table 1.1.

A decision tree is constructed in a top-down manner, starting with the most general tree consisting of only the root node, and then refining it to a more specific tree structure. A small tree consisting only of the root node is *too general*, while the most specific tree which would construct a leaf node for every single data instance would be *too specific*, as it would *overfit* the data. The art of decision tree construction is to construct a tree at the right 'generality level' which will adequately generalize the training data to enable high predictive accuracy on new instances.

The crucial step in decision tree induction is the choice of an attribute to be selected as a node in a decision tree. Typical attribute selection criteria

Fig. 1.2: A bad decision tree describing the dataset shown in Table 1.1.

use a function that measures the *purity* of a node, i.e., the degree to which the node contains only examples of a single class. This purity measure is computed for a node and all successor nodes that result from using an attribute for splitting the data. In the well-known C4.5 decision tree algorithm, which uses information-theoretic entropy as a purity measure (Quinlan, 1986), the difference between the original purity value and the sum of the purity values of the successor nodes weighted by the relative sizes of these nodes, is used to estimate the utility of this attribute, and the attribute with the largest utility is selected for expanding the tree.

To see the importance of this choice, consider a procedure that constructs decision trees simply by picking the next available attribute instead of the most informative attribute. The result is a much more complex and less comprehensible tree (Figure 1.2). Most leaves cover only a single training example, which means that this tree is overfitting the data. Consequently, the labels that are attached to the leaves are not very reliable. Although the trees in Figures 1.1 and 1.2 both classify the training data in Table 1.1 correctly, the former appears to be more trustworthy, and it has a higher chance of correctly predicting the class values of new data.[2]

Note that some of the attributes may not occur at all in the tree; for example, the tree in Figure 1.1 does not contain a test on Education. Apparently, the data can be classified without making a reference to this variable. In addition, the attributes in the upper parts of the tree (near the root) have a stronger influence on the value of the target variable than the nodes in the lower parts of the tree, in the sense that they participate in the classification of a larger number of instances.

As a result of the recursive partitioning of the data at each step of the top-down tree construction process, the number of examples that end up in each node decreases steadily. Consequently, the reliability of the chosen attributes decreases with increasing depths of the tree. As a result, overly complex models are generated, which explain the training data but do not generalize well to unseen data. This is known as *overfitting*. This is the main reason why the state-of-the-art decision tree learners employ a post-processing phase in which the generated tree is simplified by *pruning* branches and nodes near the leaves, which results in replacing some of the interior nodes of the tree with a new leaf, thereby removing the subtree that was rooted at this node. It is important to note that the leaf nodes of the new tree are no longer pure

[2] The preference for simpler models is a heuristic criterion known as *Occam's razor*, which appears to work well in practice. It is often addressed in the literature on model selection, but its utility has been the subject of discussion (Domingos, 1999; Webb, 1996).

nodes, containing only training examples of the same class labeling the leaf; instead the leaf will bear the label of the most frequent class at the leaf.

Many decision tree induction algorithms exist, the most popular being C4.5 and its variants: a commercial product SEE5, and J48, which is available in the WEKA workbench (Witten & Frank, 2005), as open source.

1.5.2 Rule set induction

Another important machine learning technique is the induction of rule sets. The learning of rule-based models has been a main research goal in the field of machine learning since its beginning in the early 1960s. Recently, rule-based techniques have also received increased attention in the statistical community (Friedman & Fisher, 1999).

A rule-based classification model consists of a set of if–then rules. Each *rule* has a conjunction of attribute values (which will in the following be called *features*) in the conditional part of the rule, and a class label in the rule consequent. As an alternative to such logical rules, *probabilistic rules* can be induced; in addition to the predicted class label, the consequent of these rules consists also of a list of probabilities or numbers of covered training instances for each possible class label (Clark & Boswell, 1991).

Rule sets are typically simpler and more comprehensible than decision trees. To see why, note that a decision tree can also be interpreted as a set of if–then rules. Each leaf in the tree corresponds to one rule, where the conditions encode the path that is taken from the root to this particular leaf, and the conclusion of the rule is the label of that leaf. Figure 1.3 shows the set of rules that corresponds to the tree in Figure 1.1. Note the rigid structure of these rules. For example, the first condition always uses the same attribute, namely, the one used at the root of the tree. Next to each rule, we show the proportion of covered examples for each class value.

The main difference between the rules generated by a decision tree and the rules generated by a rule learning algorithm is that the former rule set consists of nonoverlapping rules that span the entire instance space (i.e., each possible combination of attribute values will be covered by exactly one rule). Relaxing this constraint—by allowing for potentially overlapping rules that need not span the entire instance space—may often result in smaller rule sets. However, in this case, we need mechanisms for tie breaking (which rule to choose when more than one covers the example to be classified) and default classifications (what classification to choose when no rule covers the given example). Typically, one prefers rules with a higher ratio of correctly classified examples from the training set.

Figure 1.4 shows a particularly simple rule set which uses two different attributes in its first two rules. Note that these two rules are overlapping: several examples will be covered by more than one rule. For instance, examples 3 and 10 are covered by both the first and the third rule. These conflicts

```
IF   MaritalStatus = single
 AND Sex = female
THEN Approved = yes

IF   MaritalStatus = single
 AND Sex = male
THEN Approved = no

IF   MaritalStatus = married
THEN Approved = yes

IF   MaritalStatus = divorced
 AND HasChildren = yes
THEN Approved = no

IF   MaritalStatus = divorced
 AND HasChildren = no
THEN Approved = yes
```

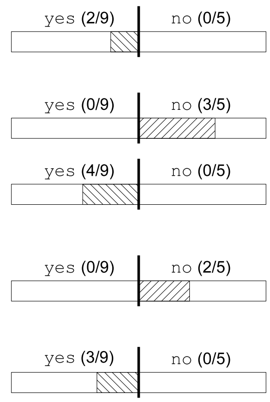

Fig. 1.3: A rule set describing the dataset shown in Table 1.1.

are typically resolved by using the more accurate rule, i.e., the rule that covers a higher proportion of examples that support its prediction (the first one in our case). Also note that this rule set makes two mistakes (the last two examples). These might be resolved by resorting to a more complex rule set (like the one in Figure 1.3) but, as stated above, it is often more advisable to sacrifice accuracy on the training set for model simplicity in order to avoid overfitting the training data. Finally, note the *default rule* at the end of the rule set. This is added for the case when certain regions of the data space are not represented in the training set.

The key ideas for learning such rule sets are quite similar to the ideas used in decision tree induction. However, instead of recursively partitioning the dataset by optimizing the purity measure over all successor nodes (in the literature, this strategy is also known as *divide-and-conquer* learning), rule learning algorithms only expand a single successor node at a time, thereby learning a complete rule that covers part of the training data. After a complete rule has been learned, all examples that are covered by this rule are removed from the training set, and the procedure is repeated with the remaining examples (this strategy is also known as *separate-and-conquer* learning). Again, pruning is a good idea for rule learning, which means that the rules only need to cover examples that are *mostly* from the same class. It turns out to be advantageous to prune rules immediately after they have been learned, that is before successive rules are learned (Fürnkranz, 1997).

```
IF   MaritalStatus = married
THEN Approved = yes

IF   Sex = female
THEN Approved = yes

IF   Sex = male
THEN Approved = no

DEFAULT  Approved = yes
```

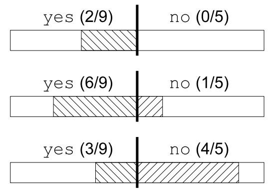

Fig. 1.4: A smaller rule set describing the dataset shown in Table 1.1.

1.5.3 Rule sets versus decision trees

There are several aspects which make rule learning attractive. First of all, decision trees are often quite complex and hard to understand. Quinlan (1993) has noted that even pruned decision trees may be too cumbersome, complex, and inscrutable to provide insight into the domain at hand and has consequently devised procedures for simplifying decision trees into pruned production rule sets (Quinlan, 1987a, 1993). Additional evidence for this comes from Rivest (1987), showing that decision lists (ordered rule sets) with at most k conditions per rule are strictly more expressive than decision trees of depth k. A similar result has been proved by Boström (1995).

Moreover, the restriction of decision tree learning algorithms to nonoverlapping rules imposes strong constraints on learnable rules. One problem resulting from this constraint is the *replicated subtree problem* (Pagallo & Haussler, 1990); it often happens that identical subtrees have to be learned at various places in a decision tree, because of the fragmentation of the example space imposed by the restriction to nonoverlapping rules. Rule learners do not make such a restriction and are thus less susceptible to this problem. An extreme example for this problem has been provided by Cendrowska (1987), who showed that the minimal decision tree for the concept `x` defined as

```
IF A = 3 AND B = 3 THEN Class = x
IF C = 3 AND D = 3 THEN Class = x
```

has 10 interior nodes and 21 leafs assuming that each attribute `A` ... `D` can be instantiated with three different values.

Finally, propositional rule learning algorithms extend naturally to the framework of *inductive logic programming* framework, where the goal is basically the induction of a rule set in first-order logic, e.g., in the form of a Prolog program.[3] First-order background knowledge can also be used for deci-

[3] Prolog is a programming language, enabling knowledge representation in first-order logic (Lloyd, 1987; Sterling & Shapiro, 1994). We will briefly return to

sion tree induction (Watanabe & Rendell, 1991; Lavrač et al., 1991; Kramer, 1996; Blockeel & De Raedt, 1998), but once more, Watanabe & Rendell (1991) have noted that first-order decision trees are usually more complex than first-order rules.

1.6 Descriptive data mining: Induction of patterns

While a decision tree and a set of rules represent a model (a theory) that can be used for classification and/or prediction, the goal of data analysis may be different. Instead of model construction, the goal may be the discovery of individual patterns/rules describing regularities in the data. This form of data analysis is referred to as *descriptive induction* and is frequently used in exploratory data analysis.

As opposed to decision tree and rule set induction, which result in classification models, *association rule learning* is an unsupervised learning method, with no class labels assigned to the examples. Another method for unsupervised learning is *clustering*, while *subgroup discovery*—aimed at finding descriptions of interesting population subgroups—is a descriptive induction method for pattern learning, but is at the same time a form of supervised learning due to a defined property of interest acting as a class.

1.6.1 Association rule learning

The problem of inducing *association rules* (Agrawal et al., 1995) has received much attention in the data mining community. It is defined as follows: given a set of transactions (examples), where each transaction is a set of items, an association rule is an expression of the form B $\rightarrow$ H, where B and H are sets of items, and B $\rightarrow$ H is interpreted as IF B THEN H, meaning that the transactions in a database which contain B tend to contain H as well.

Figure 1.5 shows three examples for association rules that could be discovered in the dataset of Table 1.1. The first rule states that in this dataset, all people with a university education were female. This rule is based on four observations in the dataset. The fraction of entries in the database that satisfy all conditions (both in body and head) is known as the *support* of the rule. Thus, the support of the rule is the ratio of the number of records having true values for all items in B and H to the number of all records in the database. As 4 of a total of 14 persons are both female and have university education, the support of the first rule is $4/14 \approx 0.286$.

The second rule also has a support of $4/14$, because four people in the database do not approve and are male. However, in this case, the strength of the rule is not as strong as in the previous case, because only $4/5 = 0.8$ of

learning in first-order logic in Section 1.7; a systematic treatment of relational rule learning can be found in Chapter 5.

```
IF   Education = university
THEN Sex = female

IF   Approved = no
THEN Sex = male

IF   Education = secondary
 AND MaritalStatus = divorced
THEN HasChildren = no
 AND Approved = yes
```

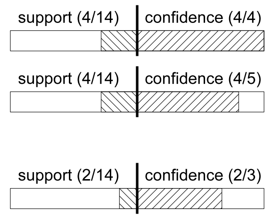

Fig. 1.5: Three rules induced by an association rule learning algorithm.

all persons that do not approve were actually male. This value is called the *confidence* of the rule. It is calculated as the ratio of the number of records having true values for all items in B and H to the number of records having true values for all items in B.

Unlike with classification rules, the head of an association rule may also contain a conjunction of conditions. This is illustrated by the third rule, which states that divorced people with secondary education typically have no children and approve.

In all rules there is no distinction between the class attribute and all other attributes: the class attribute may appear on any side of the rule or not at all. In fact, typically association rules are learned from databases with binary features (called *items*) without any dedicated class attribute. Thus association rule discovery is an unsupervised learning task. Most algorithms, such as the well-known APRIORI algorithm (Agrawal et al., 1995), find all association rules that satisfy minimum support and minimum confidence constraints.

An in-depth survey of association rule discovery is beyond the scope of this book, and, indeed, the subject has already been covered in other monographs (Adamo, 2000; Zhang & Zhang, 2002). We will occasionally touch upon the topic when it seems appropriate (e.g., the level-wise search algorithm, which forms the basis of APRIORI and related techniques, is briefly explained in Section 6.3.2), but for a systematic treatment of the subject we refer the reader to the literature.

1.6.2 Subgroup discovery

In subgroup discovery the task is to find sufficiently large population subgroups that have a significantly different class distribution than the entire population (the entire dataset). Subgroup discovery results in individual rules, where the rule conclusion is a class (the property of interest). The main difference between learning of classification rules and subgroup discovery is that the latter induces single rules (subgroups) of interest, which aim at revealing interesting properties of groups of instances, not necessarily at forming a rule set used for classification.

```
IF   MaritalStatus = single
 AND Sex = male
THEN Approved = no

IF   MaritalStatus = single
THEN Approved = no

IF   Sex = female
THEN Approved = yes
```

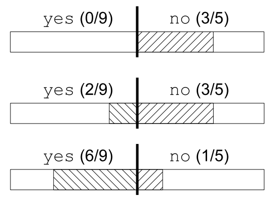

Fig. 1.6: Three subgroup descriptions induced by a subgroup discovery algorithm.

Figure 1.6 shows three subgroup descriptions that have been induced with the MAGNUM OPUS descriptive rule learning system (Webb, 1995).[4] While the first and third rules could also be found by classification rule algorithms (cf. Figure 1.4), the second rule would certainly not be found because it has a comparably low predictive quality. There are almost as many single persons that approve than there are singles that do not approve. Nevertheless, this rule can be considered to be an interesting subgroup because the class distribution of covered instances (2 yes and 3 no) is significantly different than the distribution in the entire dataset (9 yes and 5 no). Conversely, a classification rule algorithm would not find the first rule because if we accept the second rule for classification, adding the first one does not improve classification performance, i.e., it is *redundant* with respect to the second rule. Finally, note that these three rules do not cover all the examples. While it is typically considered important that each rule covers a significant number of examples, it is not necessary that each example be covered by some rule, because the rules will not be used for prediction.

