Machine Learning

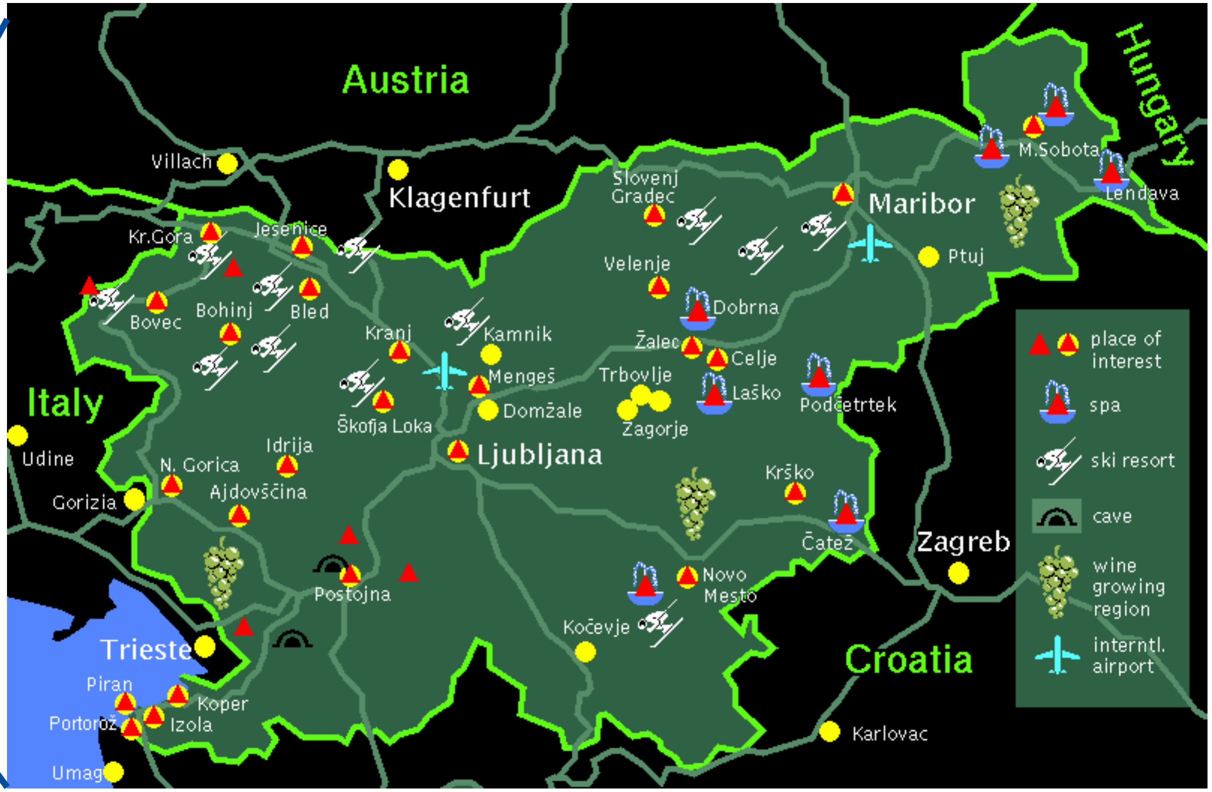
Part of programme
Applied Statistics with Network Analysis
HSE University, Moscow
January-February 2021

Nada Lavrač, Ljupčo Todorovski

Jožef Stefan Institute, University of Ljubljana
Ljubljana, Slovenia

Slovenia – Ljubljana (capital)

Europe



Ljubljana, Slovenia

Jožef Stefan Institute, Ljubljana, Slovenia

- **Jožef Stefan Institute (JSI, founded in 1949)**
  - named after a distinguished physicist Stefan (1835-1893)
  - leading national research organization $j = \sigma T^4$ s
    and technology (~700 researchers and students)

- **Jožef Stefan International Postgraduate School (founded in 2004)**
  - Offers four MSc and PhD programs (in English): ICT, nanotechnologies, ecotechnologies and sensor technologies

Department of Knowledge Technologies at JSI

Department of Knowledge Technologies

Knowledge Technologies
- Making AI techniques operational for practical problems

Staff
- 35 researchers, 10 students

Main research areas
- Machine Learning and data Mining
- Text Mining and Human Language Technologies
- Web Services and Semantic Web
- Ontologies and Knowledge Management
- Decision Support Systems

Applications
- Medicine, Bioinformatics, Public Health
- Ecology, Finance, …

Nada Lavrač

Research areas
- Machine Learning
- Text Mining
- Web Services
- Semantic Web

Applications
- Medicine, Bioinformatics
- Public Health
- Media News analysis

Teaching
- JSI Postgraduate School
- University of Nova Gorica
- University of Ljubljana
- HSE ☺

Ljupčo Todorovski

Research areas

– Machine learning
– Meta learning
– Symbolic regression (equation discovery)
– Time series and dynamical systems

Applications

– Bioinformatics
– Environmental sciences
– Public administration

Teaching

– University of Ljubljana
    - Faculty of Public Administration
    - Faculty of Mathematics and Physics
– HSE ☺

Machine Learning 2020/2021 Logistics: Course participants

Home page: https://github.com/ljupco-todorovski/hse-moscow-ml

ZOOM link: https://fmf-uni-lj-si.zoom.us/j/97756216461

Nada Lavrač: nada.lavrac@ijs.si, Ljupčo Todorovski: ljupco.todorovski@fu.uni-lj.si

HSE 1st year MSc students - masna2020group@gmail.com To be listed later – tour de table ZOOM presentation of individual students
HSE 2nd year MSc students - hsemasna@yandex.ru To be listed later – tour de table ZOOM presentation of individual students

HSE Course Schedule – 2020/21

Every Tuesday and Thursday 17:30 – 20:30 Moscow time, via ZOOM

5 x 3 hours on January 14, 19, 21, 26, 28

5 x 3 hours on February 2, 4, 9, 11, 16

Possible exceptions:

- to be communicated later

Machine Learning: Credits and Coursework

Credits:

- 4 ECTS ?

Requirements:

- Attending lectures
- Attending practical exercises: Jupyter Python notebooks, R

Exam requirements:

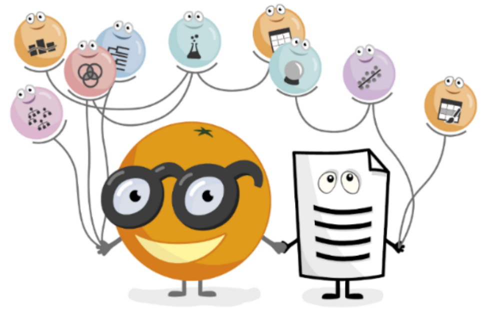
- To be communicated later

Machine Learning course: Supporting material

- Supporting material on videolectures.net:

  Seminar: AI for Industry and Society, Ljubljana 2020
  - http://videolectures.net/AIindustrySeminar2019/
  - Marko Robnik Šikonja: Artificial Intelligence: Techniques, Trends and Applications
  - Nada Lavrač: Data Science, Machine Learning and Big Data: Current trends
  - Blaž Zupan: Data Science with the OrangeToolbox

orange



- Open source machine learning and data visualization toolbox
  - https://orange.biolab.si/
  - http://file.biolab.si/datasets/
  - https://www.youtube.com/channel/UClKKWBe2SCAEyv7ZNGhIe4g
- Interactive data analysis workflows
- Visual programming
- Based on numpy, scipy and **scikit-learn**
- GUI: Qt framework

Machine Learning

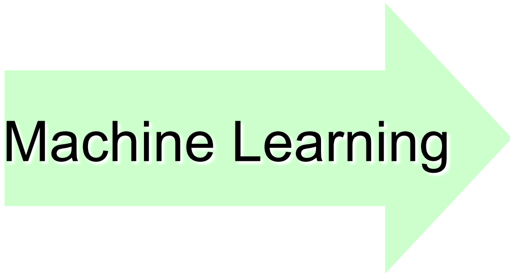
- Machine Learning (ML) – computer algorithms/machines that learn predictive models from class-labeled data
  - early rule learning algorithms: AQ (Michalski 1969), …
  - early decision tree learning algorithms since 1970s: ID3 (Quinlan 1979), …
  - early regression tree learners CART (Breiman et al. 1984), …
  - …

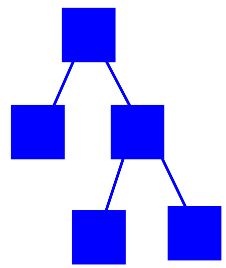
Machine Learning

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses
O1	17	myope	no	reduced	NONE
O2	23	myope	no	normal	SOFT
O3	22	myope	yes	reduced	NONE
O4	27	myope	yes	normal	HARD
O5	19	hypermetrope	no	reduced	NONE
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	SOFT
O15	43	hypermetrope	yes	reduced	NONE
O16	39	hypermetrope	yes	normal	NONE
O17	54	myope	no	reduced	NONE
O18	62	myope	no	normal	NONE
O19-O23	...	...	...	...	...
O24	56	hypermetrope	yes	normal	NONE

data

knowledge discovery
from data



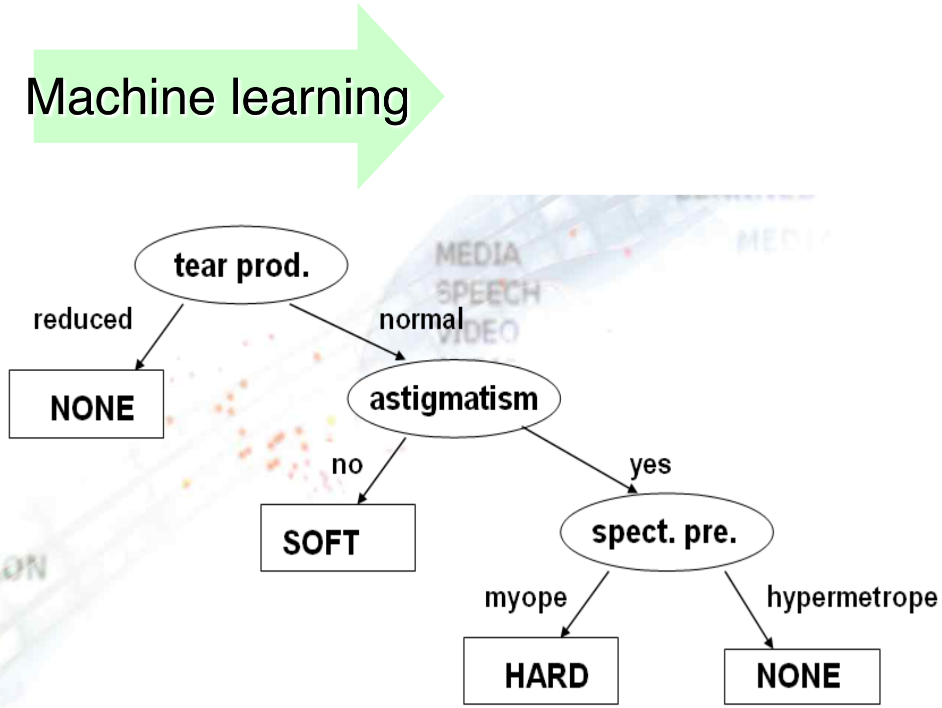


classification model

Given: class-labeled data (e.g., transaction data table, relational database, text documents, Web pages, …)
Find: a classification model, able to predict new instances