Subgroup discovery and related techniques are covered in depth in Chapter 11 of this book.

1.7 Relational data mining

Both predictive and descriptive data mining are usually performed on a single database relation, consisting of examples represented with values for a fixed number of attributes. However, in practice, the data miner often has to face more complex scenarios. Suppose that data is stored in several tables, e.g.,

[4] The rules are taken from (Kralj Novak et al., 2009).

it has a *relational database* form. In this case the data has to be transformed into a single table in order to be able to use standard data mining techniques. The most common data transformation approach is to select one table as the main table to be used for learning, and try to incorporate the contents of other tables by aggregating the information contained in the tables into summary attributes, which are added to the main table. The problem with such transformations is that some information may be lost while the aggregation may also introduce artifacts, possibly leading to inappropriate data mining results. What one would like to do is to leave data conceptually unchanged and rather use data mining tools that can deal with multirelational data.

Integrating data from multiple tables through joins or aggregation can cause loss of meaning or information. Suppose we are given two relations: `customer(CID,Name,Age,SpendALot)` encodes the ID, name, and age of a customer, and the information whether this customer spends a lot, and `purchase(CID,ProdID,Date,Value,PaymentMode)` encodes a single purchase by a customer with a given ID. Each customer can make multiple purchases, and we are interested in characterizing customers that spend a lot. Integrating the two relations via a natural join will result in a relation `purchase1(CID,Name,Age,SpendALot,ProdID,Date,Value,PaymentMode)`. However, this is problematic because now each row corresponds to a purchase and not to a customer, and we intend to analyze our information with respect to customers. An alternative would be to aggregate the information contained in the `purchase` relation. One possible aggregation could be the relation `customer1(CID,Name,Age,NofPurchases,TotalValue,SpendALot)`, which aggregates the number of purchases and their total value into new attributes. Naturally, some information has been lost during the aggregation process.

The following pattern can be discovered by a relational rule learning system if the relations `customer` and `purchase` are considered together.

```
customer(CID,Name,Age,yes) :-
    Age > 30,
    purchase(CID,PID,D,Value,PM),
    PM = creditcard,
    Value > 100.
```

This pattern, written in a Prolog-like syntax, says: *'a customer spends a lot if she is older than 30, has purchased a product of value more than 100 and paid for it by credit card.'* It would not be possible to induce such a pattern from either of the relations `purchase1` and `customer1` considered on their own.

We will return to relational learning in Chapter 5, where we take a feature-based view on the problem.

1.8 Conclusion

This chapter briefly described several aspects of machine learning and data mining, aiming to provide the background and basic understanding of the topics presented in this book. To conclude, let us make some speculations about future developments in data mining.

With regard to data mining research, every year the research community addresses new open problems and new problem areas, for many of which data mining is able to provide value-added answers and results. Because of the interdisciplinary nature of data mining, there is a big inflow of new knowledge, widening the spectrum of problems that can be solved by the use of this technology. Another reason why data mining has a scientific and commercial future was given by Friedman (1998): "Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it."

To achieve its full commercial exploitation, data mining is still lacking the standardization to the degree of, for example, the standardization available for database systems. There are initiatives in this direction, which will diminish the monopoly of expensive closed-architecture systems. For data mining to be truly successful it is important that data mining tools become available in major database products as well as in standard desktop applications (e.g., spreadsheets). Other important recent developments are open source data mining services, tools for online construction of data mining workflows, as well as the terminology and ingredients of data mining through the development of a data mining ontology (Lavrač et al., 2008, 2009).

In the future, we envisage intensive development and increased usage of data mining in specific domain areas, such as bioinformatics, multimedia, text and web data analysis. On the other hand, as data mining can be used for building surveillance systems, recent research also concentrates on developing algorithms for mining databases without compromising sensitive information (Agrawal & Srikant, 2000). A shift towards automated use of data mining in practical systems is also expected to become very common.

Bibliography

Adamo, J.-M. (2000). *Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms.* Springer-Verlag. */14/*

Adé, H., De Raedt, L., Bruynooghe, M. (1995). Declarative bias for specific-to-general ILP systems. *Machine Learning*, 20(1-2):119–154. Special Issue on Bias Evaluation and Selection. */133/*

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A. I. (1995). Fast discovery of association rules. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) *Advances in Knowledge Discovery and Data Mining*, pp. 307–328. AAAI Press. */13, 14, 121, 248/*

Agrawal, R., Srikant, R. (2000). Privacy-preserving data mining. In Chen, W., Naughton, J. F., Bernstein, P. A. (eds.) *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD-2000)*, pp. 439–450. ACM, Dallas, TX. */17/*

Aha, D. W., Kibler, D., Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6:37–66. */1/*

Ali, K. M., Pazzani, M. J. (1993). HYDRA: A noise-tolerant relational concept learning algorithm. In Bajcsy, R. (ed.) *Proceedings of the 13th Joint International Conference on Artificial Intelligence (IJCAI-93)*, pp. 1064–1071. Morgan Kaufmann, Chambèry, France. */150, 220/*

Allwein, E. L., Schapire, R. E., Singer, Y. (2000). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research*, 1:113–141. */235, 236/*

An, A., Cercone, N. (1998). ELEM2: A learning system for more accurate classifications. In Mercer, R. E., Neufeld, E. (eds.) *Proceedings of the 12th Biennial Conference of the Canadian Society for Computational Studies of Intelligence*, pp. 426–441. Springer-Verlag. */155/*

Asuncion, A., Newman, D. (2007). UCI machine learning repository. */48, 219/*

Atzmüller, M., Puppe, F. (2005). Semi-automatic visual subgroup mining using VIKAMINE. *Journal of Universal Computer Science*, 11(11):1752–1765. Special Issue on Visual Data Mining. */261/*

Atzmüller, M., Puppe, F. (2006). SD-Map - a fast algorithm for exhaustive subgroup discovery. In *Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-06)*, pp. 6–17. */249/*

Atzmüller, M., Puppe, F., Buscher, H.-P. (2005a). Exploiting background knowledge for knowledge-intensive subgroup discovery. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, pp. 647–652. */249/*

Atzmüller, M., Puppe, F., Buscher, H.-P. (2005b). Profiling examiners using intelligent subgroup mining. In *Proceedings of the 10th Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-05)*, pp. 46–51. */250/*

AUMANN, Y., LINDELL, Y. (1999). A statistical theory for quantitative association rules. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99)*, pp. 261–270. */261/*

AZEVEDO, P. J., JORGE, A. J. (2010). Ensembles of jittered association rule classifiers. *Data Mining and Knowledge Discovery*, 12(5):421–453. Special Issue on Global Modeling using Local Patterns. */185/*

BADEA, L. (2001). A refinement operator for theories. In ROUVEIROL, C., SEBAG, M. (eds.) *Proceedings of the 11th International Conference on Inductive Logic Programming (ILP-01)*, pp. 1–14. Springer-Verlag, Strasbourg, France. */183/*

BAY, S. D. (2000). Multivariate discretization of continuous variables for set mining. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2000)*, pp. 315–319. */252/*

BAY, S. D., PAZZANI, M. J. (2001). Detecting group differences: Mining contrast sets. *Data Mining and Knowledge Discovery*, 5(3):213–246. */247, 251, 256, 257, 258, 259/*

BAYARDO JR., R., AGRAWAL, R. (1999). Mining the most interesting rules. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-97)*, pp. 145–154. */168/*

BAYARDO JR., R. J. (1997). Brute-force mining of high-confidence classification rules. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97)*, pp. 123–126. */55, 122, 185/*

BAYARDO JR., R. J. (1998). Efficiently mining long patterns from databases. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data (SIGMOD-98)*, pp. 85–93. */251/*

BENNETT, K. P., BUJA, A., FREUND, W. S. Y., SCHAPIRE, R. E., FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R., BICKEL, P. J., RITOV, Y., BÜHLMANN, P., YU, B. (2008). Responses to Mease & Wyner (2008). *Journal of Machine Learning Research*, 9:157–194. */2/*

BERGADANO, F., GIORDANA, A., SAITTA, L. (1988). Automated concept acquisition in noisy environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:555–578. */119/*

BERGADANO, F., MATWIN, S., MICHALSKI, R. S., ZHANG, J. (1992). Learning two-tiered descriptions of flexible concepts: The POSEIDON system. *Machine Learning*, 8:5–43. */49, 68, 200, 202/*

BERTHOLD, M. R., CEBRON, N., DILL, F., GABRIEL, T. R., KÖTTER, T., MEINL, T., OHL, P., THIEL, K., WISWEDEL, B. (2009). KNIME — The Konstanz information miner. Version 2.0 and beyond. *SIGKDD explorations*, 11:26–31. */5/*

BILLARI, F. C., FÜRNKRANZ, J., PRSKAWETZ, A. (2006). Timing, sequencing, and quantum of life course events: A machine learning approach. *European Journal of Population*, 22(1):37–65. */268, 269, 270/*

BISHOP, C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, UK. */1/*

BISSON, G. (1992). Conceptual clustering in a first order logic representation. In NEUMANN, B. (ed.) *Proceedings of the 10th European Conference on Artificial Intelligence (ECAI-92)*, pp. 458–462. John Wiley & Sons, Vienna, Austria. */245/*

BLASZCZYNSKI, J., STEFANOWSKI, J., ZAJAC, M. (2009). Ensembles of abstaining classifiers based on rule sets. In RAUCH, J., RAS, Z. W., BERKA, P., ELOMAA, T. (eds.) *Proceedings of the 18th International Symposium on Foundations of Intelligent Systems (ISMIS-09)*, pp. 382–391. Springer, Prague, Czech Republic. */223/*

BLOCKEEL, H., DE RAEDT, L. (1998). Top-down induction of first-order logical decision trees. *Artificial Intelligence*, 101(1–2):285–297. */13/*

BLOCKEEL, H., DE RAEDT, L., RAMON, J. (1998). Top-down induction of clustering trees. In SHAVLIK, J. (ed.) *Proceedings of the 15th International Conference on Machine Learning*, pp. 55–63. Morgan Kaufmann, Madison, WI. */245/*

BLOCKEEL, H., VANSCHOREN, J. (2007). Experiment databases: Towards an improved experimental methodology in machine learning. In KOK, J. N., KORONACKI, J., DE MÁNTARAS, R. L., MATWIN, S., MLADENIC, D., SKOWRON, A. (eds.) *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-07)*, pp. 6–17. Springer, Warsawa, Poland. */48/*

BOSE, R. C., RAY CHAUDHURI, D. K. (1960). On a class of error correcting binary group codes. *Information and Control*, 3(1):68–79. */235/*

BOSTRÖM, H. (1995). Covering vs. divide-and-conquer for top-down induction of logic programs. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, pp. 1194–1200. */12/*

BOSTRÖM, H. (2004). Pruning and exclusion criteria for un-ordered incremental reduced error pruning. In FÜRNKRANZ, J. (ed.) *Proceedings of the ECML/PKDD Workshop on Advances in Inductive Rule Learning,*, pp. 17–29. Pisa, Italy. */157, 208/*

BOSTRÖM, H. (2007). Maximizing the area under the ROC curve with decision lists and rule sets. In *Proceedings of the 7th SIAM International Conference on Data Mining (SDM-07)*, pp. 27–34. Minneapolis. */242/*

BOTTA, M., GIORDANA, A. (1993). SMART+: A multi-strategy learning tool. In BAJCSY, R. (ed.) *Proceedings of the 13th Joint International Conference on Artificial Intelligence (IJCAI-93)*, pp. 937–944. Morgan Kaufmann, Chambèry, France. */163, 164/*

BOTTA, M., GIORDANA, A., SAITTA, L. (1992). Comparison of search strategies in learning relations. In NEUMANN, B. (ed.) *Proceedings of the 10th European Conference on Artificial Intelligence (ECAI-92)*, pp. 451–455. John Wiley & Sons, Vienna, Austria. */119, 164/*

BOULESTEIX, A.-L., TUTZ, G., STRIMMER, K. (2003). A CART-based approach to discover emerging patterns in microarray data. *Bioinformatics*, 19(18):2465–2472. */254/*

BRADLEY, R. A., TERRY, M. E. (1952). The rank analysis of incomplete block designs — I. The method of paired comparisons. *Biometrika*, 39:324–345. */230/*

BRATKO, I. (1990). *Prolog Programming for Artificial Intelligence*. Addison-Wesley, Wokingham, 2nd edn. */103/*

BRATKO, I. (1999). Refining complete hypotheses in ILP. In DŽEROSKI, S., FLACH, P. (eds.) *Proceedings of the 9th International Workshop on Inductive Logic Programming (ILP-99)*, pp. 44–55. Springer-Verlag. */183/*

BRATKO, I., MUGGLETON, S. H. (1995). Applications of inductive logic programming. *Communications of the ACM*, 38(11):65–70. */53/*

BREIMAN, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140. */2, 233/*

BREIMAN, L. (2001a). Random forests. *Machine Learning*, 45(1):5–32. */2/*

BREIMAN, L. (2001b). Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231. With comments by D. R. Cox, B. Efron, B. Hoadley, and E. Parzen, and a rejoinder by the author. */2/*

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R., STONE, C. (1984). *Classification and Regression Trees.* Wadsworth & Brooks, Pacific Grove, CA. */1, 3, 148, 200, 211, 244/*

BRIN, S., MOTWANI, R., SILVERSTEIN, C. (1997). Beyond market baskets: Generalizing association rules to correlations. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 265–276. */155/*

BRINGMANN, B., NIJSSEN, S., ZIMMERMANN, A. (2009). Pattern-based classification: A unifying perspective. In KNOBBE, A., FÜRNKRANZ, J. (eds.) *From Local Patterns to Global Models: Proceedings of the ECML/PKDD-09 Workshop (LeGo-09)*, pp. 36–50. Bled, Slovenia. */55, 186/*

BRUHA, I., FRANEK, F. (1996). Comparison of various routines for unknown attribute value processing: The covering paradigm. *International Journal of Pattern Recognition and Artificial Intelligence*, 10(8):939–955. */88/*

BRUNK, C. A., PAZZANI, M. J. (1991). An investigation of noise-tolerant relational concept learning algorithms. In *Proceedings of the 8th International Workshop on Machine Learning (ML-91)*, pp. 389–393. Morgan Kaufmann, Evanston, Illinois. */201, 202/*

BUNTINE, W., NIBLETT, T. (1992). A further comparison of splitting rules for decision-tree induction. *Machine Learning*, 8:75–85. */135/*

CAI, Y., CERCONE, N., HAN, J. (1991). Attribute-oriented induction in relational databases. In PIATETSKY-SHAPIRO, G., FRAWLEY, W. J. (eds.) *Knowledge Discovery in Databases*, pp. 213–228. MIT Press. */70/*

CAMERON-JONES, R. M. (1996). The complexity of batch approaches to reduced error rule set induction. In FOO, N., GOEBEL, R. (eds.) *Proceedings of the 4th Pacific Rim International Conference on Artificial Intelligence (PRICAI-96)*, pp. 348–359. Springer-Verlag, Cairns, Australia. */202, 204, 209/*

CAMERON-JONES, R. M., QUINLAN, J. R. (1993). Avoiding pitfalls when learning recursive theories. In BAJCSY, R. (ed.) *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*, pp. 1050–1057. Chambéry, France. */111/*

CARDOSO, J. S., DA COSTA, J. F. P. (2007). Learning to classify ordinal data: The data replication method. *Journal of Machine Learning Research*, 8:1393–1429. */235/*

CENDROWSKA, J. (1987). PRISM: An algorithm for inducing modular rules. *International Journal of Man-Machine Studies*, 27:349–370. */12, 23, 34, 49, 146, 164/*

CERI, S., GOTTLOB, G., TANCA, L. (1989). What you always wanted to know about Datalog (and never dared to ask). *IEEE Transactions on Knowledge and Data Engineering*, 1(1):146–166. */108, 110/*

CERI, S., GOTTLOB, G., TANCA, L. (1990). *Logic Programming and Databases*. Surveys in Computer Science, Springer-Verlag, Berlin. */108, 110/*

CESTNIK, B. (1990). Estimating probabilities: A crucial task in Machine Learning. In AIELLO, L. (ed.) *Proceedings of the 9th European Conference on Artificial Intelligence (ECAI-90)*, pp. 147–150. Pitman, Stockholm, Sweden. */150/*

CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C., WIRTH, R. (2000). Crisp-dm 1.0: Step-by-step data mining guide. SPSS. Available from `http://www.the-modeling-agency.com/crisp-dm.pdf`. */4/*

CHEN, S. F., GOODMAN, J. T. (1998). *An Empirical Study of Smoothing Techniques for Language Modeling*. Tech. Rep. TR-10-98, Computer Science Group, Harvard University, Cambridge, MA. */215/*

CHOW, M., MOLER, J., MIAN, S. (2001). Identifying marker genes in transcription profiling data using a mixture of feature relevance experts. *Physiological Genomics*, 3(5):99–111. */274/*

CLARK, P., BOSWELL, R. (1991). Rule induction with CN2: Some recent improvements. In *Proceedings of the 5th European Working Session on Learning (EWSL-91)*, pp. 151–163. Springer-Verlag, Porto, Portugal. */10, 20, 50, 146, 150, 170, 199, 228/*

CLARK, P., NIBLETT, T. (1987). Induction in noisy domains. In BRATKO, I., LAVRAČ, N. (eds.) *Progress in Machine Learning*. Sigma Press, Wilmslow, UK. */49/*

Clark, P., Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning*, 3(4):261–283. */1, 20, 43, 44, 49, 50, 67, 68, 118, 146, 195/*

Cloete, I., Van Zyl, J. (2006). Fuzzy rule induction in a set covering framework. *IEEE Transactions on Fuzzy Systems*, 14(1):93–110. */94/*

Cohen, W. W. (1993). Efficient pruning methods for separate-and-conquer rule learning systems. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*, pp. 988–994. Morgan Kaufmann, Chambéry, France. */201, 202, 203, 204, 209, 211/*

Cohen, W. W. (1995). Fast effective rule induction. In Prieditis, A., Russell, S. (eds.) *Proceedings of the 12th International Conference on Machine Learning (ML-95)*, pp. 115–123. Morgan Kaufmann, Lake Tahoe, CA. */20, 45, 52, 67, 68, 145, 163, 189, 198, 210, 211, 213/*

Cohen, W. W. (1996). Learning trees and rules with set-valued features. In *Proceedings of the 13th National Conference on Artificial Intelligene (AAAI-96)*, pp. 709–716. AAAI Press. */69/*

Cohen, W. W., Schapire, R. E., Singer, Y. (1999). Learning to order things. *Journal of Artificial Intelligence Research*, 10:243–270. */243/*

Cohen, W. W., Singer, Y. (1999). A simple, fast, and effective rule learner. In *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI-99)*, pp. 335–342. AAAI/MIT Press, Menlo Park, CA. */178, 186/*

Cook, D. J., Holder, L. B. (1994). Substructure discovery using minimum description length and background knowledge. *Journal of Artificial Intelligence Research*, 1:231–255. */245/*

Cootes, A. P., Muggleton, S. H., Sternberg, M. J. (2003). The automatic discovery of structural principles describing protein fold space. *Journal of Molecular Biology*, 330(4):527–532. */53/*

Crammer, K., Singer, Y. (2002). On the learnability and design of output codes for multiclass problems. *Machine Learning*, 47(2-3):201–233. */235/*

Daly, O., Taniar, D. (2005). Exception rules in data mining. In Khosrow-Pour, M. (ed.) *Encyclopedia of Information Science and Technology*, vol. II, pp. 1144–1148. Idea Group. */261/*

Dasarathy, B. V. (ed.) (1991). *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press, Los Alamitos, CA. */1/*

Davis, J., Burnside, E., Castro Dutra, I. d., Page, D., Santos Costa, V. (2004). Using Bayesian classifiers to combine rules. In *Proceedings of the 3rd SIGKDD Workshop on Multi-Relational Data Mining (MRDM-04)*. */222/*

De Raedt, L. (1992). *Interactive Theory Revision: An Inductive Logic Programming Approach*. Academic Press. */133/*

De Raedt, L. (ed.) (1995). *Advances in Inductive Logic Programming*, *Frontiers in Artificial Intelligence and Applications*, vol. 32. IOS Press. */96/*

De Raedt, L. (1996). Induction in logic. In Michalski, R., Wnek, J. (eds.) *Proceedings of the 3rd International Workshop on Multistrategy Learning (MSL-96)*, pp. 29–38. Machine Learning and Inference Laboratory, George Mason University, Fairfax,VA. */174/*

De Raedt, L. (2008). *Logical and Relational Learning*. Springer-Verlag. */1, 129/*

De Raedt, L., Dehaspe, L. (1997). Clausal discovery. *Machine Learning*, 26(2/3):99–146. Special Issue on *Inductive Logic Programming*. */96, 173/*

De Raedt, L., Van Laer, W. (1995). Inductive constraint logic. In Jantke, K., Shinohara, T., Zeugmann, T. (eds.) *Proceedings of the 5th Workshop on Algorithmic Learning Theory (ALT-95)*, pp. 80–94. Springer-Verlag. */50, 173/*

Dehaspe, L., Toivonen, H. (2001). Discovery of relational association rules. In Džeroski, S., Lavrač, N. (eds.) *Relational Data Mining*, pp. 189–212. Springer. */96/*

Dekel, O., Manning, C. D., Singer, Y. (2004). Log-linear models for label ranking. In Thrun, S., Saul, L. K., Schölkopf, B. (eds.) *Advances in Neural Information Processing Systems (NIPS-03)*, pp. 497–504. MIT Press, Cambridge, MA. */239/*

del Jesus, M. J., González, P., Herrera, F., Mesonero, M. (2007). Evolutionary fuzzy rule induction process for subgroup discovery: A case study in marketing. *IEEE Transactions on Fuzzy Systems*, 15(4):578–592. */249, 250/*

Dembczyński, K., Kotłowski, W., Słowiński, R. (2008). Solving regression by learning an ensemble of decision rules. In Rutkowski, L., Tadeusiewicz, R., Zadeh, L. A., Zurada, J. M. (eds.) *Proceedings of the 9th International Conference on Artificial Intelligence and Soft Computing (ICAISC-08)*, pp. 533–544. Springer-Verlag, Zakopane, Poland. */215, 245/*

Dembczyński, K., Kotłowski, W., Słowiński, R. (2010). ENDER – A statistical framework for boosting decision rules. *Data Mining and Knowledge Discovery*, 12(5):385–420. Special Issue on Global Modeling using Local Patterns. */178, 186/*

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30. */91/*

Demšar, J., Zupan, B., Leban, G. (2004). Orange: From experimental machine learning to interactive data mining. White Paper, Faculty of Computer and Information Science, University of Ljubljana. Available from `http://http://orange.biolab.si/`. */5/*

Dietterich, T. G. (2000). Ensemble methods in machine learning. In Kittler, J., Roli, F. (eds.) *First International Workshop on Multiple Classifier Systems*, pp. 1–15. Springer-Verlag. */180/*

Dietterich, T. G., Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286. */233/*

Domingos, P. (1996a). Linear-time rule induction. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pp. 96–101. AAAI Press. */182/*

Domingos, P. (1996b). Unifying instance-based and rule-based induction. *Machine Learning*, 24:141–168. */181, 182/*

Domingos, P. (1999). The role of Occam's Razor in knowledge discovery. *Data Mining and Knowledge Discovery*, 3(4):409–425. */9/*

Dong, G., Li, J. (1999). Efficient mining of emerging patterns: Discovering trends and differences. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99)*, pp. 43–52. San Diego, CA. */247, 253, 256, 258/*

Dong, G., Zhang, X., Wong, L., Li, J. (1999). CAEP: Classification by aggregating emerging patterns. In *Proceedings of the 2nd International Conference on Discovery Science (DS-99)*, pp. 30–42. */253, 259/*

Dubois, D., Prade, H. (1980). *Fuzzy Sets and Systems.* Academic Press, New York. */93/*

Duda, R. O., Hart, P. E., Stork, D. G. (2000). *Pattern Classification.* John Wiley and Sons, 2nd edn. */1/*

Džeroski, S., Bratko, I. (1992). Handling noise in Inductive Logic Programming. In Muggleton, S. H., Furukawa, K. (eds.) *Proceedings of the 2nd International Workshop on Inductive Logic Programming (ILP-92)*, pp. 109–125. No. TM-1182 in ICOT Technical Memorandum, Institue for New Generation Computer Technology, Tokyo, Japan. */50, 118, 150, 195/*

Džeroski, S., Cestnik, B., Petrovski, I. (1993). Using the m-estimate in rule induction. *Journal of Computing and Information Technology*, 1:37–46. */45/*

Džeroski, S., Lavrač, N. (eds.) (2001). *Relational Data Mining: Inductive Logic Programming for Knowledge Discovery in Databases*. Springer-Verlag. */7, 95, 96, 129/*

Džeroski, S., Schulze-Kremer, S., Heidtke, K. R., Siems, K., Wettschereck, D., Blockeel, H. (1998). Diterpene structure elucidation from 13CNMR spectra with inductive logic programming. *Applied Artificial Intelligence*, 12(5):363–383. Special Issue on First-Order Knowledge Discovery in Databases. */120/*

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis*. AcademicPress, New York. */58/*

Eineborg, M., Boström, H. (2001). Classifying uncovered examples by rule stretching. In Rouveirol, C., Sebag, M. (eds.) *Proceedings of the Eleventh International Conference on Inductive Logic Programming (ILP-01)*, pp. 41–50. Springer Verlag, Strasbourg, France. */223/*

Elmasri, R., Navathe, S. B. (2006). *Fundamentals of Database Systems*. Addison Wesley, 5th edn. */107/*

Escalera, S., Pujol, O., Radeva, P. (2006). Decoding of ternary error correcting output codes. In Trinidad, J. F. M., Carrasco-Ochoa, J. A., Kittler, J. (eds.) *Proceedings of the 11th Iberoamerican Congress in Pattern Recognition (CIARP-06)*, pp. 753–763. Springer-Verlag, Cancun, Mexico. */236/*

Esposito, F., Malerba, D., Semeraro, G. (1993). Decision tree pruning as a search in the state space. In Brazdil, P. (ed.) *Proceedings of the 6th European Conference on Machine Learning (ECML-93)*, pp. 165–184. Springer-Verlag, Vienna, Austria. */200/*

Esposito, F., Semeraro, G., Fanizzi, N., Ferilli, S. (2000). Multistrategy theory revision: Induction and abduction in inthelex. *Machine Learning*, 38(1–2):133–156. */183/*

Everitt, B., Hothorn, T. (2006). *A Handbook of Statistical Analyses Using R*. Chapman & Hall/CRC, Boca Raton, FL. */5/*

Fan, H., Fan, M., Ramamohanarao, K., Liu, M. (2006). Further improving emerging pattern based classifiers via bagging. In *Proceedings of the 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-06)*, pp. 91–96. */253/*

Fan, H., Ramamohanarao, K. (2003a). A Bayesian approach to use emerging patterns for classification. In *Proceedings of the 14th Australasian Database Conference (ADC-03)*, pp. 39–48. */253/*

Fan, H., Ramamohanarao, K. (2003b). Efficiently mining interesting emerging patterns. In *Proceeding of the 4th International Conference on Web-Age Information Management (WAIM-03)*, pp. 189–201. */253/*

Fawcett, T., Niculescu-Mizil, A. (2007). PAV and the ROC convex hull. *Machine Learning*, 68(1):97–106. */215/*

Fawcett, T. E. (2001). Using rule sets to maximize ROC performance. In *Proceedings of the IEEE International Conference on Data Mining (ICDM-01)*, pp. 131–138. */242/*

Fawcett, T. E. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874. */60/*

Fawcett, T. E. (2008). PRIE: A system for generating rulelists to maximize ROC performance. *Data Mining and Knowledge Discovery*, 17(2):207–224. */242/*

Fayyad, U. M., Irani, K. B. (1992). On the handling of continuous-valued attributes in decision tree generation. *Machine Learning*, 8(2):87–102. */82/*

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37–54. */4/*

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) (1995). *Advances in Knowledge Discovery and Data Mining.* AAAI Press, Menlo Park. */2/*