Machine learning: An illustrative example

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses
O1	17	myope	no	reduced	NONE
O2	23	myope	no	normal	SOFT
O3	22	myope	yes	reduced	NONE
O4	27	myope	yes	normal	HARD
O5	19	hypermetrope	no	reduced	NONE
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	SOFT
O15	43	hypermetrope	yes	reduced	NONE
O16	39	hypermetrope	yes	normal	NONE
O17	54	myope	no	reduced	NONE
O18	62	myope	no	normal	NONE
O19-O23	...	...	...	...	...
O24	56	hypermetrope	yes	normal	NONE



lenses=NONE ← tear production=reduced

lenses=NONE ← tear production=normal AND astigmatism=yes AND spect. pre.=hypermetrope

lenses=SOFT ← tear production=normal AND astigmatism=no

lenses=HARD ← tear production=normal AND astigmatism=yes AND spect. pre.=myope

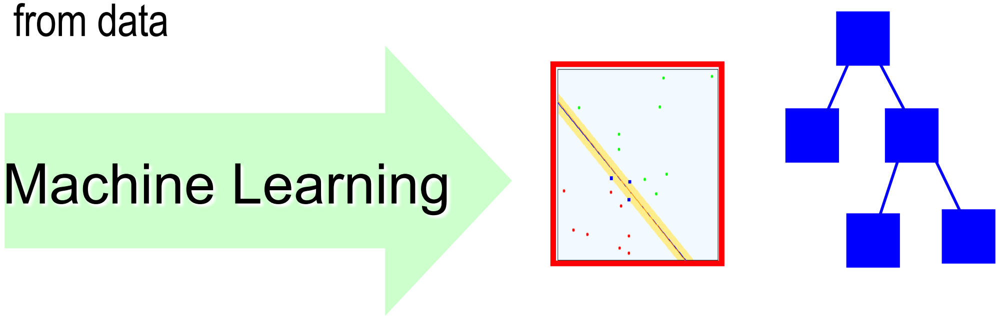
lenses=NONE ←

Machine Learning

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses
O1	17	myope	no	reduced	NONE
O2	23	myope	no	normal	SOFT
O3	22	myope	yes	reduced	NONE
O4	27	myope	yes	normal	HARD
O5	19	hypermetrope	no	reduced	NONE
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	SOFT
O15	43	hypermetrope	yes	reduced	NONE
O16	39	hypermetrope	yes	normal	NONE
O17	54	myope	no	reduced	NONE
O18	62	myope	no	normal	NONE
O19-O23	...	...	...	...	...
O24	56	hypermetrope	yes	normal	NONE

data

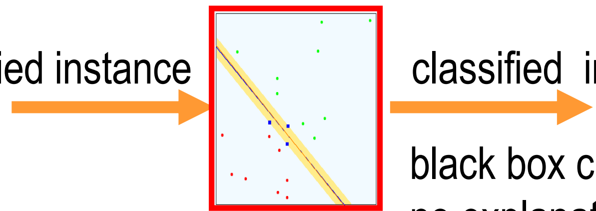
knowledge discovery
from data



classification models

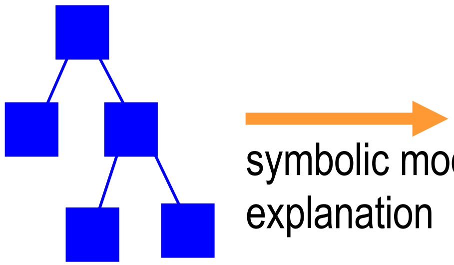
Given: class-labeled data (e.g., transaction data table, relational database, text documents, Web pages, …)
Find: a classification model, able to predict new instances

new unclassified instance

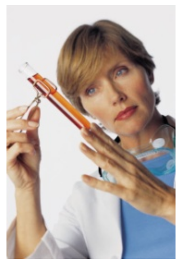


classified instance

black box classifier
no explanation



symbolic model
explanation

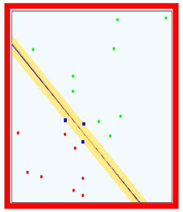


Why learn and use black-box models

Given: the learned classification model
(e.g., linear classifier, deep neural network, …)

Find: - the class label for a new unlabeled instance

new unclassified instance



classified instance

Advantages:

- best classification results in image recognition and other complex classification tasks

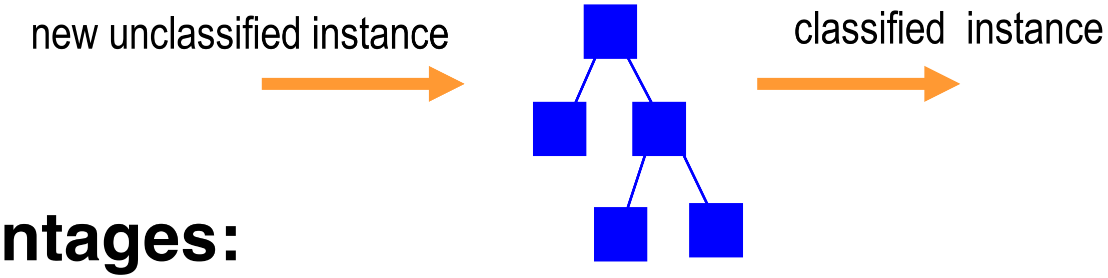
Drawbacks:

- poor interpretability of results
- can not be used for pattern analysis

Why learn and use symbolic models

Given: the learned classification model
(a decision tree or a set of rules)

Find: - the class label for a new unlabeled instance



Advantages:

- use the model for the explanation of classifications of new data instances
- use the discovered patterns for data exploration

Drawbacks:

- lower accuracy than deep NNs

First Generation Machine Learning

- **First machine learning algorithms for**
  – Decision tree and rule learning in 1970s and early 1980s by Quinlan, Michalski et al., Breiman et al., …
- **Characterized by**
  – Learning from data stored in a single data table
  – Relatively small set of instances and attributes
- **Lots of ML research followed in 1980s**
  – Numerous conferences ICML, ECML, … and ML sessions at AI conferences IJCAI, ECAI, AAAI, …
  – Extended set of learning tasks and algorithms addressed

Multi-class Learning Task

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses
O1	17	myope	no	reduced	NONE
O2	23	myope	no	normal	SOFT
O3	22	myope	yes	reduced	NONE
O4	27	myope	yes	normal	HARD
O5	19	hypermetrope	no	reduced	NONE
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	SOFT
O15	43	hypermetrope	yes	reduced	NONE
O16	39	hypermetrope	yes	normal	NONE
O17	54	myope	no	reduced	NONE
O18	62	myope	no	normal	NONE
O19-O23	...	...	no	...	...
O24	56	hypermetrope	no	normal	NONE

Several class labels of training examples of a single Target attribute

Binary Classification

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses
O1	17	myope	no	reduced	NO
O2	23	myope	no	normal	YES
O3	22	myope	yes	reduced	NO
O4	27	myope	yes	normal	YES
O5	19	hypermetrope	no	reduced	NO
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	YES
O15	43	hypermetrope	yes	reduced	NO
O16	39	hypermetrope	yes	normal	NO
O17	54	myope	no	reduced	NO
O18	62	myope	no	normal	NO
O19-O23	...	...	...	...	...
O24	56	hypermetrope	yes	normal	NO

Binary classes

- positive vs. negative examples of Target class
- Concept learning – binary classification and class description
  - for Subgroup discovery – exploring patterns characterizing groups of instances of target class

Multi-target Classification

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses	Pilot
O1	17	myope	no	reduced	NO	NO
O2	23	myope	no	normal	YES	NO
O3	22	myope	yes	reduced	NO	NO
O4	27	myope	yes	normal	YES	NO
O5	19	hypermetrope	no	reduced	NO	NO
O6-O13	...	...	...	...	...	...
O14	35	hypermetrope	no	normal	YES	YES
O15	43	hypermetrope	yes	reduced	NO	NO
O16	39	hypermetrope	yes	normal	NO	NO
O17	54	myope	no	reduced	NO	NO
O18	62	myope	no	normal	NO	YES
O19-O23	...	...	...	...	...	...
O24	56	hypermetrope	yes	normal	NO	NO

Multi target classification

- each example belongs to several Target classes

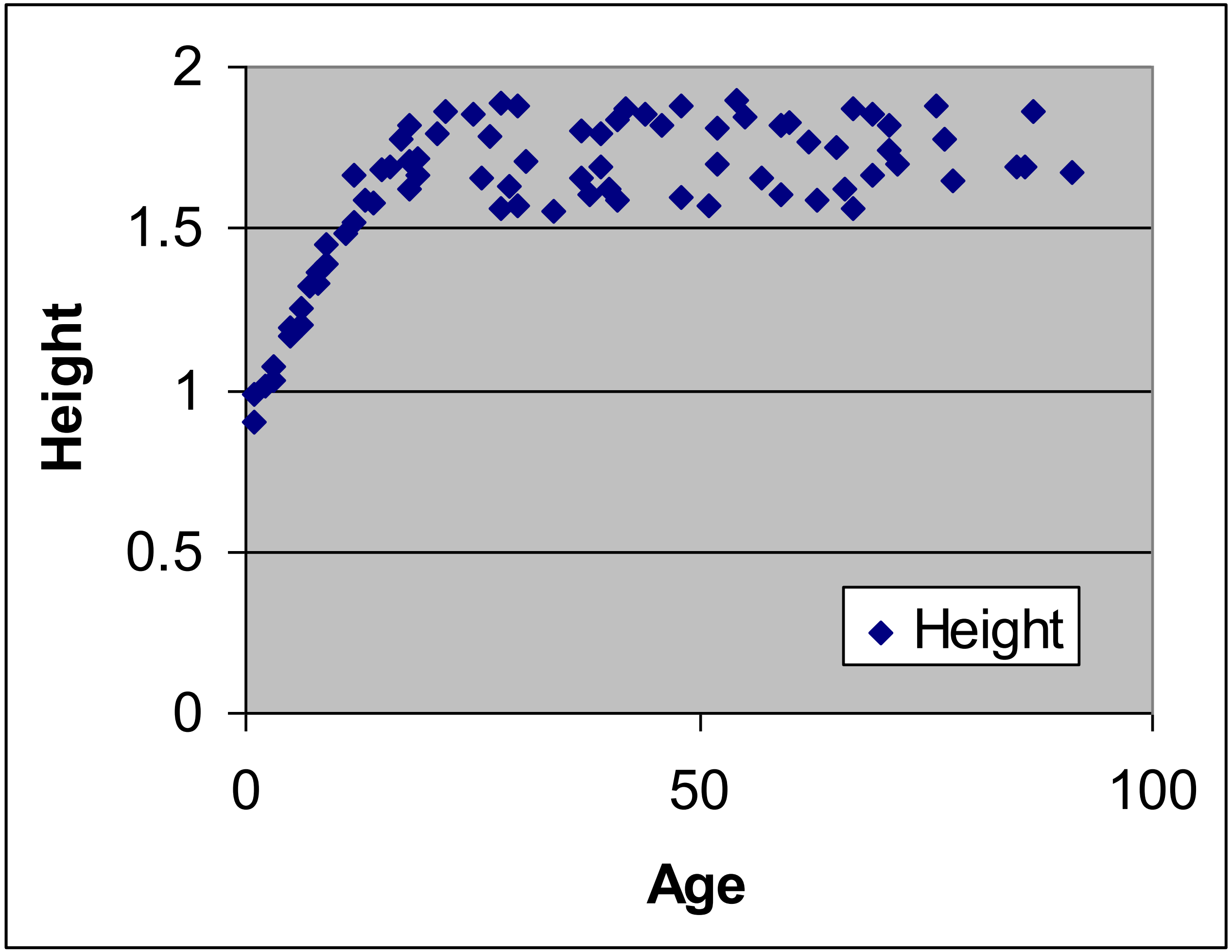
Learning from Numeric Class Data

Person	Age	Spect. presc.	Astigm.	Tear prod.	LensPrice
O1	17	myope	no	reduced	0
O2	23	myope	no	normal	8
O3	22	myope	yes	reduced	0
O4	27	myope	yes	normal	5
O5	19	hypermetrope	no	reduced	0
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	5
O15	43	hypermetrope	yes	reduced	0
O16	39	hypermetrope	yes	normal	0
O17	54	myope	no	reduced	0
O18	62	myope	no	normal	0
O19-O23	...	...	...	...	...
O24	56	hypermetrope	yes	normal	0