Fensel, D., Wiese, M. (1993). Refinement of rule sets with JoJo. In Brazdil, P. (ed.) *Proceedings of the 6th European Conference on Machine Learning (ECML-93)*, pp. 378–383. Springer-Verlag. */128/*

Fensel, D., Wiese, M. (1994). From JoJo to Frog: Extending a bi-directional strategy to a more flexible three-directional search. In Globig, C., Althoff, K.-D. (eds.) *Beiträge zum 7. Fachgruppentreffen Maschinelles Lernen*, pp. 37–44. Zentrum für Lernende Systeme und Anwendungen, University of Kaiserslautern. Technical Report LSA-95-01. */128/*

Ferri, C., Flach, P., Hernández, J. (2002). Learning decision trees using the area under the ROC curve. In Sammut, C., Hoffmann, A. (eds.) *Proceedings of the 19th International Conference on Machine Learning (ICML-02)*, pp. 139–146. Morgan Kaufmann, Sydney, Australia. */214/*

Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2(2):139–172. */245/*

Flach, P. (1993). Predicate invention in inductive data engineering. In Brazdil, P. B. (ed.) *Proceedings of the 6th European Conference on Machine Learning (ECML-93)*, pp. 83–94. Springer-Verlag, Vienna, Austria. */96/*

Flach, P. (1994). *Simply Logical – Intelligent Reasoning by Example.* John Wiley. */103/*

Flach, P. (1997). Normal forms for inductive logic programming. In Lavrač, N., Džeroski, S. (eds.) *Proceedings of the 7th International Workshop on Inductive Logic Programming (ILP-97)*, pp. 149–156. Prague, Czech Republic. */174/*

Flach, P. (2003). The geometry of ROC space: Using ROC isometrics to understand machine learning metrics. In Fawcett, T., Mishra, N. (eds.) *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 194–201. AAAI Press, Washington, DC. */149, 177/*

Flach, P., Giraud-Carrier, C., Lloyd, J. (1998). Strongly typed inductive concept learning. In *Proceedings of the 8th International Conference on Inductive Logic Programming (ILP-98)*, pp. 185–194. Springer. */96/*

Flach, P., Lachiche, N. (1999). 1BC: A first-order Bayesian classifier. In *Proceedings of the 9th International Workshop on Inductive Logic Programming (ILP-99)*, pp. 92–103. Springer. */101/*

Flach, P., Lachiche, N. (2001). Confirmation-guided discovery of first-order rules with Tertius. *Machine Learning*, 42(1/2):61–95. */96/*

Flach, P., Lavrač, N. (2003). Rule induction. In Berthold, M., Hand, D. J. (eds.) *Intelligent Data Analysis*, pp. 229–267. Springer-Verlag, 2nd edn. */19, 95/*

Flach, P., Wu, S. (2005). Repairing concavities in ROC curves. In Kaelbling, L. P., Saffiotti, A. (eds.) *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, pp. 702–707. Professional Book Center, Edinburgh, Scotland. */214/*

Fodor, J., Roubens, M. (1994). *Fuzzy Preference Modelling and Multicriteria Decision Support.* Kluwer Academic Publishers. */223/*

Frank, E., Hall, M. (2001). A simple approach to ordinal classification. In Raedt, L. D., Flach, P. (eds.) *Proceedings of the 12th European Conference on Machine Learning (ECML-01)*, pp. 145–156. Springer-Verlag, Freiburg, Germany. */241, 243/*

Frank, E., Witten, I. H. (1998). Generating accurate rule sets without global optimization. In Shavlik, J. (ed.) *Proceedings of the 15th International Conference on Machine Learning (ICML-98)*, pp. 144–151. Morgan Kaufmann, Madison, Wisconsin. */55, 170, 183, 244/*

FREUND, Y., SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139. */2/*

FRIEDMAN, A. (1986). *Fundamentals of Logic Design and Switching Theory*. Computer Science Press. */173/*

FRIEDMAN, J. H. (1996). *Another Approach to Polychotomous Classification*. Tech. rep., Department of Statistics, Stanford University, Stanford, CA. */230/*

FRIEDMAN, J. H. (1998). Data mining and statistics: What's the connection? In *Computing Science and Statistics: Proceedings of the 29th Symposium on the Interface*. Interface Foundation of North America. */2, 17/*

FRIEDMAN, J. H., FISHER, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, 9(2):123–143. */2, 10, 20, 161, 261/*

FRIEDMAN, J. H., POPESCU, B. E. (2008). Predictive learning via rule ensembles. *Annals of Applied Statistics*, 2:916–954. */215, 245/*

FRIEDMAN, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32:675–701. */91/*

FRIEDMAN, N., GEIGER, D., GOLDSZMIDT, M. (1997). Bayesian networks classifiers. *Machine Learning*, 29:131–161. */222/*

FÜRNKRANZ, J. (1994a). FOSSIL: A robust relational learner. In BERGADANO, F., DE RAEDT, L. (eds.) *Proceedings of the 7th European Conference on Machine Learning (ECML-94)*, pp. 122–137. Springer-Verlag, Catania, Italy. */52, 155, 197, 198/*

FÜRNKRANZ, J. (1994b). Top-down pruning in relational learning. In COHN, A. G. (ed.) *Proceedings of the 11th European Conference on Artificial Intelligence (ECAI-94)*, pp. 453–457. John Wiley & Sons, Amsterdam, The Netherlands. */211/*

FÜRNKRANZ, J. (1995). A tight integration of pruning and learning (extended abstract). In LAVRAČ, N., WROBEL, S. (eds.) *Proceedings of the 8th European Conference on Machine Learning (ECML-95)*, pp. 291–294. Springer-Verlag, Heraclion, Greece. */212/*

FÜRNKRANZ, J. (1997). Pruning algorithms for rule learning. *Machine Learning*, 27(2):139–171. */11, 187, 197, 209, 210, 211/*

FÜRNKRANZ, J. (2002a). A pathology of bottom-up hill-climbing in inductive rule learning. In CESA-BIANCHI, N., NUMAO, M., REISCHUK, R. (eds.) *Proceedings of the 13th European Conference on Algorithmic Learning Theory (ALT-02)*, pp. 263–277. Springer-Verlag, Lübeck, Germany. */125, 126/*

FÜRNKRANZ, J. (2002b). Round robin classification. *Journal of Machine Learning Research*, 2:721–747. */217, 230, 232, 233/*

FÜRNKRANZ, J. (2003). Round robin ensembles. *Intelligent Data Analysis*, 7(5):385–404. */233, 241/*

FÜRNKRANZ, J. (2005). From local to global patterns: Evaluation issues in rule learning algorithms. In MORIK, K., BOULICAUT, J.-F., SIEBES, A. (eds.) *Local Pattern Detection*, pp. 20–38. Springer-Verlag. */186/*

FÜRNKRANZ, J., FLACH, P. (2003). An analysis of rule evaluation metrics. In FAWCETT, T., MISHRA, N. (eds.) *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 202–209. AAAI Press, Washington, DC. */61/*

FÜRNKRANZ, J., FLACH, P. (2004). An analysis of stopping and filtering criteria for rule learning. In BOULICAUT, J.-F., ESPOSITO, F., GIANNOTTI, F., PEDRESCHI, D. (eds.) *Proceedings of the 15th European Conference on Machine Learning (ECML-04)*, pp. 123–133. Springer-Verlag, Pisa, Italy. */61, 168/*

Fürnkranz, J., Flach, P. (2005). ROC 'n' rule learning – Towards a better understanding of covering algorithms. *Machine Learning*, 58(1):39–77. */57, 58, 60, 135, 149, 152, 155, 187, 193, 242/*

Fürnkranz, J., Hüllermeier, E. (2003). Pairwise preference learning and ranking. In Lavrač, N., Gamberger, D., Blockeel, H., Todorovski, L. (eds.) *Proceedings of the 14th European Conference on Machine Learning (ECML-03)*, pp. 145–156. Springer-Verlag, Cavtat, Croatia. */238, 239/*

Fürnkranz, J., Hüllermeier, E. (eds.) (2010a). *Preference Learning*. Springer-Verlag. */237, 242/*

Fürnkranz, J., Hüllermeier, E. (2010b). Preference learning and ranking by pairwise comparison. In Fürnkranz, J., Hüllermeier, E. (eds.) *Preference Learning*, pp. 65–82. Springer-Verlag. */217/*

Fürnkranz, J., Hüllermeier, E., Loza Mencía, E., Brinker, K. (2008). Multilabel classification via calibrated label ranking. *Machine Learning*, 73(2):133–153. */240/*

Fürnkranz, J., Hüllermeier, E., Vanderlooy, S. (2009). Binary decomposition methods for multipartite ranking. In Buntine, W. L., Grobelnik, M., Mladenić, D., Shawe-Taylor, J. (eds.) *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD-09)*, vol. Part I, pp. 359–374. Springer-Verlag, Bled, Slovenia. 2009. */243/*

Fürnkranz, J., Knobbe, A. (2010). Special issue on global modeling using local patterns. *Data Mining and Knowledge Discovery*, 12(5). */184/*

Fürnkranz, J., Sima, J. F. (2010). On exploiting hierarchical label structure with pairwise classifiers. *SIGKDD Explorations*, 12(2):21–25. Special Issue on Mining Unexpected Results. */231, 241/*

Fürnkranz, J., Widmer, G. (1994). Incremental Reduced Error Pruning. In Cohen, W. W., Hirsh, H. (eds.) *Proceedings of the 11th International Conference on Machine Learning (ML-94)*, pp. 70–77. Morgan Kaufmann, New Brunswick, NJ. */142, 207, 210/*

Gamberger, D., Lavrač, N. (2000). Confirmation rule sets. In Zighed, D. A., Komorowski, J., Žytkow, J. (eds.) *Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD-00)*, pp. 34–43. Springer, Berlin, Lyon, France. */178, 186/*

Gamberger, D., Lavrač, N. (2002). Expert-guided subgroup discovery: Methodology and application. *Journal of Artificial Intelligence Research*, 17:501–527. */61, 73, 148, 149, 180, 249, 250, 259, 263, 268/*

Gamberger, D., Lavrač, N., Fürnkranz, J. (2008). Handling unknown and imprecise attribute values in propositional rule learning: A feature-based approach. In Ho, T.-B., Zhou, Z.-H. (eds.) *Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence (PRICAI-08)*, pp. 636–645. Springer-Verlag, Hanoi, Vietnam. */88/*

Gamberger, D., Lavrač, N., Krstačić, G. (2002a). Confirmation rule induction and its applications to coronary heart disease diagnosis and risk group discovery. *Journal of Intelligent and Fuzzy Systems*, 12(1):35–48. */223/*

Gamberger, D., Lavrač, N., Wettschereck., D. (2002b). Subgroup visualization: A method and application in population screening. In *Proceedings of the 7th International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-02)*, pp. 31–35. */261, 263/*

Gamberger, D., Lavrač, N., Zelezny, F., Tolar, J. (2004). Induction of comprehensible models for gene expression datasets by subgroup discovery methodology. *Journal of Biomedical Informatics*, 37(4):269–284. */94, 280/*

Garriga, G. C., Kralj, P., Lavrač, N. (2006). Closed sets for labeled data. In *Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-06)*, pp. 163 – 174. */260/*

Geibel, P., Wysotzki, F. (1996). Learning relational concepts with decision trees. In Saitta, L. (ed.) *Proceedings of the 13th International Conference on Machine Learning (ICML-96)*, pp. 166–174. Morgan Kaufmann Publishers. */103/*

Gelfand, S., Ravishankar, C., Delp, E. (1991). An iterative growing and pruning algorithm for classification tree design. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):163–174. */200/*

Geng, L., Hamilton, H. J. (2006). Interestingness measures for data mining: A survey. *ACM Computing Surveys*, 38(3). */136/*

Georgeff, M. P., Wallace, C. S. (1984). A general criterion for inductive inference. In O'Shea, T. (ed.) *Proceedings of the Sixth European Conference on Artificial Intelligence (ECAI-84)*, pp. 473–482. Elsevier, Amsterdam. */163/*

Ghani, R. (2000). Using error-correcting codes for text classification. In *Proceedings of the 17th International Conference on Machine Learning (ICML-00)*, pp. 303–310. Morgan Kaufmann Publishers. */235/*

Giordana, A., Sale, C. (1992). Learning structured concepts using genetic algorithms. In Sleeman, D., Edwards, P. (eds.) *Proceedings of the 9th International Workshop on Machine Learning (ML-92)*, pp. 169–178. Morgan Kaufmann, Edinburgh. */123, 164/*

Goethals, B. (2005). Frequent set mining. In Maimon, O., Rokach, L. (eds.) *The Data Mining and Knowledge Discovery Handbook*, pp. 377–397. Springer-Verlag. */122/*

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA. */123/*

Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaaseenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M., Bloomfield, C., Lander, E. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537. */274/*

Gönen, M., Heller, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92(4):965–970. */243/*

Grant, J., Minker, J. (1992). The impact of logic programming on databases. *Communications of the ACM*, 35(3):66–81. */110/*

Groff, J. R., Weinberg, P. N. (2002). *SQL, the complete reference*. McGraw-Hill Osborne Media, 2nd edn. */108/*

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD explorations*, 11(1):10–18. */5/*

Han, J., Cai, Y., Cercone, N. (1992). Knowledge discovery in databases: An attribute-oriented approach. In *Proceedings of the 18th Conference on Very Large Data Bases (VLDB-92)*, pp. 547–559. Vancouver, Canada. */70/*

Han, J., Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers. */2, 70/*

Han, J., Pei, J., Yin, Y., Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1):53–87. */122/*

Hand, D. J. (2002). Pattern detection and discovery. In Hand, D. J., Adams, N. M., Bolton, R. J. (eds.) *Pattern detection and discovery: Proceedings of the ESF Exploratory Workshop*, pp. 1–12. Springer-Verlag. */184/*

Har-Peled, S., Roth, D., Zimak, D. (2002). Constraint classification: A new approach to multiclass classification. In Cesa-Bianchi, N., Numao, M., Reischuk, R. (eds.) *Proceedings of the 13th International Conference on Algorithmic Learning Theory (ALT-02)*, pp. 365–379. Springer, Lübeck, Germany. */239/*

Hart, P. E., Nilsson, N. J., Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107. */120/*

Hastie, T., Tibshirani, R. (1998). Classification by pairwise coupling. In Jordan, M., Kearns, M., Solla, S. (eds.) *Advances in Neural Information Processing Systems 10 (NIPS-97)*, pp. 507–513. MIT Press. */230, 239/*

Hastie, T., Tibshirani, R., Friedman, J. H. (2001). *The Elements of Statistical Learning*. Springer-Verlag. */1/*

Helft, N. (1989). Induction as nonmonotonic inference. In Brachman, R. J., Levesque, H. J., Reiter, R. (eds.) *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR-89)*, pp. 149–156. Morgan Kaufmann, Toronto, Canada. */111/*

Hernández-Orallo, J., Ramírez-Quintana, M. (1999). A complete schema for inductive functional logic programming. In Džeroski, S., Flach, P. (eds.) *Proceedings of the 9th International Workshop on Inductive Logic Programming (ILP-99)*, pp. 116–127. Springer-Verlag, Bled, Slovenia. */96/*

Hilderman, R. J., Peckham, T. (2005). A statistically sound alternative approach to mining contrast sets. In *Proceedings of the 4th Australia Data Mining Conference (AusDM-05)*, pp. 157–172. */251/*

Hipp, J., Güntzer, U., Nakhaeizadeh, G. (2000). Algorithms for association rule mining – a general survey and comparison. *SIGKDD explorations*, 2(1):58–64. */122/*

Hocquenghem, A. (1959). Codes correcteurs d'erreurs. *Chiffres*, 2:147–156. In French. */235/*

Holmes, G., Hall, M., Frank, E. (1999). Generating rule sets from model trees. In Foo, N. Y. (ed.) *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence (AI-99)*, pp. 1–12. Springer, Sydney, Australia. */244/*

Holte, R., Acker, L., Porter, B. (1989). Concept learning and the problem of small disjuncts. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, pp. 813–818. Morgan Kaufmann, Detroit, MI. */159, 210/*

Hoos, H. H., Stützle, T. (2004). *Stochastic Local Search: Foundations and Applications*. Morgan Kaufmann, San Francisco, CA. */183/*

Hsu, C.-W., Lin, C.-J. (2002). A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425. */230/*

Hühn, J., Hüllermeier, E. (2009). FR3: A fuzzy rule learner for inducing reliable classifiers. *IEEE Transactions on Fuzzy Systems*, 17(1):138–149. */223, 230/*

Hühn, J., Hüllermeier, E. (2009). Furia: An algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 19(3):293–319. */52, 93/*

Hüllermeier, E. (2011). Fuzzy sets in machine learning and data mining. *Applied Soft Computing*, 11(2):1493–1505. */93/*

Hüllermeier, E., Fürnkranz, J. (2010). On predictive accuracy and risk minimization in pairwise label ranking. *Journal of Computer and System Sciences*, 76(1):49–62. */239/*

Hüllermeier, E., Fürnkranz, J., Cheng, W., Brinker, K. (2008). Label ranking by learning pairwise preferences. *Artificial Intelligence*, 172:1897–1916. */238, 239/*

Hüllermeier, E., Vanderlooy, S. (2009). Why fuzzy decision trees are good rankers. *IEEE Transactions on Fuzzy Systems*, 17(6):1233–1244. */215/*

Janssen, F., Fürnkranz, J. (2009). A re-evaluation of the over-searching phenomenon in inductive rule learning. In Park, H., Parthasarathy, S., Liu, H., Obradovic, Z. (eds.) *Proceedings of the SIAM International Conference on Data Mining (SDM-09)*, pp. 329–340. Sparks, Nevada. */119, 156/*

Janssen, F., Fürnkranz, J. (2010). On the quest for optimal rule learning heuristics. *Machine Learning*, 78(3):343–379. */144, 159, 165, 166/*

Janssen, F., Fürnkranz, J. (2011). Heuristic rule-based regression via dynamic reduction to classification. In Walsh, T. (ed.) *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 1330–1335. AAAI Press, Barcelona, Spain. */245/*

Jenkole, J., Kralj, P., Lavrač, N., Sluga, A. (2007). A data mining experiment on manufacturing shop floor data. In *Proceedings of the 40th CIRP International Seminar on Manufacturing Systems*. */250/*

Joachims, T. (2002). Optimizing search engines using clickthrough data. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-02)*, pp. 133–142. ACM Press. */243/*

Joachims, T. (2006). Training linear SVMs in linear time. In Eliassi-Rad, T., Ungar, L. H., Craven, M., Gunopulos, D. (eds.) *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06*, pp. 217–226. */243/*

Joshi, S., Ramakrishnan, G., Srinivasan, A. (2008). Feature construction using theory-guided sampling and randomised search. In Zelezný, F., Lavrac, N. (eds.) *Proceedings of the 18th International Conference on Inductive Logic Programming (ILP-08)*, pp. 140–157. Springer, Prague, Czech Republic. */53/*

Jovanoski, V., Lavrač, N. (2001). Classification rule learning with APRIORI-C. In Brazdil, P., Jorge, A. (eds.) *Proceedings of the 10th Portuguese Conference on Artificial Intelligence (EPIA 2001)*, pp. 44–51. Springer-Verlag, Porto, Portugal. */55, 122, 185, 191/*

Karalič, A., Bratko, I. (1997). First order regression. *Machine Learning*, 26(2/3):147–176. Special Issue on *Inductive Logic Programming*. */244/*

Kaufman, K. A., Michalski, R. S. (2000). An adjustable rule learner for pattern discovery using the AQ methodology. *Journal of Intelligent Information Systems*, 14:199–216. */49/*

Kavšek, B., Lavrač, N. (2006). Apriori-SD: Adapting association rule learning to subgroup discovery. *Applied Artificial Intelligence*, 20(7):543–583. */122, 249, 259/*

King, R. D., Whelan, K. E., Jones, F. M., Reiser, P., Bryant, C., Muggleton, S., Kell, D., Oliver, S. (2004). Functional genomic hypothesis generation and experimentation by a robot. *Nature*, 427:247–252. */53/*

Kirkpatrick, S., Gelatt, C., Vecchi, M. (1983). Optimization by simulated annealing. *Science*, 220:671–680. */122/*

Kirsten, M., Wrobel, S., Horvath, T. (2001). Distance based approaches to relational learning and clustering. In Džeroski, S., Lavrač, N. (eds.) *Relational Data Mining*, pp. 213–232. Springer. */96/*

Kittler, J., Ghaderi, R., Windeatt, T., Matas, J. (2003). Face verification via error correcting output codes. *Image and Vision Computing*, 21(13-14):1163–1169. */235/*

Klösgen, W. (1992). Problems for knowledge discovery in databases and their treatment in the statistics interpreter EXPLORA. *International Journal of Intelligent Systems*, 7(7):649–673. */157, 159/*

Klösgen, W. (1996). Explora: A multipattern and multistrategy discovery assistant. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) *Advances in Knowledge Discovery and Data Mining*, chap. 10, pp. 249–271. AAAI Press. */143, 157, 168, 247, 249, 250/*

KLÖSGEN, W., MAY, M. (2002). Spatial subgroup mining integrated in an object-relational spatial database. In *Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-02)*, pp. 275–286. */249/*

KLÖSGEN, W., MAY, M., PETCH, J. (2003). Mining census data for spatial effects on mortality. *Intelligent Data Analysis*, 7(6):521–540. */250/*

KNERR, S., PERSONNAZ, L., DREYFUS, G. (1990). Single-layer learning revisited: A stepwise procedure for building and training a neural network. In FOGELMAN SOULIÉ, F., HÉRAULT, J. (eds.) *Neurocomputing: Algorithms, Architectures and Applications*, *NATO ASI Series*, vol. F68, pp. 41–50. Springer-Verlag. */230/*

KNERR, S., PERSONNAZ, L., DREYFUS, G. (1992). Handwritten digit recognition by neural networks with single-layer training. *IEEE Transactions on Neural Networks*, 3(6):962–968. */230/*

KNOBBE, A. J., CRÉMILLEUX, B., FÜRNKRANZ, J., SCHOLZ, M. (2008). From local patterns to global models: The LeGo approach to data mining. In FÜRNKRANZ, J., KNOBBE, A. J. (eds.) *From Local Patterns to Global Models: Proceedings of the ECML/PKDD-08 Workshop (LeGo-08)*, pp. 1–16. Antwerp, Belgium. */184, 185/*

KOLLER, D., SAHAMI, M. (1997). Hierarchically classifying documents using very few words. In *Proceedings of the 14th International Conference on Machine Learning (ICML-97)*, pp. 170–178. Nashville. */241/*

KONG, E. B., DIETTERICH, T. G. (1995). Error-correcting output coding corrects bias and variance. In *Proceedings of the 12th International Conference on Machine Learning (ICML-95)*, pp. 313–321. Morgan Kaufmann. */235/*

KONONENKO, I., KOVAČIČ, M. (1992). Learning as optimization: Stochastic generation of multiple knowledge. In SLEEMAN, D., EDWARDS, P. (eds.) *Proceedings of the 9th International Workshop on Machine Learning (ML-92)*, pp. 257–262. Morgan Kaufmann. */122, 128/*

KOTSIANTIS, S., ZAHARAKIS, I., PINTELAS, P. (2006). Supervised machine learning: A review of classification techniques. *Artificial Intelligence Review*, 26:159–190. */6/*

KOVAČIČ, M. (1991). Markovian neural networks. *Biological Cybernetics*, 64:337–342. */122/*

KOVAČIČ, M. (1994a). MDL-heuristics in ILP revised. In *Proceedings of the ML-COLT-94 Workshop on Applications of Descriptional Complexity to Inductive, Statistical, and Visual Inference.* */163/*

KOVAČIČ, M. (1994b). *Stochastic Inductive Logic Programming.* Ph.D. thesis, Department of Computer and Information Science, University of Ljubljana. */122, 163/*

KRALJ, P., GRUBEŠIČ, A., TOPLAK, N., GRUDEN, K., LAVRAČ, N., GARRIGA, G. C. (2006). Application of closed itemset mining for class labeled data in functional genomics. *Informatica Medica Slovenica*, 11(1):40–45. */260/*

KRALJ, P., LAVRAČ, N., GAMBERGER, D., KRSTAČIĆ, A. (2007a). Contrast set mining for distinguishing between similar diseases. In *Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME-07)*, pp. 109–118. */250/*

KRALJ, P., LAVRAČ, N., GAMBERGER, D., KRSTAČIĆ, A. (2007b). Contrast set mining through subgroup discovery applied to brain ischaemia data. In *Proceedings of the 11th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD-07)*, pp. 579–586. */250, 253/*

KRALJ, P., LAVRAČ, N., ZUPAN, B. (2005). Subgroup visualization. In *Proceedings of the 8th International Multiconference Information Society (IS-05)*, pp. 228–231. Ljubljana, Slovenia. */261, 264/*

Kralj Novak, P., Lavrač, N., Webb, G. I. (2009). Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research*, 10:377–403. */15, 168, 247, 258/*

Kramer, S. (1996). Structural regression trees. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI-96)*, pp. 812–819. AAAI Press. */13, 244/*

Kramer, S., Frank, E. (2000). Bottom-up propositionalization. In *Proceedings of the ILP-2000 Work-In-Progress Track*, pp. 156–162. Imperial College, London. */103/*

Kramer, S., Lavrač, N., Flach, P. (2001). Propositionalization approaches to relational data mining. In Džeroski, S., Lavrač, N. (eds.) *Relational Data Mining*, pp. 262–291. Springer-Verlag, Berlin. */54, 101/*

Kramer, S., Pfahringer, B., Helma, C. (2000). Stochastic propositionalization of non-determinate background knowledge. In *Proceedings of the 8th International Conference on Inductive Logic Programming (ILP-2000)*, pp. 80–94. Springer-Verlag. */103/*

Kressel, U. H.-G. (1999). Pairwise classification and support vector machines. In Schölkopf, B., Burges, C., Smola, A. (eds.) *Advances in Kernel Methods: Support Vector Learning*, chap. 15, pp. 255–268. MIT Press, Cambridge, MA. */230/*

Krogel, M. A., Rawles, S., Železný, F., Flach, P., Lavrač, N., Wrobel, S. (2003). Comparative evaluation of approaches to propositionalization. In Horvath, T., Yamamoto, A. (eds.) *Proceedings of the 13th International Conference on Inductive Logic Programming (ILP-2003)*, pp. 197–214. Springer-Verlag. */102/*

Kullback, S., Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86. */147/*

Landwehr, N., Kersting, K., De Raedt, L. (2007). Integrating Naive Bayes and FOIL. *Journal of Machine Learning Research*, 8:481–507. */222/*

Langford, J., Oliveira, R., Zadrozny, B. (2006). Predicting conditional quantiles via reduction to classification. In *Proceedings of the 22nd Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pp. 257–264. AUAI Press, Arlington, Virginia. */244/*

Langley, P. (1996). *Elements of Machine Learning*. Morgan Kaufmann. */1/*

Langley, P., Simon, H. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, 38(11):54–64. */267/*

Lavrač, N., Cestnik, B., Džeroski, S. (1992a). Search heuristics in empirical Inductive Logic Programming. In *Logical Approaches to Machine Learning, Workshop Notes of the 10th European Conference on AI*. Vienna, Austria. */165/*

Lavrač, N., Cestnik, B., Džeroski, S. (1992b). Use of heuristics in empirical Inductive Logic Programming. In Muggleton, S. H., Furukawa, K. (eds.) *Proceedings of the 2nd International Workshop on Inductive Logic Programming (ILP-92)*. No. TM-1182 in ICOT Technical Memorandum, Institute for New Generation Computer Technology, Tokyo, Japan. */165/*

Lavrač, N., Cestnik, B., Gamberger, D., Flach, P. A. (2004a). Decision support through subgroup discovery: Three case studies and the lessons learned. *Machine Learning*, 57(1-2):115–143. Special issue on Data Mining Lessons Learned. */250/*

Lavrač, N., Džeroski, S. (1994a). *Inductive Logic Programming: Techniques and Applications*. Ellis Horwood. */1, 7, 54, 67, 77, 95, 96, 101, 129/*

Lavrač, N., Džeroski, S. (1994b). Weakening the language bias in LINUS. *Journal of Experimental and Theoretical Artificial Intelligence*, 6:95–119. */110/*