Numeric class values – regression analysis

Example regression problem

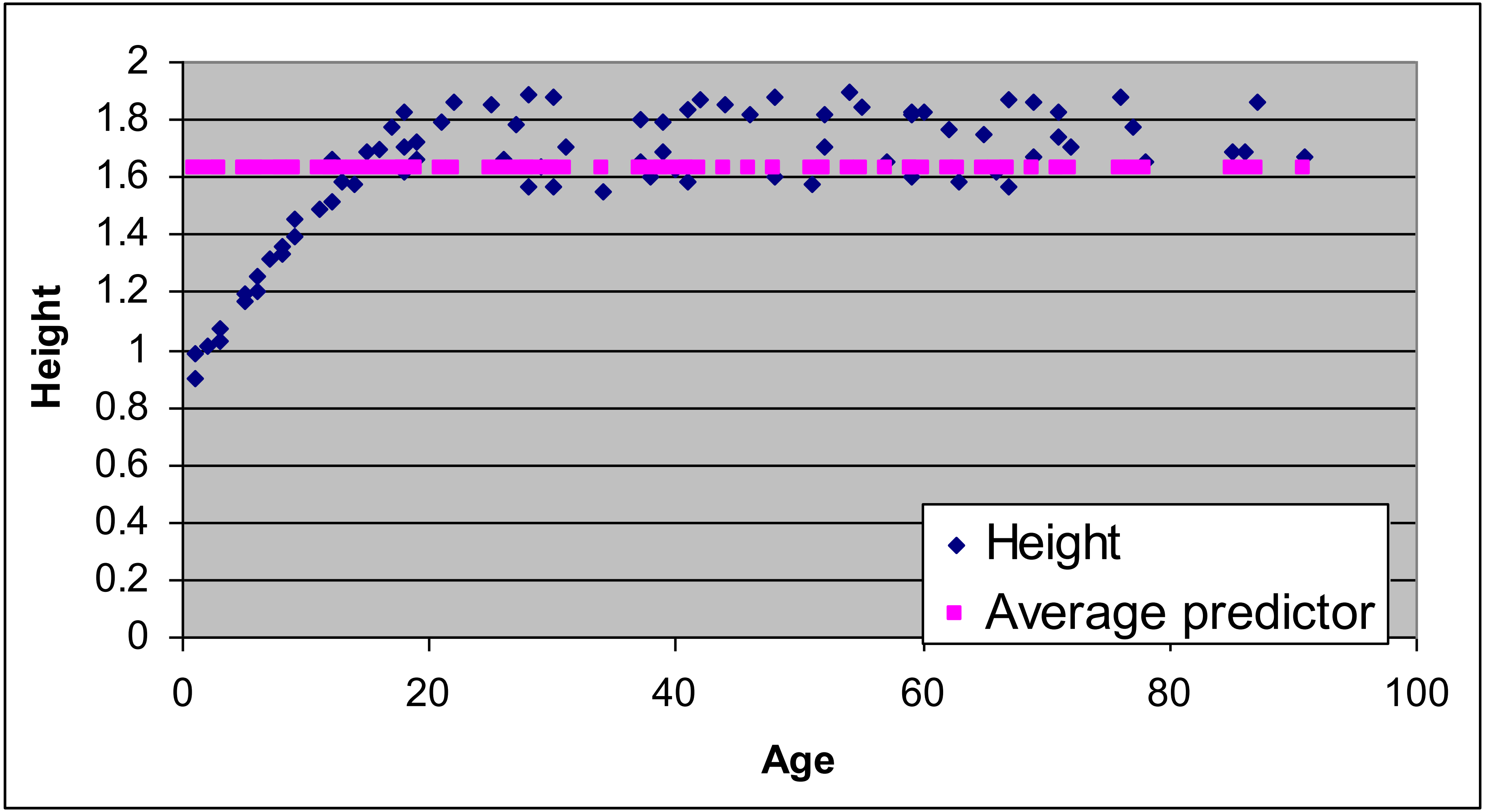
- data about 80 people: Age and Height



Age	Height
3	1.03
5	1.19
6	1.26
9	1.39
15	1.69
19	1.67
22	1.86
25	1.85
41	1.59
48	1.60
54	1.90
71	1.82
...	...