LAVRAČ, N., DŽEROSKI, S., GROBELNIK, M. (1991). Learning nonrecursive definitions of relations with LINUS. In *Proceedings of the 5th European Working Session on Learning (EWSL-91)*, pp. 265–281. Springer-Verlag, Porto, Portugal. */13, 54, 67, 77, 101/*

LAVRAČ, N., FLACH, P. (2001). An extended transformation approach to inductive logic programming. *ACM Transactions on Computational Logic*, 2(4):458–494. */67, 101/*

LAVRAČ, N., FLACH, P., ZUPAN, B. (1999a). Rule evaluation measures: A unifying view. In DŽEROSKI, S., FLACH, P. (eds.) *Proceedings of the 9th International Workshop on Inductive Logic Programming (ILP-99)*, pp. 174–185. Springer-Verlag. */143, 168/*

LAVRAČ, N., FÜRNKRANZ, J., GAMBERGER, D. (2010). Explicit feature construction and manipulation for covering rule learning algorithms. In KORONACKI, J., RAS, Z., WIERZCHON, S. T., KACPRZYK, J. (eds.) *Advances in Machine Learning II — Dedicated to the Memory of Professor Ryszard S. Michalski*, pp. 121–146. Springer-Verlag. */65, 88/*

LAVRAČ, N., GAMBERGER, D., JOVANOSKI, V. (1999b). A sudy of relevance for learning in deductive databases. *The Journal of Logic Programming*, 40(2/3):215–249. */78/*

LAVRAČ, N., GROBELNIK, M. (2003). Data mining. In MLADENIĆ, D., LAVRAČ, N., BOHANEC, M., MOYLE, S. (eds.) *Data Mining and Decision Support: Integration and Collaboration*, pp. 3–14. Kluwer Academic Publishers. */1/*

LAVRAČ, N., KAVŠEK, B., FLACH, P., TODOROVSKI, L. (2004b). Subgroup discovery with CN2-SD. *Journal of Machine Learning Research*, 5:153–188. */143, 167, 178, 179, 180, 249, 250, 259/*

LAVRAČ, N., KOK, J., DE BRUIN, J., PODPEČAN, V. (eds.) (2008). *Proceedings of the ECML-PKDD-08 Workshop on Third Generation Generation Data Mining: Towards Service-Oriented Knowledge Discovery (SoKD-08)*. Antwerp, Belgium. */17/*

LAVRAČ, N., KRALJ, P., GAMBERGER, D., KRSTAČIĆ, A. (2007). Supporting factors to improve the explanatory potential of contrast set mining: Analyzing brain ischaemia data. In *Proceedings of the 11th Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON-07)*, pp. 157–161. */250/*

LAVRAČ, N., PODPEČAN, V., KOK, J., DE BRUIN, J. (eds.) (2009). *Proceedings of the ECML-PKDD-09 Workshop on Service-Oriented Knowledge Discovery (SoKD-09)*. Bled, Slovenia. */17/*

LI, J., DONG, G., RAMAMOHANARAO, K. (2000). Instance-based classification by emerging patterns. In *Proceedings of the 14th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-2000)*, pp. 191–200. */253/*

LI, J., DONG, G., RAMAMOHANARAO, K. (2001a). Making use of the most expressive jumping emerging patterns for classification. *Knowledge and Information Systems*, 3(2):1–29. */253/*

LI, J., LIU, H., DOWNING, J. R., YEOH, A. E.-J., WONG, L. (2003). Simple rules underlying gene expression profiles of more than six subtypes of acute lymphoblastic leukemia (ALL) patients. *Bioinformatics*, 19(1):71–78. */254/*

LI, J., WONG, L. (2002a). Geography of differences between two classes of data. In *Proceedings of the 6th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD-02)*, pp. 325–337. */274, 275/*

LI, J., WONG, L. (2002b). Identifying good diagnostic gene groups from gene expression profiles using the concept of emerging patterns. *Bioinformatics*, 18(10):1406–1407. */254/*

Li, W., Han, J., Pei, J. (2001b). CMAR: Accurate and efficient classification based on multiple class-association rules. In *Proceedings of the IEEE Conference on Data Mining (ICDM-01)*, pp. 369–376. */55, 185/*

Lin, H.-T., Lin, C.-J., Weng, R. C. (2007). A note on Platt's probabilistic outputs for support vector machines. *Machine Learning*, 68(3):267–276. */215/*

Lin, J., Keogh, E. (2006). Group SAX: Extending the notion of contrast sets to time series and multimedia data. In *Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-06)*, pp. 284–296. */252/*

Lindgren, T., Boström, H. (2004). Resolving rule conflicts with double induction. *Intelligent Data Analysis*, 8(5):457–468. */222/*

Liu, B., Hsu, W., Han, H.-S., Xia, Y. (2000a). Mining changes for real-life applications. In *In Proceedings of the 2nd International Conference on Data Warehousing and Knowledge Discovery (DaWaK-2000)*, pp. 337–346. */260/*

Liu, B., Hsu, W., Ma, Y. (1998). Integrating classification and association rule mining. In Agrawal, R., Stolorz, P., Piatetsky-Shapiro, G. (eds.) *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD-98)*, pp. 80–86. */54, 122, 184, 191/*

Liu, B., Hsu, W., Ma, Y. (2001). Discovering the set of fundamental rule changes. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-01)*, pp. 335–340. */260/*

Liu, B., Ma, Y., Wong, C.-K. (2000b). Improving an exhaustive search based rule learner. In Zighed, D. A., Komorowski, H. J., Zytkow, J. M. (eds.) *Proceedings of the 4th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-2000)*, pp. 504–509. Lyon, France. */54, 184, 191/*

Lloyd, J. W. (1987). *Foundations of Logic Programming*. Springer-Verlag, Berlin, 2nd, extended edn. */12, 103, 105/*

Loza Mencía, E., Park, S.-H., Fürnkranz, J. (2009). Efficient voting prediction for pairwise multilabel classification. In *Proceedings of the 17th European Symposium on Artificial Neural Networks (ESANN-09)*, pp. 117–122. d-side publications, Bruges, Belgium. */241/*

Lu, B.-L., Ito, M. (1999). Task decomposition and module combination based on class relations: A modular neural network for pattern classification. *IEEE Transactions on Neural Networks*, 10(5):1244–1256. */230/*

Macskassy, S. A., Provost, F., Rosset, S. (2005). ROC confidence bands: An empirical evaluation. In *Proceedings of the 22nd International Conference on Machine Learning (ICML-05)*, pp. 537–544. ACM Press, Bonn, Germany. */60/*

MacWilliams, F. J., Sloane, N. J. A. (1983). *The Theory of Error-Correcting Codes*. North-Holland Mathematical Library, North Holland. */233/*

Major, J. A., Mangano, J. J. (1995). Selecting among rules induced from a hurricane database. *Journal of Intelligent Information Systems*, 4(1):39–52. */157, 167/*

Manning, C. D., Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts. */215/*

May, M., Ragia, L. (2002). Spatial subgroup discovery applied to the analysis of vegetation data. In *Proceedings of the 4th International Conference on Practical Aspects of Knowledge Management (PAKM-2002)*, pp. 49–61. */250/*

Mease, D., Wyner, A. (2008). Evidence contrary to the statistical view of boosting. *Journal of Machine Learning Research*, 9:131–156. */2, 304/*

Melvin, I., Ie, E., Weston, J., Noble, W. S., Leslie, C. (2007). Multi-class protein classification using adaptive codes. *Journal of Machine Learning Research*, 8:1557–1581. */235/*

MICHALSKI, R. S. (1969). On the quasi-minimal solution of the covering problem. In *Proceedings of the 5th International Symposium on Information Processing (FCIP-69)*, vol. A3 (Switching Circuits), pp. 125–128. Bled, Yugoslavia. */3, 40, 49, 245/*

MICHALSKI, R. S. (1973). AQVAL/1 — computer implementation of a variable-valued logic system VL_1 and examples of its application to pattern recognition. In *Proceedings of the 1st International Joint Conference on Pattern Recognition*, pp. 3–17. */68/*

MICHALSKI, R. S. (1980). Pattern recognition and rule-guided inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:349–361. */3, 49, 68, 245/*

MICHALSKI, R. S. (1983). A theory and methodology of inductive learning. *Artificial Intelligence*, 20(2):111–162. */164/*

MICHALSKI, R. S., CARBONELL, J. G., MITCHELL, T. M. (eds.) (1983). *Machine Learning: An Artificial Intelligence Approach, Vol. I.* Tioga, Palo Alto, CA. */3/*

MICHALSKI, R. S., CARBONELL, J. G., MITCHELL, T. M. (eds.) (1986a). *Machine Learning: An Artificial Intelligence Approach, Vol. II.* Morgan Kaufmann, Los Altos, CA. */3/*

MICHALSKI, R. S., LARSON, J. B. (1978). *Selection of most representative training examples and incremental generation of VL1 hypotheses: the underlying methodology and the description of programs ESEL and AQ11.* Tech. Rep. 78-867, Department of Computer Science, University of Illinois at Urbana-Champaign. */49/*

MICHALSKI, R. S., MOZETIČ, I., HONG, J., LAVRAČ, N. (1986b). The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. In *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI-86)*, pp. 1041–1045. Philadelphia, PA. */1, 3, 49, 68, 118, 200/*

MICHALSKI, R. S., STEPP, R. E. (1983). Learning from observation: Conceptual clustering. In MICHALSKI, R., CARBONELL, J., MITCHELL, T. (eds.) *Machine Learning: An Artificial Intelligence Approach.* Tioga, Palo Alto, CA. */245/*

MICHIE, D., MUGGLETON, S. H., PAGE, D., SRINIVASAN, A. (1994a). *To the International Computing Community: A New East-West Challenge.* Tech. rep., Oxford University Computing laboratory, Oxford, UK. */98/*

MICHIE, D., SPIEGELHALTER, D., TAYLOR, C. C. (eds.) (1994b). *Machine Learning, Neural and Statistical Classification.* Ellis Horwood. */1/*

MIERSWA, I., WURST, M., KLINKENBERG, R., SCHOLZ, M., EULER, T. (2006). Yale: Rapid prototyping for complex data mining tasks. In UNGAR, L., CRAVEN, M., GUNOPULOS, D., ELIASSI-RAD, T. (eds.) *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 935–940. ACM, New York, NY, USA. */5/*

MINGERS, J. (1989a). An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, 4:227–243. */200/*

MINGERS, J. (1989b). An empirical comparison of selection measures for decision-tree induction. *Machine Learning*, 3:319–342. */135/*

MITCHELL, T. M. (1982). Generalization as search. *Artificial Intelligence*, 18(2):203–226. */29, 114/*

MITCHELL, T. M. (1997). *Machine Learning.* McGraw Hill. */1, 30/*

MLADENIĆ, D. (1993). Combinatorial optimization in inductive concept learning. In *Proceedings of the 10th International Conference on Machine Learning (ML-93)*, pp. 205–211. Morgan Kaufmann. */117, 122, 128/*

MOONEY, R. J. (1995). Encouraging experimental results on learning CNF. *Machine Learning*, 19:79–92. */51, 173/*

Mooney, R. J., Califf, M. E. (1995). Induction of first-order decision lists: Results on learning the past tense of english verbs. *Journal of Artificial Intelligence Research*, 3:1–24. */226/*

Morik, K., Boulicaut, J.-F., Siebes, A. (eds.) (2005). *Local Pattern Detection*. Springer-Verlag. */184/*

Muggleton, S. H. (1987). Structuring knowledge by asking questions. In Bratko, I., Lavrač, N. (eds.) *Progress in Machine Learning*, pp. 218–229. Sigma Press, Wilmslow, England. */127/*

Muggleton, S. H. (1988). A strategy for constructing new predicates in first order logic. In *Proceedings of the 3rd European Working Session on Learning (EWSL-88)*, pp. 123–130. */127/*

Muggleton, S. H. (1991). Inverting the resolution principle. In Hayes, J. E., Michie, D., Tyugu, E. (eds.) *Machine Intelligence 12*, chap. 7, pp. 93–103. Clarendon Press, Oxford. */127/*

Muggleton, S. H. (ed.) (1992). *Inductive Logic Programming*. Academic Press Ltd., London. */1, 7, 96/*

Muggleton, S. H. (1995). Inverse entailment and Progol. *New Generation Computing*, 13(3,4):245–286. Special Issue on Inductive Logic Programming. */52, 53, 120, 142/*

Muggleton, S. H., Buntine, W. L. (1988). Machine invention of first-order predicates by inverting resolution. In *Proceedings of the 5th International Conference on Machine Learning (ML-88)*, pp. 339–352. */127/*

Muggleton, S. H., Feng, C. (1990). Efficient induction of logic programs. In *Proceedings of the 1st Conference on Algorithmic Learning Theory*, pp. 1–14. Tokyo, Japan. */110, 131, 132, 133/*

Muggleton, S. H., Firth, J. (2001). Relational rule induction with CProgol4.4: A tutorial introduction. In Džeroski, S., Lavrač, N. (eds.) *Relational Data Mining*, chap. 7, pp. 160–188. Springer-Verlag, Berlin. */53/*

Muggleton, S. H., Santos, J. C. A., Tamaddoni-Nezhad, A. (2009). ProGolem: A system based on relative minimal generalisation. In De Raedt, L. (ed.) *Proceedings of the 19th International Conference on Inductive Logic Programming (ILP-09)*, pp. 131–148. Springer, Leuven, Belgium. */133/*

Mutter, S., Hall, M., Frank, E. (2004). Using classification to evaluate the output of confidence-based association rule mining. In Webb, G. I., Yu, X. (eds.) *Proceedings of the Australian Joint Conference on Artificial Intelligence (AI-04)*, pp. 538–549. Springer-Verlag, Cairns, Australia. */55, 185/*

Nemenyi, P. (1963). *Distribution-free multiple comparisons*. Ph.D. thesis, Princeton University. */91/*

Niblett, T., Bratko, I. (1987). Learning decision rules in noisy domains. In Bramer, M. A. (ed.) *Research and Development in Expert Systems III*, pp. 25–34. Cambridge University Press, Brighton, U.K. */150, 200/*

Niculescu-Mizil, A., Caruana, R. (2005a). Obtaining calibrated probabilities from boosting. In *Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence (UAI-05)*, p. 413. AUAI Press, Edinburgh, Scotland. */215/*

Niculescu-Mizil, A., Caruana, R. (2005b). Predicting good probabilities with supervised learning. In Raedt, L. D., Wrobel, S. (eds.) *Proceedings of the 22nd International Conference on Machine Learning (ICML 2005)*, pp. 625–632. ACM, Bonn, Germany. */215/*

Pagallo, G., Haussler, D. (1990). Boolean feature discovery in empirical learning. *Machine Learning*, 5:71–99. */12, 40, 145, 199, 204/*

Park, S.-H., Fürnkranz, J. (2007). Efficient pairwise classification. In Kok, J. N., Koronacki, J., López de Mántaras, R., Matwin, S., Mladenić, D., Skowron, A. (eds.) *Proceedings of 18th European Conference on Machine Learning (ECML-07)*, pp. 658–665. Springer-Verlag, Warsaw, Poland. */232/*

Park, S.-H., Fürnkranz, J. (2009). Efficient decoding of ternary error-correcting output codes for multiclass classification. In Buntine, W. L., Grobelnik, M., Mladenić, D., Shawe-Taylor, J. (eds.) *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD-09)*, vol. Part II, pp. 189–204. Springer-Verlag, Bled, Slovenia. */217, 236/*

Pazzani, M., Merz, C. J., Murphy, P., Ali, K., Hume, T., Brunk, C. (1994). Reducing misclassification costs. In Cohen, W. W., Hirsh, H. (eds.) *Proceedings of the 11th International Conference on Machine Learning (ML-94)*, pp. 217–225. Morgan Kaufmann, New Brunswick, NJ. */221/*

Peña Castillo, L., Wrobel, S. (2004). A comparative study on methods for reducing myopia of hill-climbing search in multirelational learning. In Brodley, C. E. (ed.) *Proceedings of the 21st International Conference on Machine Learning (ICML-2004)*. ACM Press, Banff, Alberta, Canada. */112, 117/*

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann. */1/*

Pechter, R. (2009). What's PMML and what's new in PMML 4.0. *SIGKDD explorations*, 11:19–25. */5/*

Pelleg, D., Moore, A. (2001). Mixtures of rectangles: Interpretable soft clustering. In Brodley, C. E., Danyluk, A. P. (eds.) *Proceedings of the 18th International Conference on Machine Learning (ICML-01)*, pp. 401–408. Morgan Kaufmann, Williamstown, MA. */245/*

Pfahringer, B. (1995a). A new MDL measure for robust rule induction (extended abstract). In Lavrač, N., Wrobel, S. (eds.) *Proceedings of the 8th European Conference on Machine Learning (ECML-95)*, pp. 331–334. Springer-Verlag, Heraclion, Greece. */163, 192, 206/*

Pfahringer, B. (1995b). *Practical Uses of the Minimum Description Length Principle in Inductive Learning*. Ph.D. thesis, Technische Universität Wien. */163/*

Pfahringer, B., Holmes, G., Wang, C. (2005). Millions of random rules. In Fürnkranz, J. (ed.) *Proceedings of the ECML/PKDD Workshop on Advances in Inductive Rule Learning*. Pisa, Italy. */183/*

Piatetsky-Shapiro, G. (1991). Discovery, analysis, and presentation of strong rules. In Piatetsky-Shapiro, G., Frawley, W. J. (eds.) *Knowledge Discovery in Databases*, pp. 229–248. MIT Press. */143, 155, 157, 166, 167/*

Piatetsky-Shapiro, G., Frawley, W. J. (eds.) (1991). *Knowledge Discovery in Databases*. MIT Press. */2, 3/*

Pietraszek, T. (2007). On the use of ROC analysis for the optimization of abstaining classifiers. *Machine Learning*, 68(2):137–169. */223/*

Pimenta, E., Gama, J., de Leon Ferreira de Carvalho, A. C. P. (2008). The dimension of ECOCs for multiclass classification problems. *International Journal on Artificial Intelligence Tools*, 17(3):433–447. */235/*

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In Smola, A., Bartlett, P., Schölkopf, B., Schuurmans, D. (eds.) *Advances in Large Margin Classifiers*. MIT Press. */215/*

Platt, J. C., Cristianini, N., Shawe-Taylor, J. (2000). Large margin DAGs for multiclass classification. In Solla, S. A., Leen, T. K., Müller, K.-R. (eds.) *Advances in Neural Information Processing Systems 12 (NIPS-99)*, pp. 547–553. MIT Press. */243/*

Plotkin, G. D. (1970). A note on inductive generalisation. In Meltzer, B., Michie, D. (eds.) *Machine Intelligence 5*, pp. 153–163. Elsevier North-Holland, New York. */132/*

Plotkin, G. D. (1971). A further note on inductive generalisation. In Meltzer, B., Michie, D. (eds.) *Machine Intelligence 6*, pp. 101–124. Elsevier North-Holland, New York. */133/*

Pompe, U., Kovačič, M., Kononenko, I. (1993). SFOIL: Stochastic approach to inductive logic programming. In *Proceedings of the 2nd Slovenian Conference on Electrical Engineering and Computer Science (ERK-93)*, vol. B, pp. 189–192. Portorož, Slovenia. */123, 191/*

Prati, R. C., Flach, P. A. (2005). Roccer: An algorithm for rule learning based on ROC analysis. In Kaelbling, L. P., Saffiotti, A. (eds.) *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, pp. 823–828. Professional Book Center, Edinburgh, Scotland. */242/*

Price, D., Knerr, S., Personnaz, L., Dreyfus, G. (1995). Pairwise neural network classifiers with probabilistic outputs. In Tesauro, G., Touretzky, D., Leen, T. (eds.) *Advances in Neural Information Processing Systems 7 (NIPS-94)*, pp. 1109–1116. MIT Press. */230/*

Provost, F., Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, 42(3):203–231. */58, 59/*

Provost, F. J., Domingos, P. (2003). Tree induction for probability-based ranking. *Machine Learning*, 52(3):199–215. */215/*

Pujol, O., Radeva, P., Vitriá, J. (2006). Discriminant ECOC: A heuristic method for application dependent design of error correcting output codes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):1007–1012. */235/*

Quevedo, J. R., Montañés, E., Luaces, O., del Coz, J. J. (2010). Adapting decision DAGs for multipartite ranking. In Balcázar, J. L., Bonchi, F., Gionis, A., Sebag, M. (eds.) *Proceedings of the Euorpean Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD-10), Part III*, pp. 115–130. Springer, Barcelona, Spain. */243/*

Quinlan, J. R. (1979). Discovering rules by induction from large collections of examples. In Michie, D. (ed.) *Expert Systems in the Micro Electronic Age*, pp. 168–201. Edinburgh University Press. */3/*

Quinlan, J. R. (1983). Learning efficient classification procedures and their application to chess end games. In Michalski, R. S., Carbonell, J. G., Mitchell, T. M. (eds.) *Machine Learning. An Artificial Intelligence Approach*, pp. 463–482. Tioga, Palo Alto, CA. */50, 146, 170/*

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1:81–106. */1, 3, 7, 9, 67, 146/*

Quinlan, J. R. (1987a). Generating production rules from decision trees. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI-87)*, pp. 304–307. Morgan Kaufmann. */12, 55, 182, 211, 221/*

Quinlan, J. R. (1987b). Simplifying decision trees. *International Journal of Man-Machine Studies*, 27:221–234. */199, 200/*

Quinlan, J. R. (1990). Learning logical definitions from relations. *Machine Learning*, 5:239–266. */7, 51, 67, 96, 97, 120, 129, 160, 170, 191, 192/*

Quinlan, J. R. (1991). Determinate literals in inductive logic programming. In *Proceedings of the 8th International Workshop on Machine Learning (ML-91)*, pp. 442–446. */51, 110/*

Quinlan, J. R. (1992). Learning with continuous classes. In Adams, N., Sterling, L. (eds.) *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, pp. 343–348. World Scientific, Hobart, Tasmania. */244/*

QUINLAN, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA. */V, 12, 52, 55, 182, 206, 221/*

QUINLAN, J. R. (1994). The minimum description length principle and categorical theories. In COHEN, W., HIRSH, H. (eds.) *Proceedings of the 11th International Conference on Machine Learning (ML-94)*, pp. 233–241. Morgan Kaufmann, New Brunswick, NJ. */163, 206/*

QUINLAN, J. R. (1995). MDL and categorical theories (continued). In PRIEDITIS, A., RUSSELL, S. J. (eds.) *Proceedings of the 12th International Conference on Machine Learning (ICML-95)*, pp. 464–470. Morgan Kaufmann, Tahoe City, CA. */206, 211/*

QUINLAN, J. R., CAMERON-JONES, R. M. (1995a). Induction of logic programs: FOIL and related systems. *New Generation Computing*, 13(3,4):287–312. Special Issue on Inductive Logic Programming. */51, 111/*

QUINLAN, J. R., CAMERON-JONES, R. M. (1995b). Oversearching and layered search in empirical learning. In MELLISH, C. (ed.) *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, pp. 1019–1024. Morgan Kaufmann. */119/*

RAMAKRISHNAN, G., JOSHI, S., BALAKRISHNAN, S., SRINIVASAN, A. (2008). Using ILP to construct features for information extraction from semi-structured text. In BLOCKEEL, H., RAMON, J., SHAVLIK, J. W., TADEPALLI, P. (eds.) *Proceedings of the 17th International Conference on Inductive Logic Programming (ILP-07)*, pp. 211–224. Springer, Corvallis, OR. */53/*

RAMASWAMY, S., TAMAYO, P., RIFKIN, R., MUKHERJEE, S., YEANG, C.-H., ANGELO, M., LADD, C., REICH, M., LATULIPPE, E., MESIROV, J. P., POGGIO, T., GERALD, W., LODA, M., LANDER, E. S., GOLUB, T. R. (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences*, 98(26):15149–15154. */274, 281/*

RIJNBEEK, P. R., KORS, J. A. (2010). Finding a short and accurate decision rule in disjunctive normal form by exhaustive search. *Machine Learning*, 80(1):33–62. */183/*

RIPLEY, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press. */1/*

RISSANEN, J. (1978). Modeling by shortest data description. *Automatica*, 14:465–471. */163, 192, 221/*

RIVEST, R. L. (1987). Learning decision lists. *Machine Learning*, 2:229–246. */12, 50, 116/*

ROUVEIROL, C. (1992). Extensions of inversion of resolution applied to theory completion. In MUGGLETON, S. H. (ed.) *Inductive Logic Programming*, pp. 63–92. Academic Press Ltd., London. */127, 133/*

ROUVEIROL, C. (1994). Flattening and saturation: Two representation changes for generalization. *Machine Learning*, 14:219–232. Special issue on Evaluating and Changing Representation. */104, 133/*

ROUVEIROL, C., PUGET, J. F. (1990). Beyond inversion of resolution. In *Proceedings of the 7th International Conference on Machine Learning (ML-90)*, pp. 122–130. */127/*

RÜCKERT, U., DE RAEDT, L. (2008). An experimental evaluation of simplicity in rule learning. *Artificial Intelligence*, 172(1):19–28. */183/*

RÜCKERT, U., KRAMER, S. (2003). Stochastic local search in k-term DNF learning. In FAWCETT, T., MISHRA, N. (eds.) *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 648–655. AAAI Press, Washington, DC. */183/*

RÜCKERT, U., KRAMER, S. (2008). Margin-based first-order rule learning. *Machine Learning*, 70(2-3):189–206. */215/*

RUMELHART, D. E., MCCLELLAND, J. L. (eds.) (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1: Foundations. MIT Press, Cambridge, MA. */1, 3/*

RÜPING, S. (2006). Robust probabilistic calibration. In FÜRNKRANZ, J., SCHEFFER, T., SPILIOPOULOU, M. (eds.) *Proceedings of the 17th European Conference on Machine Learning (ECML/PKDD-06)*, pp. 743–750. Springer. */215/*

SALZBERG, S. (1991). A nearest hyperrectangle learning method. *Machine Learning*, 6:251–276. */181, 223/*

SCHAPIRE, R. E., FREUND, Y., BARTLETT, P., LEE, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics*, 26(5):1651–1686. */2/*

SCHEFFER, T., WROBEL, S. (2002). Finding the most interesting patterns in a database quickly by using sequential sampling. *Journal of Machine Learning Research*, 3:833–862. */186/*

SCHMIDT, M. S., GISH, H. (1996). Speaker identification via support vector classifiers. In *Proceedings of the 21st IEEE International Conference Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)*, pp. 105–108. Atlanta, GA. */230/*

SCHÖLKOPF, B., SMOLA, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press. */1/*

SHAPIRO, E. Y. (1981). An algorithm that infers theories from facts. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI-81)*, pp. 446–451. */110/*

SHAPIRO, E. Y. (1982). *Algorithmic Program Debugging*. The MIT Press, Cambridge. */97/*

SHAPIRO, E. Y. (1991). Inductive inference of theories from facts. In LASSEZ, J. L., PLOTKIN, G. D. (eds.) *Computational Logic: Essays in Honor of Alan Robinsons*, pp. 199–255. MIT Press. */97/*

SILBERSCHATZ, A., TUZHILIN, A. (1995). On subjective measure of interestingness in knowledge discovery. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, pp. 275–281. */288/*

SIMEON, M., HILDERMAN, R. J. (2007). Exploratory quantitative contrast set mining: A discretization approach. In *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence - Vol.2 (ICTAI-07)*, pp. 124–131. */252/*

SIU, K., BUTLER, S., BEVERIDGE, T., GILLAM, J., HALL, C., KAYE, A., LEWIS, R., MANNAN, K., MCLOUGHLIN, G., PEARSON, S., ROUND, A., SCHULTKE, E., WEBB, G., WILKINSON, S. (2005). Identifying markers of pathology in SAXS data of malignant tissues of the brain. *Nuclear Instruments and Methods in Physics Research A*, 548:140–146. */253/*

SMOLA, A. J., SCHÖLKOPF, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14:199–222. */245/*

SMYTH, P., GOODMAN, R. M. (1991). Rule induction using information theory. In PIATETSKY-SHAPIRO, G., FRAWLEY, W. J. (eds.) *Knowledge Discovery in Databases*, pp. 159–176. MIT Press. */153, 183/*

SOARES, C. (2003). Is the UCI repository useful for data mining? In MOURA-PIRES, F., ABREU, S. (eds.) *Proceedings of the 11th Protuguese Conference on Artificial Intelligence (EPIA-03)*, pp. 209–223. Springer. */48/*