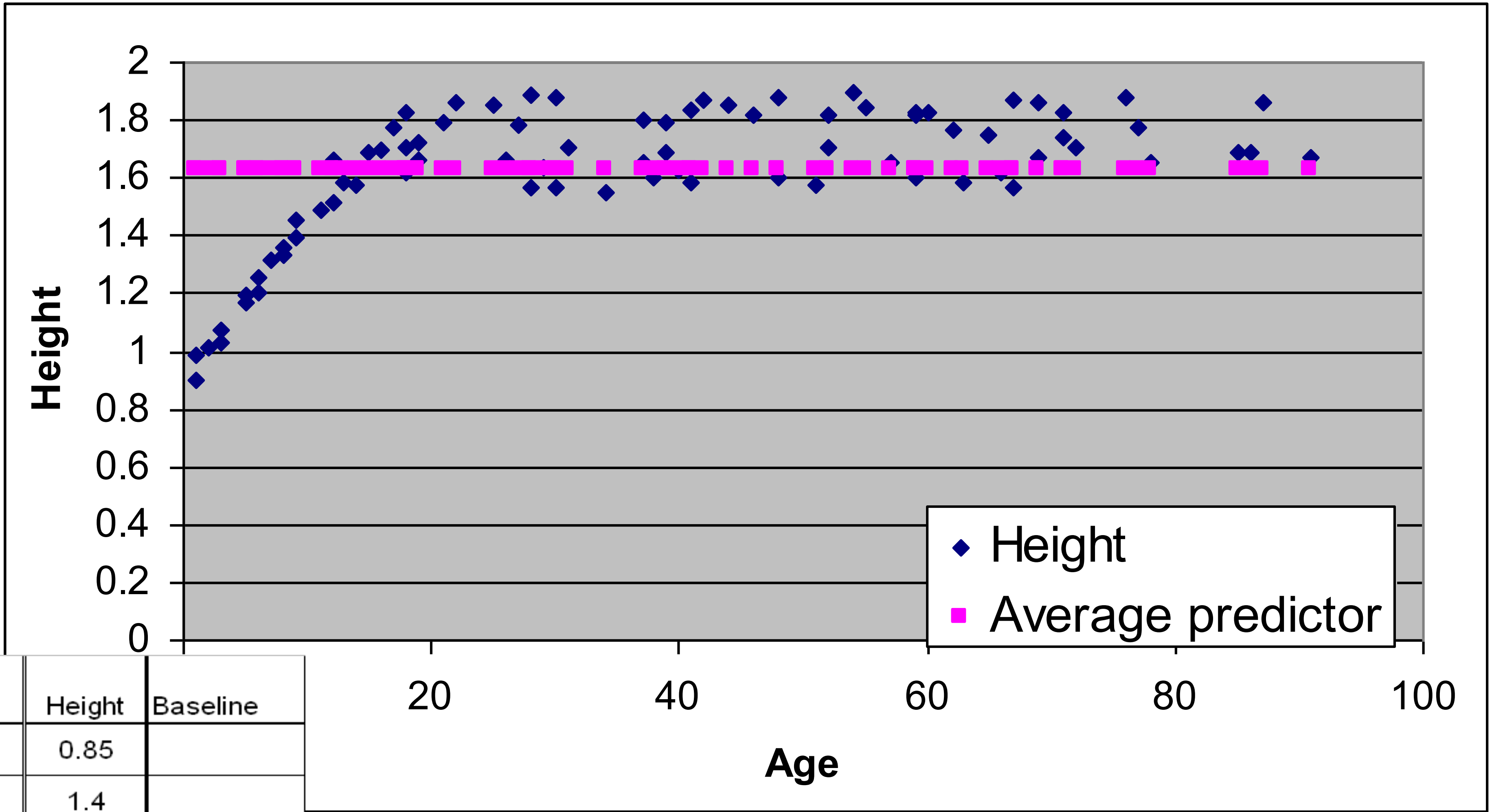
Baseline numeric model

- Average of the target variable



Baseline numeric predictor

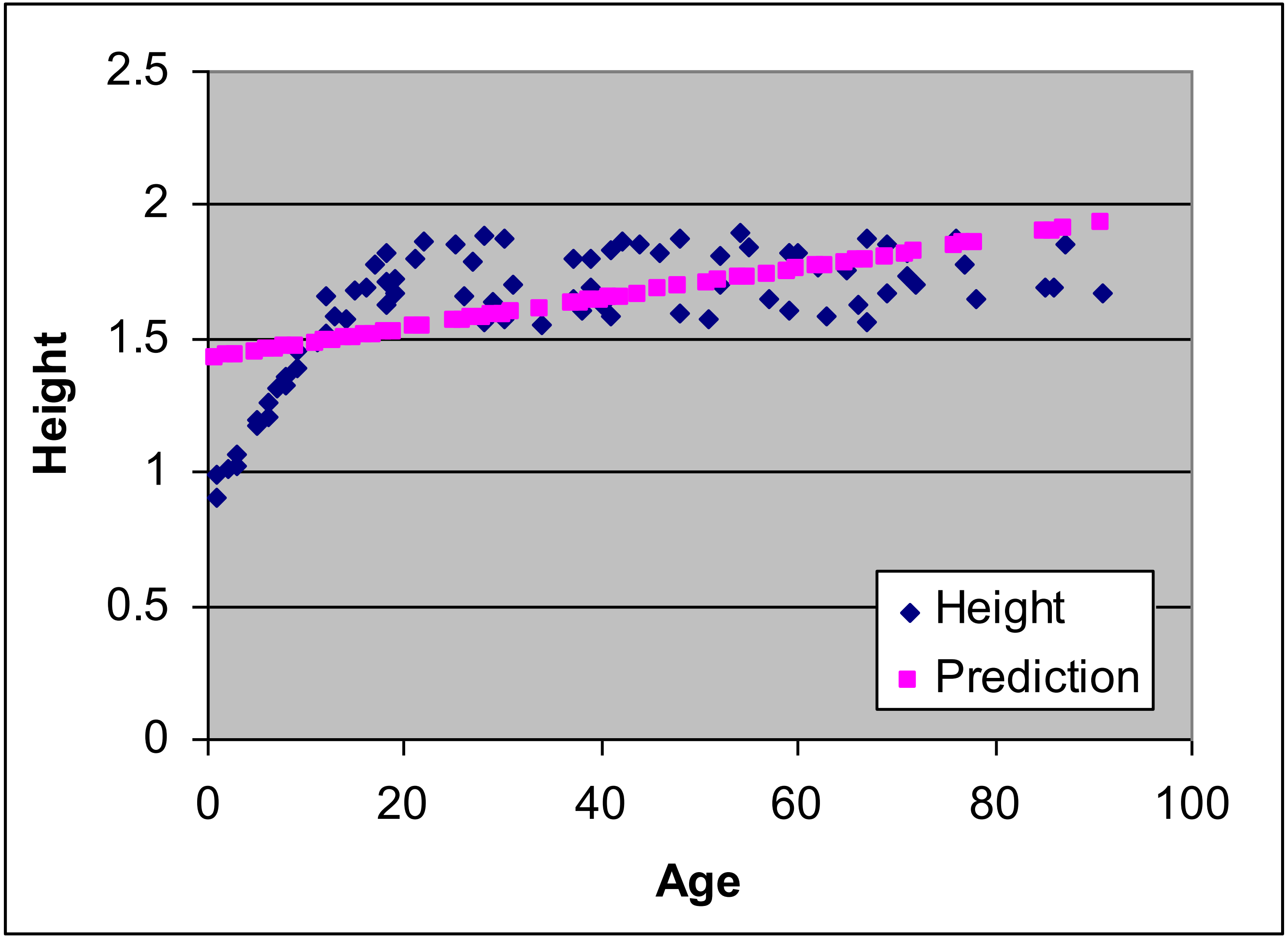
- Average of the target variable is 1.63



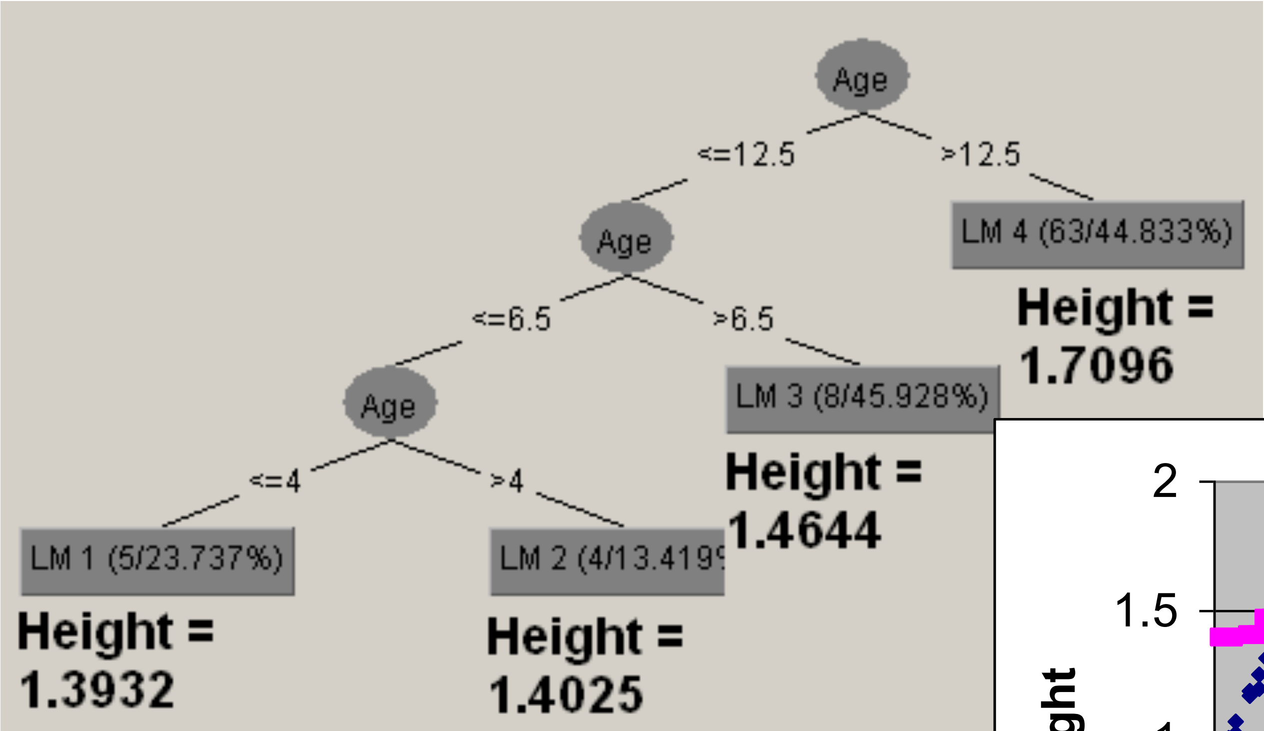
Age	Height	Baseline
2	0.85	
10	1.4	
35	1.7	
70	1.6	

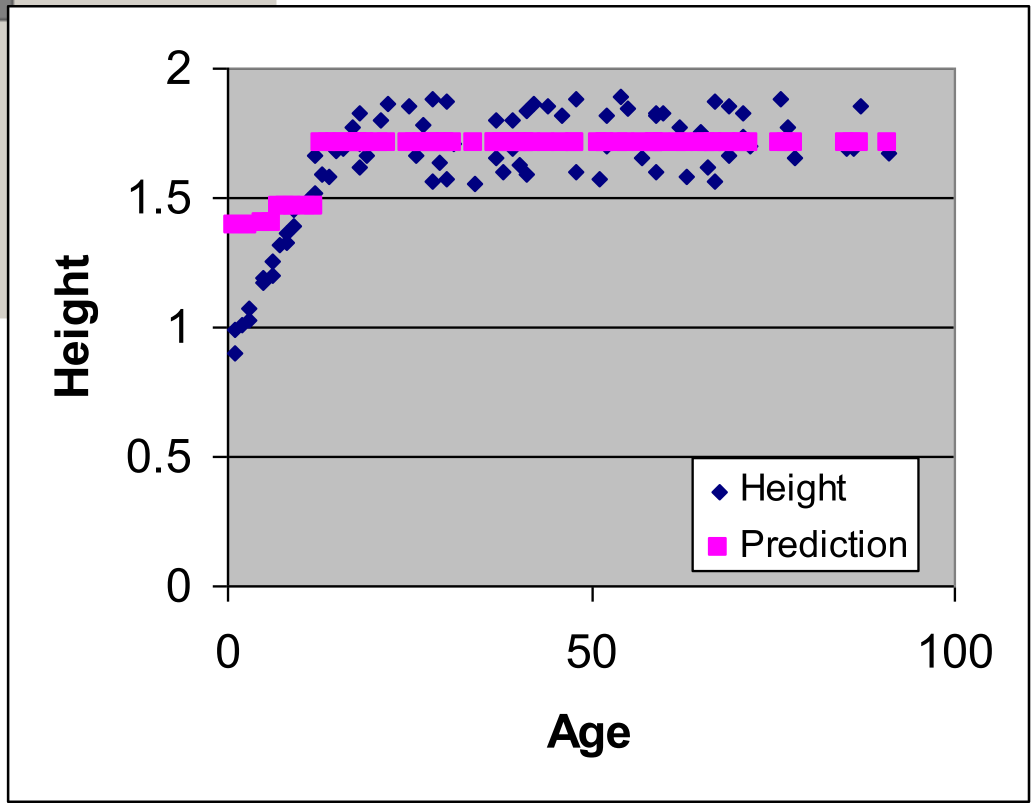
Linear Regression Model

Height = 0.0056 * Age + 1.4181

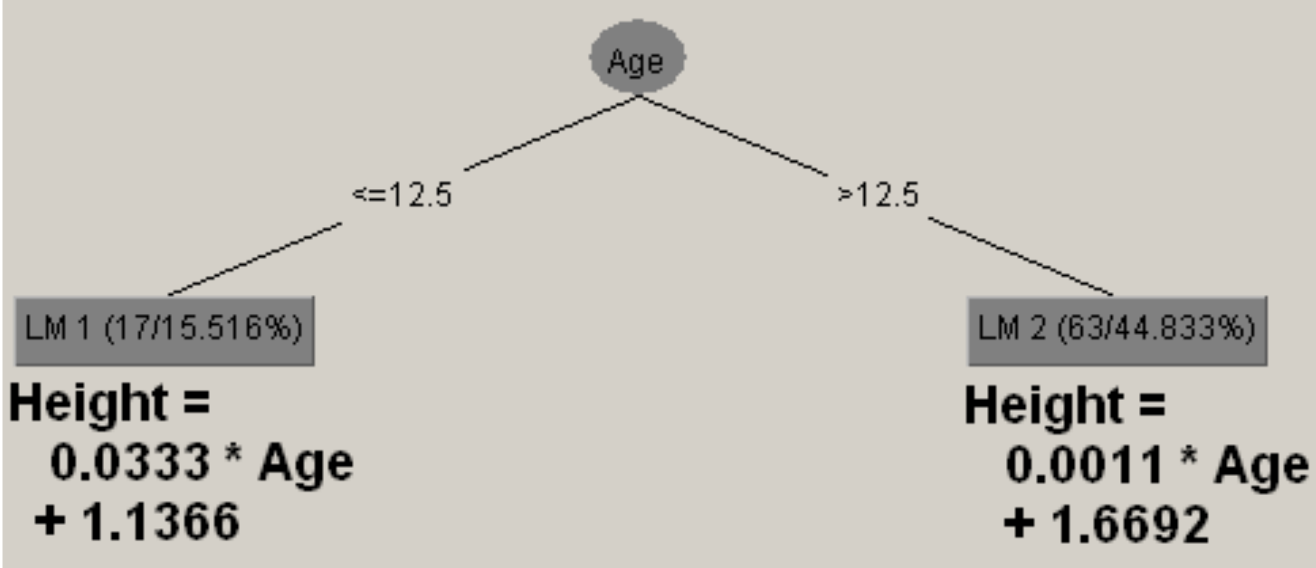


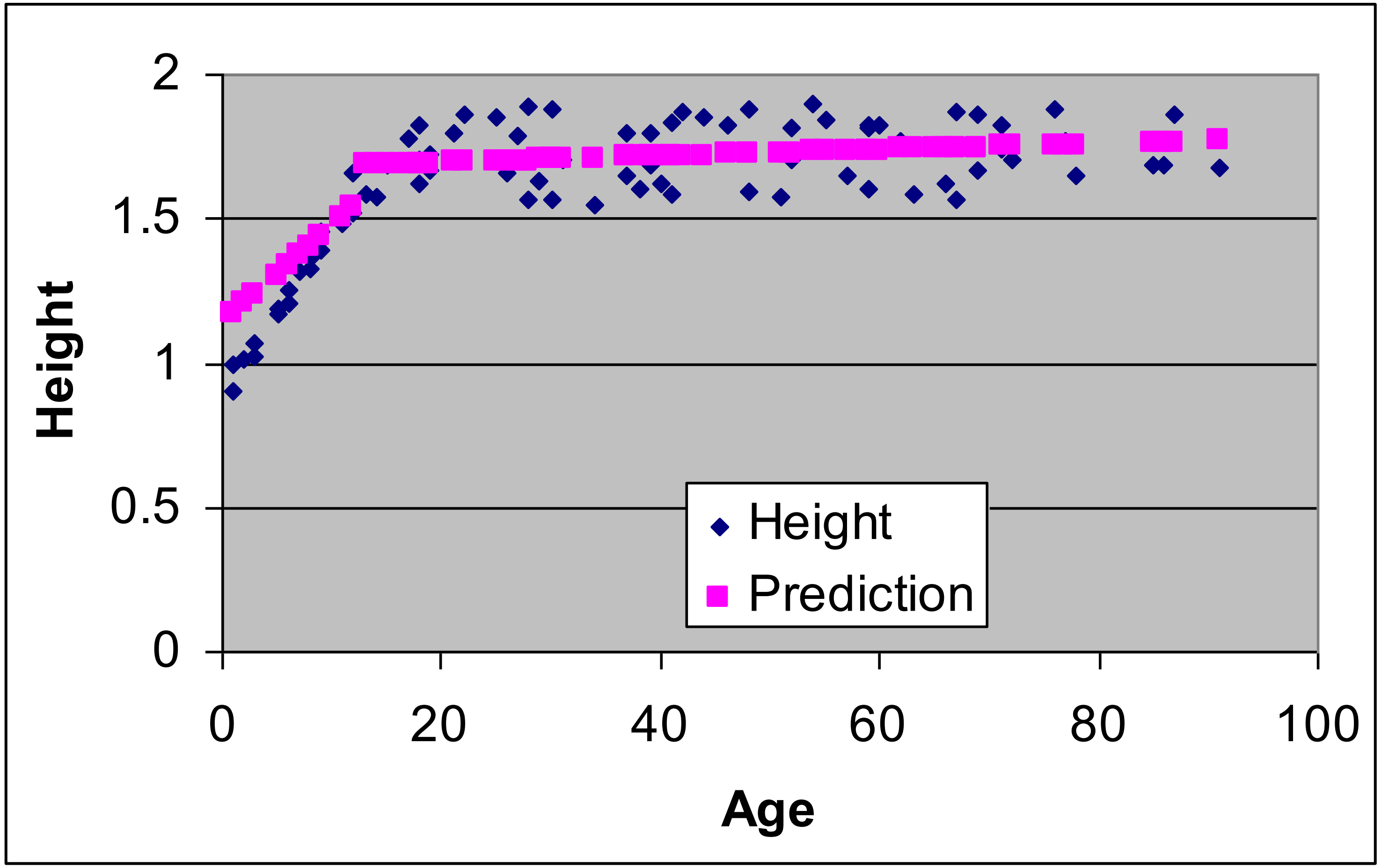
Regression tree





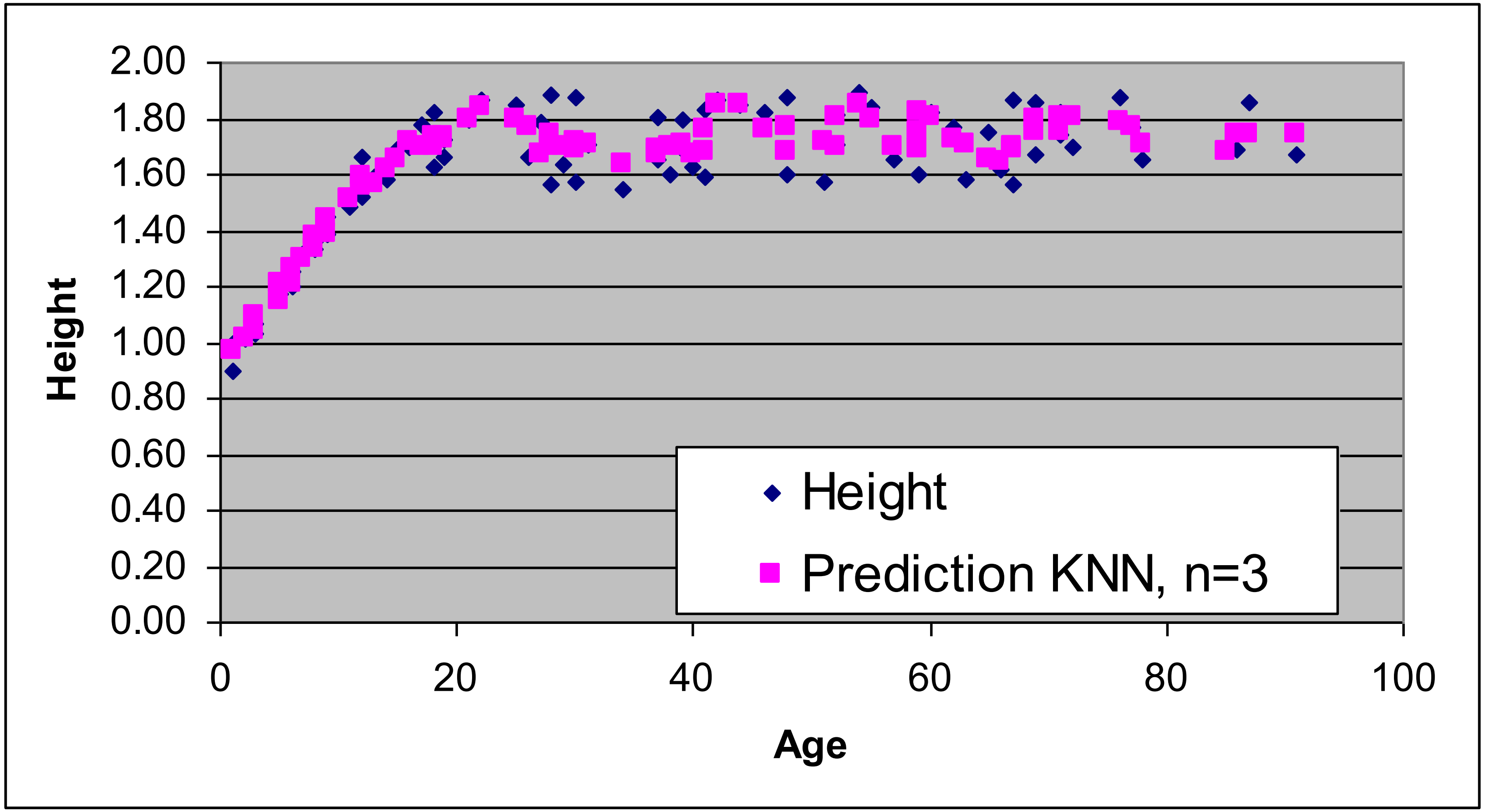
Model tree


Age
<=12.5
>12.5
LM 1 (17/15.516%)
LM 2 (63/44.833%)
Height =
0.0333 * Age
+ 1.1366
Height =
0.0011 * Age
+ 1.6692


Height
Age
Prediction
0
0.5
1
1.5
2
20
40
60
80
100

kNN – K nearest neighbors

- Looks at K closest examples (by age) and predicts the average of their target variable
- K=3



Machine Learning and Data Mining

- Machine Learning (ML) – computer algorithms/machines that learn predictive models from class-labeled data
- Data Mining (DM) – extraction of useful information from data: discovering relationships and patterns that have not previously been known,
  - sometimes used to denote the use of ML techniques applied to solving real-life data analysis problems

Data Mining

data

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses
O1	17	myope	no	reduced	NONE
O2	23	myope	no	normal	SOFT
O3	22	myope	yes	reduced	NONE
O4	27	myope	yes	normal	HARD
O5	19	hypermetrope	no	reduced	NONE
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	SOFT
O15	43	hypermetrope	yes	reduced	NONE
O16	39	hypermetrope	yes	normal	NONE
O17	54	myope	no	reduced	NONE
O18	62	myope	no	normal	NONE
O19-O23	...	...	...	...	...
O24	56	hypermetrope	yes	normal	NONE

data

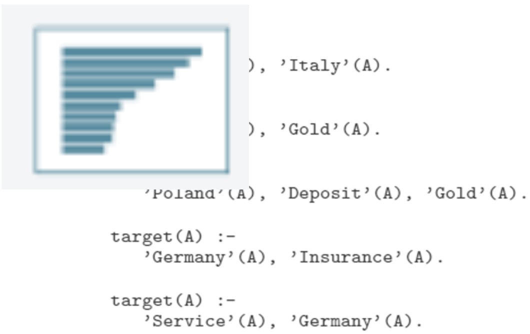
knowledge discovery
from data

Data Mining

patterns

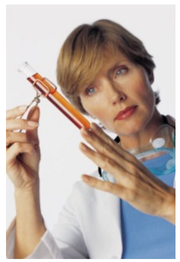
Given: class labeled or non-labeled data

Find: a set of interesting patterns, explaining the data



symbolic patterns

explanation



Knowledge Discovery in Databases

- Buzzword since 1996
- KDD is defined as “the process of identifying valid, novel, potentially useful and ultimately understandable models/patterns in data.” *

Usama M. Fayyad, Gregory Piatesky-Shapiro, Pedhraic Smyth: The KDD Process for Extracting Useful Knowledge form Volumes of Data. Comm ACM, Nov 96/Vol 39 No 11

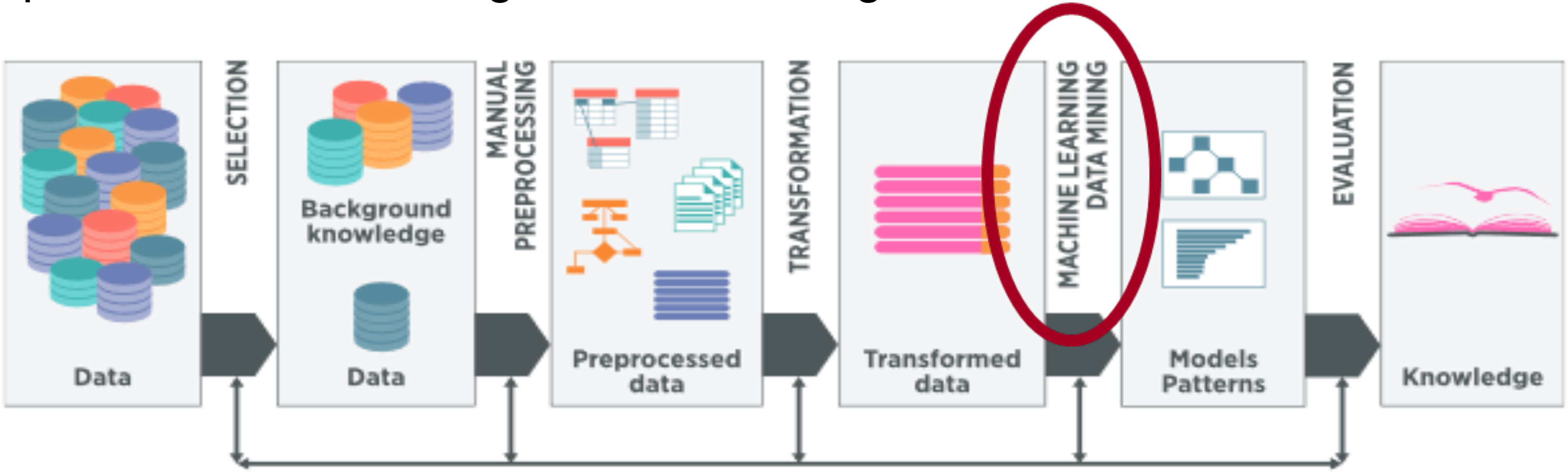
KDD Process

KDD process of discovering useful knowledge from data

- KDD is defined as “the process of identifying valid, novel, potentially useful and ultimately understandable models or patterns in data.”
- KDD process involves several phases:
  - data preparation
  - machine learning, data mining, statistics, …
  - evaluation and use of discovered patterns