SONG, H. S., KIMB, J. K., KIMA, S. H. (2001). Mining the change of customer behavior in an internet shopping mall. *Expert Systems with Applications*, 21(3):157–168. */254/*

Soulet, A., Crmilleux, B., Rioult, F. (2004). Condensed representation of emerging patterns. In *Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-04)*, pp. 127–132. */254/*

Specia, L., Srinivasan, A., Joshi, S., Ramakrishnan, G., das Graças Volpe Nunes, M. (2009). An investigation into feature construction to assist word sense disambiguation. *Machine Learning*, 76(1):109–136. */53/*

Srinivasan, A. (1999). The Aleph manual. `http://web.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/`. */53/*

Srinivasan, A., King, R. D. (1997). Feature construction with inductive logic programming: A study of quantitative predictions of biological activity by structural attributes. In Muggleton, S. (ed.) *Proceedings of the 6th International Workshop, on Inductive Logic Programming (ILP-96)*, pp. 89–104. Springer, Stockholm, Sweden. */54/*

Stepp, R. E., Michalski, R. S. (1986). Conceptual clustering of structured objects: A goal-oriented approach. *Artificial Intelligence*, 28(1):43–69. */245/*

Sterling, L., Shapiro, E. (1994). *The Art of Prolog — Advanced Programming Techniques*. The MIT Press, 2nd edn. */12, 103/*

Sternberg, M. J., Muggleton, S. H. (2003). Structure activity relationships (SAR) and pharmacophore discovery using inductive logic programming (ILP). *QSAR and Combinatorial Science*, 22(5):527–532. */53/*

Sulzmann, J.-N., Fürnkranz, J. (2008). A comparison of techniques for selecting and combining class association rules. In Fürnkranz, J., Knobbe, A. J. (eds.) *From Local Patterns to Global Models: Proceedings of the ECML/PKDD-08 Workshop (LeGo-08)*, pp. 154–168. Antwerp, Belgium. */185/*

Sulzmann, J.-N., Fürnkranz, J. (2009). An empirical comparison of probability estimation techniques for probabilistic rules. In Gama, J., Santos Costa, V., Jorge, A., Brazdil, P. B. (eds.) *Proceedings of the 12th International Conference on Discovery Science (DS-09)*, pp. 317–331. Springer-Verlag. Winner of Carl Smith Award for Best Student Paper. */215/*

Sulzmann, J.-N., Fürnkranz, J. (2011). Rule stacking: An approach for compressing an ensemble of rule sets into a single classifier. In Elomaa, T., Hollmèn, J., Mannila, H. (eds.) *Proceedings of the 14th International Conference on Discovery Science (DS-11)*. */232/*

Suzuki, E. (2006). Data mining methods for discovering interesting exceptions from an unsupervised table. *Journal of Universal Computer Science*, 12(6):627–653. */261/*

Tan, P.-N., Kumar, V., Srivastava, J. (2002). Selecting the right interestingness measure for association patterns. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-02)*, pp. 32–41. Edmonton, Alberta. */136, 168/*

Tan, P.-N., Kumar, V., Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems*, 29(4):293–313. */136/*

Theron, H., Cloete, I. (1996). BEXA: A covering algorithm for learning propositional concept descriptions. *Machine Learning*, 24:5–40. */50, 68, 118, 150, 195, 211/*

Todorovski, L., Flach, P., Lavrač, N. (2000). Predictive performance of weighted relative accuracy. In Zighed, D., Komorowski, J., Zytkow, J. (eds.) *Proceedings of the 4th European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD-2000)*, pp. 255–264. Springer-Verlag, Lyon, France. */143, 165, 198/*

Torgo, L. (1995). Data fitting with rule-based regression. In Zizka, J., Brazdil, P. B. (eds.) *Proceedings of the 2nd International Workshop on Artificial Intelligence Techniques (AIT-95)*. Springer-Verlag, Brno, Czech Republic. */244/*

TORGO, L. (2010). *Data Mining with R: Learning with Case Studies*. Data Mining and Knowledge Discovery Series, Chapman & Hall / CRC. /5/

TORGO, L., GAMA, J. (1997). Regression using classification algorithms. *Intelligent Data Analysis*, 1(4). /245/

ULLMAN, J. D. (1988). *Principles of Database and Knowledge Base Systems*, vol. I. Computer Science Press, Rockville, MA. /107, 108, 109/

VAN HORN, K. S., MARTINEZ, T. R. (1993). The BBG rule induction algorithm. In *Proceedings of the 6th Australian Joint Conference on Artificial Intelligence (AI-93)*, pp. 348–355. Melbourne, Australia. /226/

VAN RIJSBERGEN, C. J. (1979). *Information Retrieval*. Butterworths, London, UK, 2nd edn. /148/

VAPNIK, V. (1995). *The Nature of Statististical Learning Theory*. Springer-Verlag, New York. /1/

VENTURINI, G. (1993). SIA: A supervised inductive algorithm with genetic search for learning attributes based concepts. In BRAZDIL, P. (ed.) *Proceedings of the 6th European Conference on Machine Learning (ECML-93)*, pp. 280–296. Springer-Verlag, Vienna, Austria. /123, 127, 163, 164/

WALLACE, C. S., BOULTON, D. M. (1968). An information measure for classification. *Computer Journal*, 11:185–194. /163, 192/

WANG, B., ZHANG, H. (2006). Improving the ranking performance of decision trees. In FÜRNKRANZ, J., SCHEFFER, T., SPILIOPOULOU, M. (eds.) *Proceedings of the 17th European Conference on Machine Learning (ECML-06)*, pp. 461–472. Springer-Verlag, Berlin, Germany. /215/

WANG, K., ZHOU, S., FU, A. W.-C., YU, J. X. (2003). Mining changes of classification by correspondence tracing. In *Proceedings of the 3rd SIAM International Conference on Data Mining (SDM-03)*, pp. 95–106. /260/

WATANABE, L., RENDELL, L. (1991). Learning structural decision trees from examples. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, pp. 770–776. /13/

WEBB, G. I. (1992). *Learning Disjunctive Class Descriptions by Least Generalisation*. Tech. Rep. TR C92/9, Deakin University, School of Computing & Mathematics, Geelong, Australia. /127, 140/

WEBB, G. I. (1993). Systematic search for categorical attribute-value data-driven machine learning. In ROWLES, C., LIU, H., FOO, N. (eds.) *Proceedings of the 6th Australian Joint Conference of Artificial Intelligence (AI'93)*, pp. 342–347. World Scientific, Melbourne. /119, 150/

WEBB, G. I. (1994). Recent progress in learning decision lists by prepending inferred rules. In *Proceedings of the 2nd Singapore International Conference on Intelligent Systems*, pp. B280–B285. /226/

WEBB, G. I. (1995). OPUS: An efficient admissible algorithm for unordered search. *Journal of Artificial Intelligence Research*, 5:431–465. /15, 54, 120, 252/

WEBB, G. I. (1996). Further experimental evidence aganst the utility of Occam's razor. *Journal of Artificial Intelligence Research*, 4:397–417. /9/

WEBB, G. I. (2000). Efficient search for association rules. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2000)*, pp. 99–107. Boston, MA. /54, 122/

WEBB, G. I. (2001). Discovering associations with numeric variables. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-01)*, pp. 383–388. /261/

WEBB, G. I. (2007). Discovering significant patterns. *Machine Learning*, 68(1):1–33. /251, 252, 259/

WEBB, G. I., BRKIČ, N. (1993). Learning decision lists by prepending inferred rules. In *Proceedings of the AI'93 Workshop on Machine Learning and Hybrid Systems*. Melbourne, Australia. */226/*

WEBB, G. I., BUTLER, S. M., NEWLANDS, D. (2003). On detecting differences between groups. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-03)*, pp. 256–265. */251, 253, 259/*

WEBB, G. I., ZHANG, S. (2005). k-optimal rule discovery. *Data Mining and Knowledge Discovery*, 10(1):39–79. */54/*

WEISS, S. M., INDURKHYA, N. (1991). Reduced complexity rule induction. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, pp. 678–684. */128, 145, 202/*

WEISS, S. M., INDURKHYA, N. (1994). Small sample decision tree pruning. In *Proceedings of the 11th Conference on Machine Learning*, pp. 335–342. Rutgers University, New Brunswick, NJ. */212/*

WEISS, S. M., INDURKHYA, N. (1995). Rule-based machine learning methods for functional prediction. *Journal of Artificial Intelligence Research*, 3:383–403. */245/*

WEISS, S. M., INDURKHYA, N. (2000). Lightweight rule induction. In LANGLEY, P. (ed.) *Proceedings of the 17th International Conference on Machine Learning (ICML-2000)*, pp. 1135–1142. Stanford, CA. */178, 186/*

WETTSCHERECK, D. (2002). A KDDSE-independent PMML visualizer. In *Proceedings of 2nd Workshop on Integration Aspects of Data Mining, Decision Support and Meta-Learning (IDDM-02)*, pp. 150–155. */261/*

WIDMER, G. (1993). Combining knowledge-based and instance-based learning to exploit qualitative knowledge. *Informatica*, 17:371–385. Special Issue on Multistrategy Learning. */128/*

WIDMER, G. (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146(2):129–148. */182/*

WIESE, M. (1996). A bidirectional ILP algorithm. In *Proceedings of the MLnet Familiarization Workshop on Data Mining with Inductive Logic Programming (ILP for KDD)*, pp. 61–72. */128/*

WINDEATT, T., GHADERI, R. (2003). Coding and decoding strategies for multi-class learning problems. *Information Fusion*, 4(1):11–21. */235/*

WIRTH, R. (1988). Learning by failure to prove. In *Proceedings of the Third European Working Session on Learning*, pp. 237–251. Pitman, Glasgow, Scotland. */127/*

WITTEN, I. H., FRANK, E. (2005). *Data Mining — Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers, 2nd edn. */1, 2, 5, 10, 23, 52/*

WOHLRAB, L., FÜRNKRANZ, J. (2011). A review and comparison of strategies for handling missing values in separate-and-conquer rule learning. *Journal of Intelligent Information Systems*, 36(1):73–98. */88, 90, 91/*

WOLPERT, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2):241–260. */232/*

WONG, T.-T., TSENG, K.-L. (2005). Mining negative contrast sets from data with discrete attributes. *Expert Systems with Applications*, 29(2):401–407. */251, 253/*

WROBEL, S. (1996). First order theory refinement. In DE RAEDT, L. (ed.) *Advances in Inductive Logic Programming*, pp. 14–33. IOS Press, Amsterdam, Netherlands. */183/*

WROBEL, S. (1997). An algorithm for multi-relational discovery of subgroups. In *Proceedings of the 1st European Symposium on Principles of Data Min-*

ing and Knowledge Discovery (PKDD-97), pp. 78–87. Springer-Verlag, Berlin. */157, 158, 247, 249, 256, 258/*

WROBEL, S. (2001). Inductive logic programming for knowledge discovery in databases. In DŽEROSKI, S., LAVRAČ, N. (eds.) *Relational Data Mining*, pp. 74–101. Springer. */96, 249, 261, 262/*

WU, T., CHEN, Y., HAN, J. (2007). Association mining in large databases: A re-examination of its measures. In *Proceedings of the 11th European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD-07)*, pp. 621–628. Springer-Verlag, Warsawa, Poland. */168/*

WU, T.-F., LIN, C.-J., WENG, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5(Aug):975–1005. */239/*

XIONG, H., SHEKHAR, S., TAN, P.-N., , KUMAR, V. (2004). Exploiting a support-based upper bound of pearson's correlation coefficient for efficiently identifying strongly correlated pairs. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-04)*, pp. 334–343. Seattle, USA. */155/*

YANG, Y., WEBB, G. I., WU, X. (2005). Discretization methods. In MAIMON, O., ROKACH, L. (eds.) *The Data Mining and Knowledge Discovery Handbook*, pp. 113–130. Springer-Verlag. */75/*

YIN, X., HAN, J. (2003). CPAR: Classification based on predictive association rules. In BARBARÁ, D., KAMATH, C. (eds.) *Proceedings of the SIAM Conference on Data Mining (SDM-03)*, pp. 331–335. */55, 185/*

ZADEH, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353. */93/*

ZADROZNY, B., ELKAN, C. (2001). Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. In BRODLEY, C. E., DANYLUK, A. P. (eds.) *Proceedings of the 18th International Conference on Machine Learning (ICML 2001)*, pp. 609–616. Morgan Kaufmann, Williams College, Williamstown, MA, USA. */215/*

ZADROZNY, B., ELKAN, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-02)*, pp. 694–699. ACM, Edmonton, Alberta, Canada. */215/*

ZAKI, M. J., PARTHASARATHY, S., OGIHARA, M., LI, W. (1997). New algorithms for fast discovery of association rules. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97)*, pp. 283–286. Newport, CA. */122/*

ZELEZNÝ, F., LAVRAČ, N. (2006). Propositionalization-based relational subgroup discovery with RSD. *Machine Learning*, 62:33–63. */102, 249/*

ZELLE, J. M., MOONEY, R. J., KONVISSER, J. B. (1994). Combining top-down and bottom-up techniques in inductive logic programming. In COHEN, W., HIRSH, H. (eds.) *Proceedings of the 11th International Conference on Machine Learning (ML-94)*, pp. 343–351. Morgan Kaufmann, New Brunswick, NJ. */182/*

ZENKO, B. (2007). *Learning Predictive Clustering Rules*. Ph.D. thesis, University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia. */246/*

ZENKO, B., DŽEROSKI, S., STRUYF, J. (2006). Learning predictive clustering rules. In BONCHI, F., BOULICAUT, J.-F. (eds.) *Proceedings of the 4th International Workshop on Knowledge Discovery in Inductive Databases (KDID-05)*, pp. 234–250. Springer, Porto, Portugal. */246/*

ZHANG, C., ZHANG, S. (2002). *Association Rule Mining: Models and Algorithms*. Springer-Verlag. */14, 122/*

Zhou, W., Xiong, H. (2011). Checkpoint evolution for volatile correlation computing. *Machine Learning*, 83(1):103–131. */155/*

Zimmermann, A., De Raedt, L. (2009). Cluster-grouping: From subgroup discovery to clustering. *Machine Learning*, 77(1):125–159. */246/*