KDD Process

KDD process of discovering useful knowledge from data



- KDD is defined as "the process of identifying valid, novel, potentially useful and ultimately understandable models or patterns in data."
- Machine Learning (ML) / Data Mining (DM) is the key step in the KDD process
  - performed using machine learning or pattern mining techniques for extracting classification models or interesting patterns in data
  - this key step represents only 15%-25% of entire KDD process

Second Generation Machine Learning

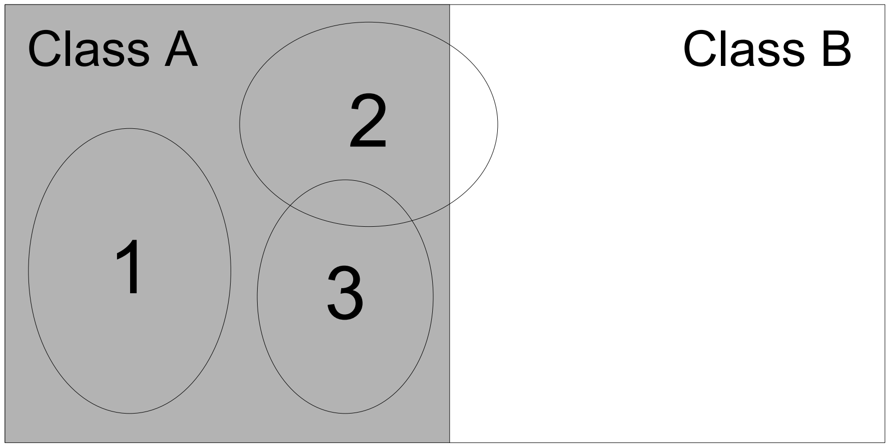
- **Developed since 1990s:**
  - Focused on data mining tasks characterized by large datasets described by large numbers of attributes
  - Industrial standard: CRISP-DM methodology (1997)

  - New conferences on practical aspects of data mining and knowledge discovery: KDD, PKDD, …
  - New learning tasks and efficient learning algorithms:
    - Learning descriptive patterns: association rule learning, subgroup discovery, …
    - Learning predictive models: Bayesian network learning,, relational data mining, statistical relational learning, SVMs, …

Subgroup Discovery

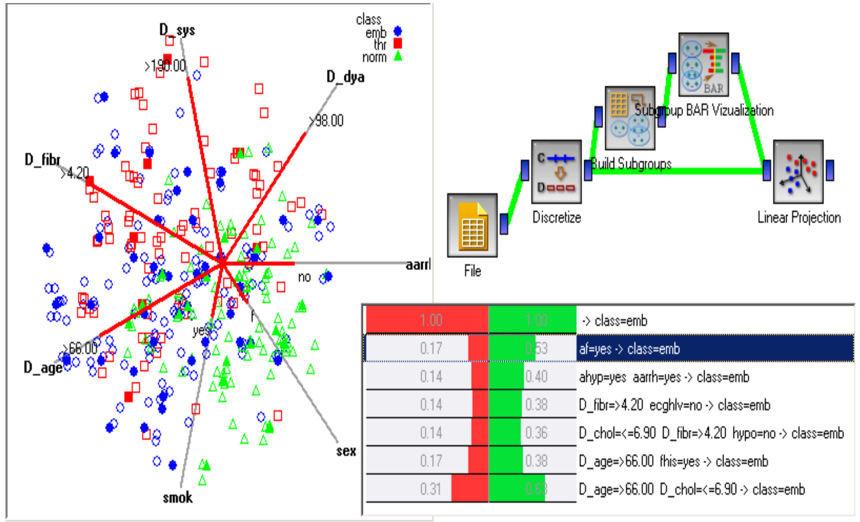
- Data transformation:
  - binary class values (positive vs. negative examples of Target class)
- Subgroup discovery:
  - a task in which individual interpretable patterns in the form of rules are induced from data, labeled by a predefined property of interest.
- SD algorithms learn several independent rules that describe groups of target class examples
  - subgroups must be large and significant

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses
O1	17	myope	no	reduced	NO
O2	23	myope	no	normal	YES
O3	22	myope	yes	reduced	NO
O4	27	myope	yes	normal	YES
O5	19	hypermetrope	no	reduced	NO
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	YES
O15	43	hypermetrope	yes	reduced	NO
O16	39	hypermetrope	yes	normal	NO
O17	54	myope	no	reduced	NO
O18	62	myope	no	normal	NO
O19-O23	...	...	...	...	...
O24	56	hypermetrope	yes	normal	NO



SD algorithms in Orange DM Platform

- **Orange** data mining toolkit
  – classification and subgroup discovery algorithms
  – data mining workflows
  – visualization



- **SD Algorithms in Orange**
  – SD (Gamberger & Lavrač, JAIR 2002)
  – Apriori-SD (Kavšek & Lavrač, AAI 2006)
  – CN2-SD (Lavrač et al., JMLR 2004)

Relational Data Mining

customer

ID	Zip	Sex	So St	Income	Age	Club	Resp
...	...	...	...	...	...	...	...
3478	34677	m	si	60-70	32	me	nr
3479	43666	f	ma	80-90	45	nm	re
...	...	...	...	...	...	...	...

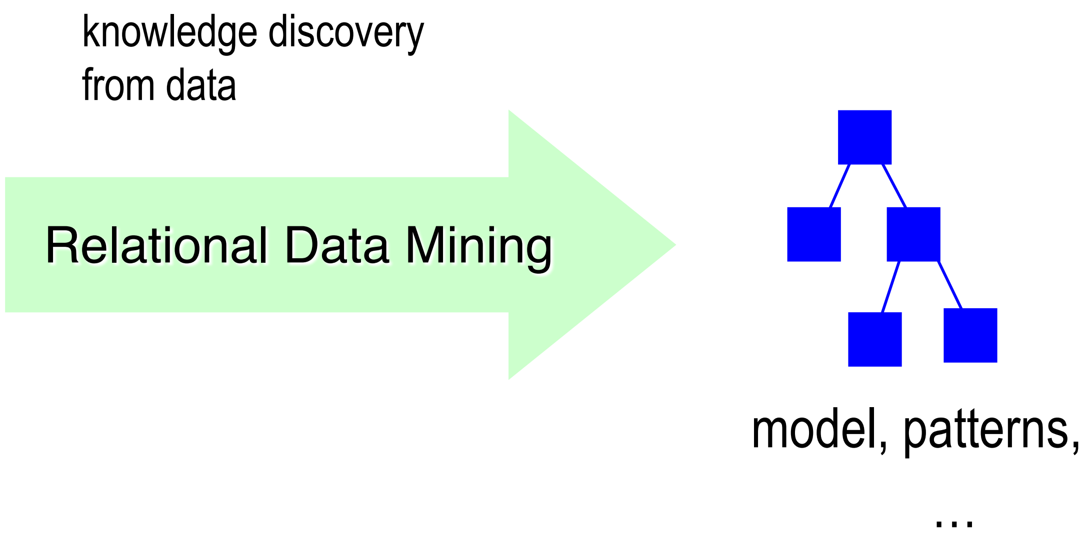
order

Customer ID	Order ID	Store ID	Delivery Mode	Paymt Mode
...	...	...	...	...
3478	2140267	12	regular	cash
3478	3446778	12	express	check
3478	4728386	17	regular	check
3479	3233444	17	express	credit
3479	3475886	12	regular	credit
...	...	...	...	...

store

Store ID	Size	Type	Location
...	...	...	...
12	small	franchise	city
17	large	indep	rural
...	...	...	...

Relational representation of customers, orders and stores.



Given: a relational database, a set of tables, sets of logical facts, a graph, …

Find: a classification model, a set of patterns

Relational Data Mining

- **ILP, relational learning, relational data mining**
  - Learning from complex relational databases

customer							
ID	Zip	Sex	SoSt	Income	Age	Club	Resp
...	...	...	...	...	...	...	...
3478	34677	m	si	60-70	32	me	nr
3479	43666	f	ma	80-90	45	nm	re
...	...	...	...	...	...	...	...

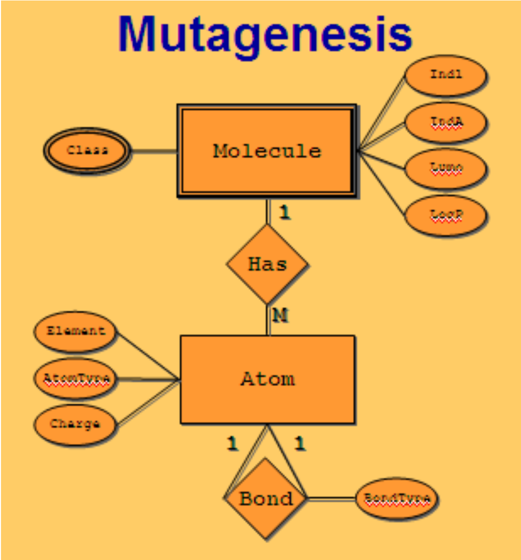
order				
Customer ID	Order ID	Store ID	Delivery Mode	Paymt Mode
...	...	...	...	...
3478	2140267	12	regular	cash
3478	3446778	12	express	check
3478	4728386	17	regular	check
3479	3233444	17	express	credit
3479	3475886	12	regular	credit
...	...	...	...	...

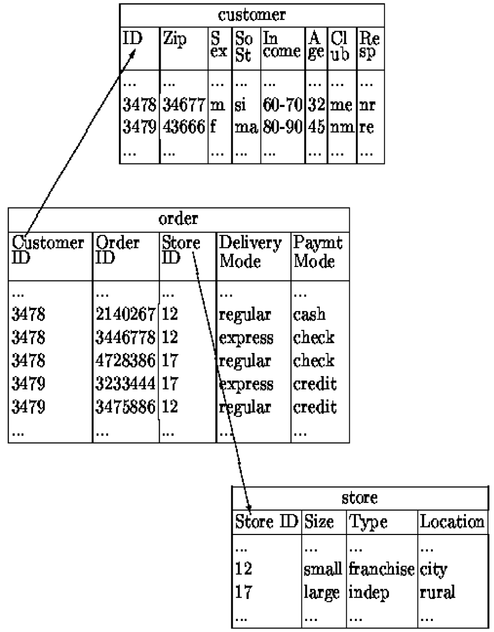
store			
Store ID	Size	Type	Location
...	...	...	...
12	small	franchise	city
17	large	indep	rural
...	...	...	...

Relational representation of customers, orders and stores.

Relational Data Mining

- **ILP, relational learning, relational data mining**
  - Learning from complex relational databases
  - Learning from complex structured data, e.g. molecules and their biochemical properties





customer							
ID	Zip	Sex	SoSt	Income	Age	Club	Resp
...	...	...	...	...	...	...	...
3478	34677	m	si	60-70	32	me	nr
3479	43666	f	ma	80-90	45	nm	re
...	...	...	...	...	...	...	...

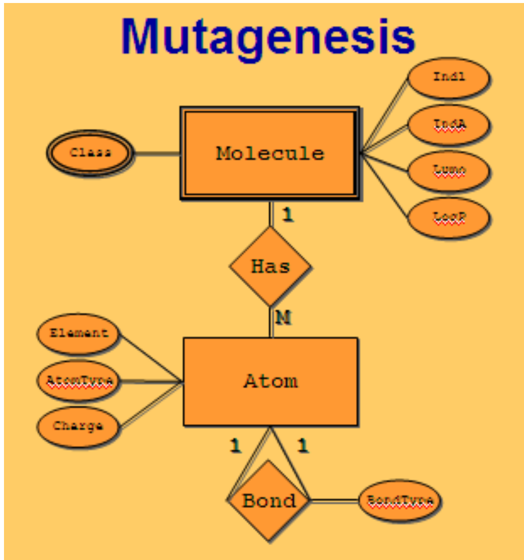
order				
Customer ID	Order ID	Store ID	Delivery Mode	Paymt Mode
...	...	...	...	...
3478	2140267	12	regular	cash
3478	3446778	12	express	check
3478	4728386	17	regular	check
3479	3233444	17	express	credit
3479	3475886	12	regular	credit
...	...	...	...	...

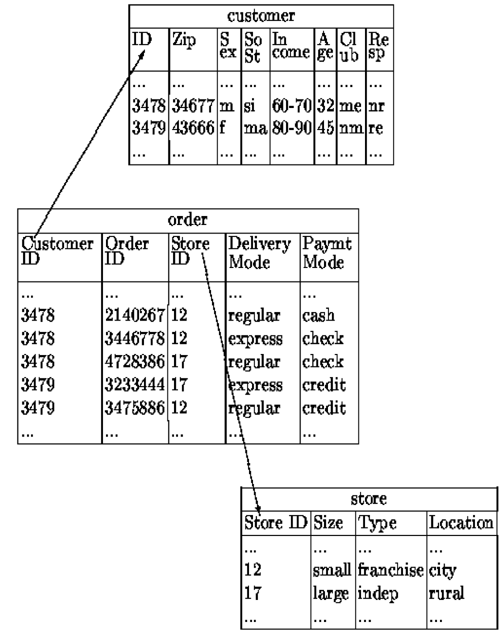
store			
Store ID	Size	Type	Location
...	...	...	...
12	small	franchise	city
17	large	indep	rural
...	...	...	...

Relational representation of customers, orders and stores.

Relational and Semantic Data Mining

- **ILP, relational learning, relational data mining**
  - Learning from complex relational databases
  - Learning from complex structured data, e.g. molecules and their biochemical properties
  - Learning by using domain knowledge in the form of ontologies = **semantic data mining**





customer

ID	Zip	Sex	SoSt	Income	Age	Club	Resp
...	...	...	...	...	...	...	...
3478	34677	m	si	60-70	32	me	nr
3479	43666	f	ma	80-90	45	nm	re
...	...	...	...	...	...	...	...

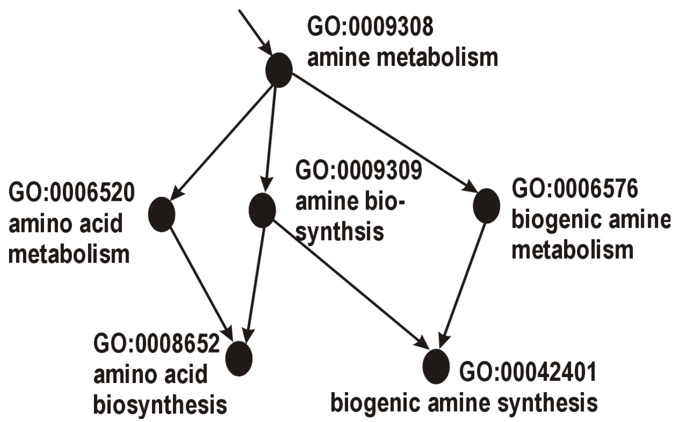
order

Customer ID	Order ID	Store ID	Delivery Mode	Paymt Mode
...	...	...	...	...
3478	2140267	12	regular	cash
3478	3446778	12	express	check
3478	4728386	17	regular	check
3479	3233444	17	express	credit
3479	3475886	12	regular	credit
...	...	...	...	...

store

Store ID	Size	Type	Location
...	...	...	...
12	small	franchise	city
17	large	indep	rural
...	...	...	...

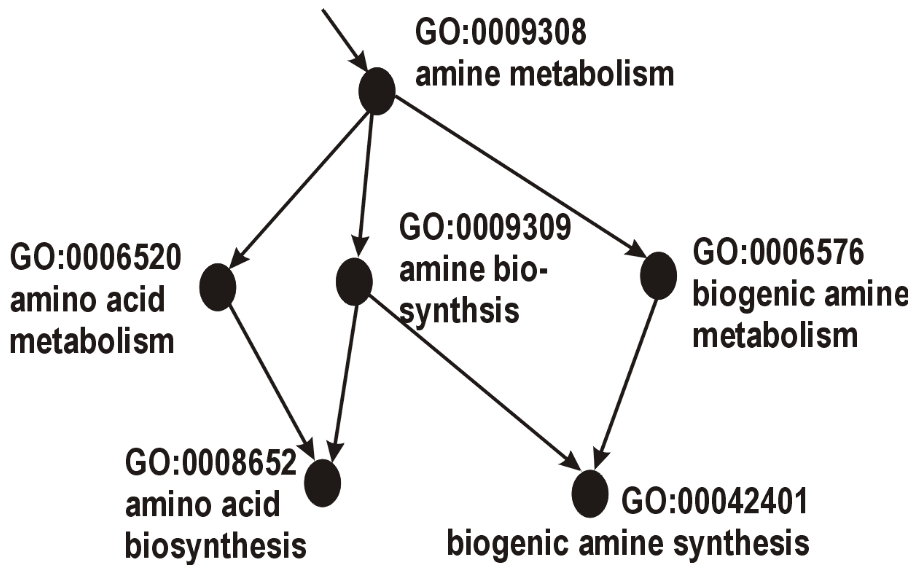
Relational representation of customers, orders and stores.



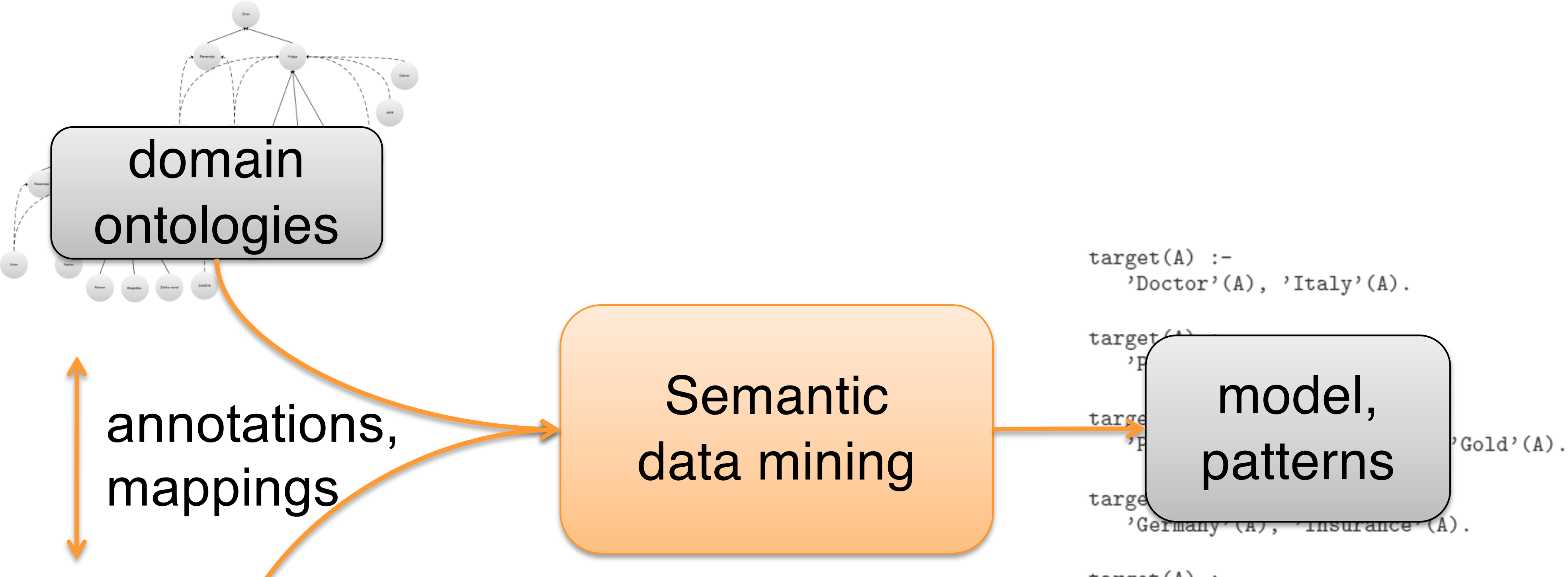
Using domain ontologies

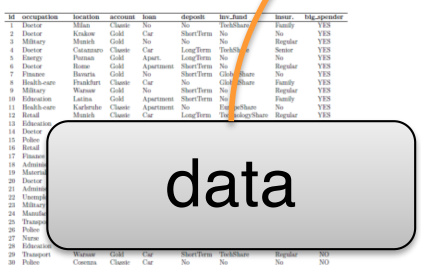
Using domain ontologies as background knowledge, e.g., using the Gene Ontology (GO)

- GO is a database of terms, describing gene sets in terms of their
  - functions (12,093)
  - processes (1,812)
  - components (7,459)
- Genes are annotated to GO terms
- Terms are connected (is_a, part_of)
- Levels represent terms generality



Semantic Data Mining: Using ontologies as background knowledge in RDM





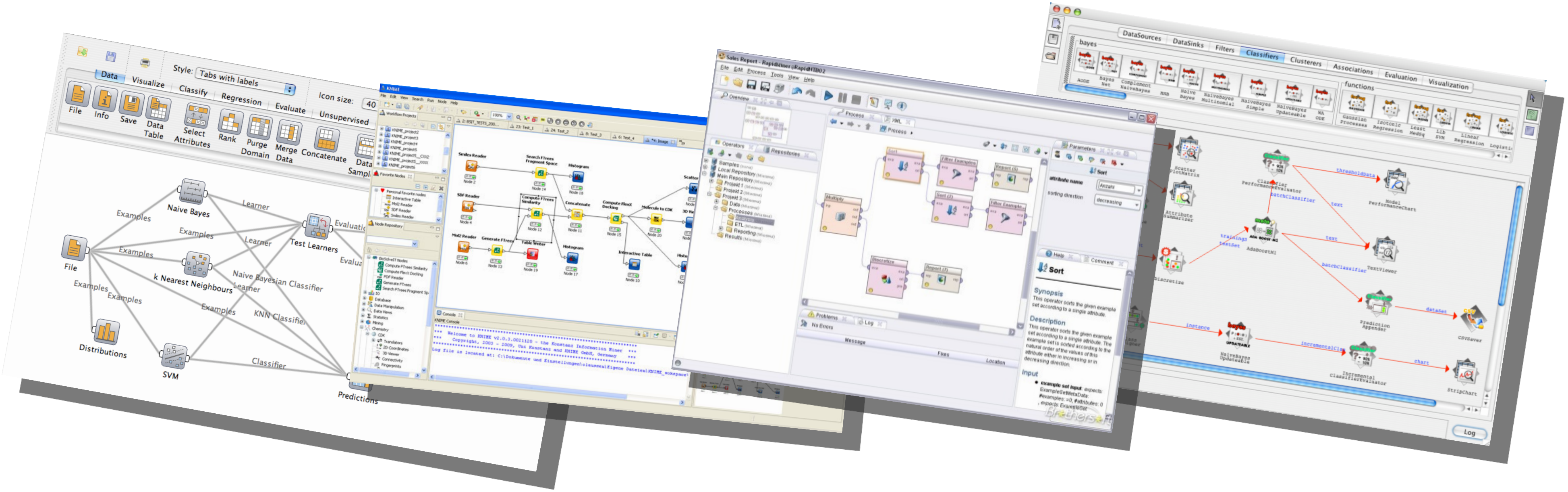
Given:

- transaction data table, relational database, text documents, Web pages, …
- one or more domain ontologies

Find: a classification model, a set of patterns

Second Generation Data Mining Platforms

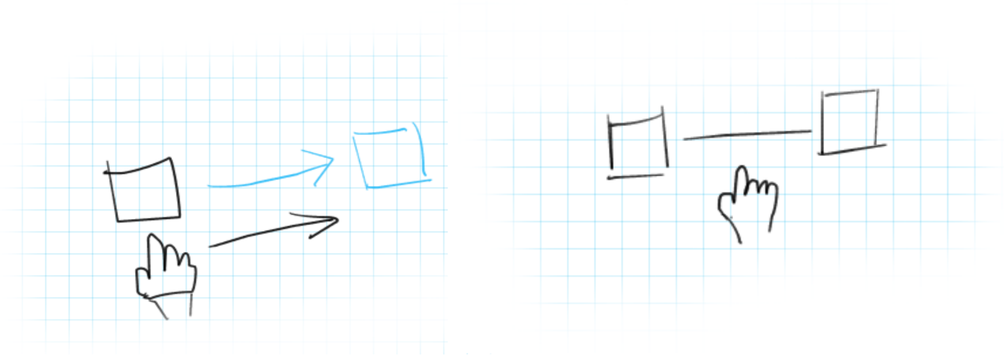
Orange, WEKA, KNIME, RapidMiner, …



– include numerous data mining algorithms
– enable data and model visualization
– like Orange, Taverna, WEKA, KNIME, RapidMiner, also enable complex **workflow** construction

Data Mining Workflows for Open Data Science

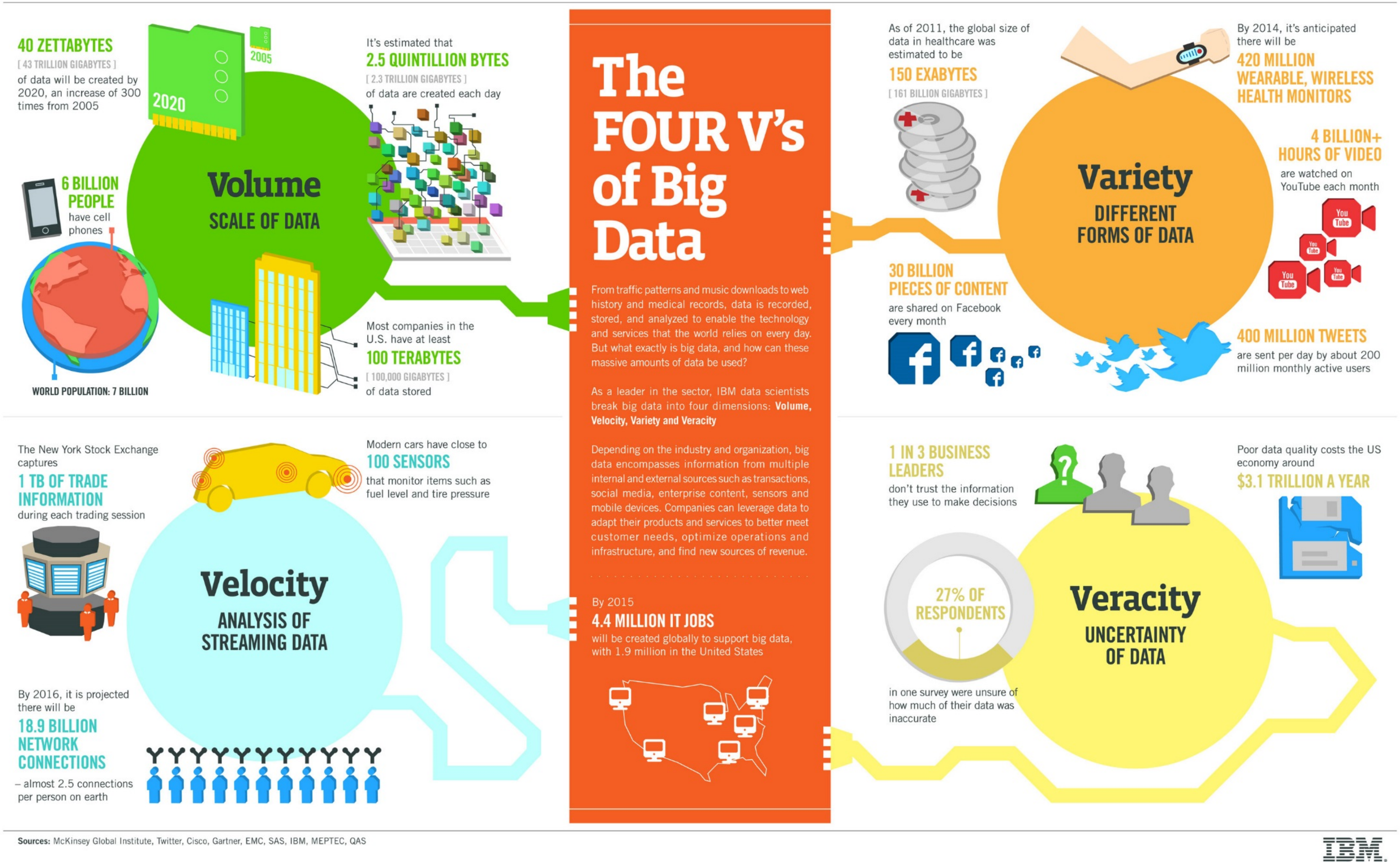
- Workflows are executable visual representations of procedures
  - divided into smaller chunks of code (components)
  - organized as sequences of connected components.
- Suitable for representing complex scientific pipelines
  - by explicitly modeling dependencies of components
- Building scientific workflows consists of simple operations on workflow elements (drag, drop, connect), suitable for non-experts



Big Data

- Big Data – Buzzword since 2008 (special issue of Nature on Big Data)
  - data and techniques for dealing with very large volumes of data, possibly dynamic data streams
  - requiring large data storage resources, special algorithms for parallel computing architectures.

The 4 Vs of Big Data



Data Science

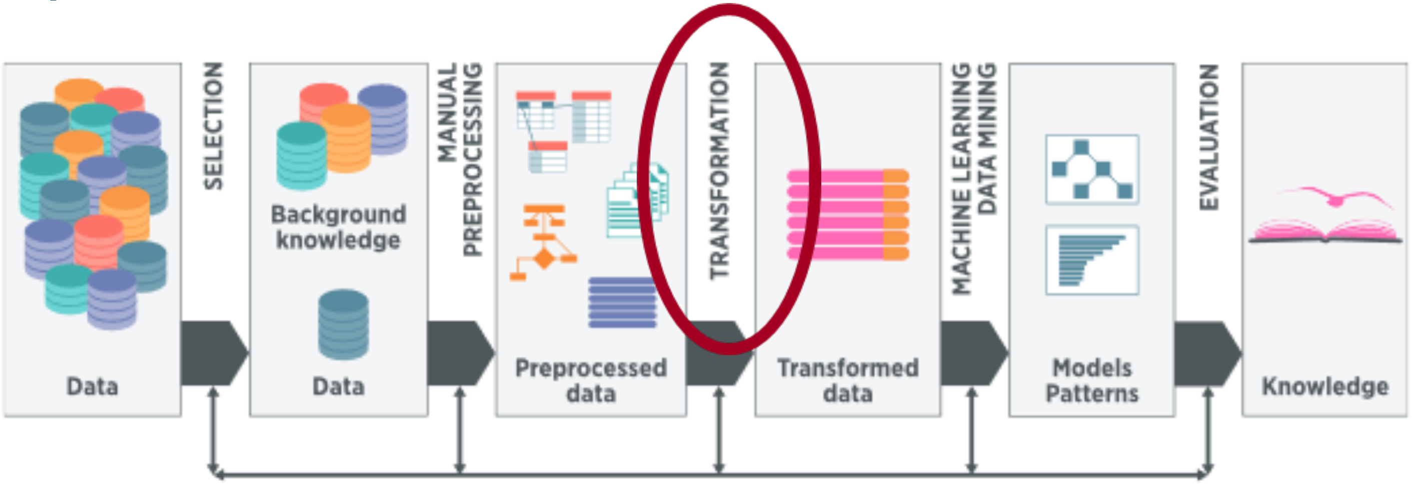
- Data Science – buzzword since 2012 when Harvard Business Review called it "The Sexiest Job of the 21st Century"
  - an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured, similar to data mining.
  - used interchangeably with earlier concepts like business analytics, business intelligence, predictive modeling, and statistics.

Third Generation Machine Learning

- **Developed since 2010s:**
  - Focused on big data analytics
  - Addressing complex data mining tasks and scenarios
  - New conferences on data science and big data analytics; e.g., IEEE Big Data, Complex networks, …
  - New learning tasks and efficient learning algorithms:
    - Analysis of dynamic data streams, Network analysis, Text mining, Semantic data analysis, …
  - Lots of emphasis on automated **data transformation, i.e. representation learning**

Representation Learning

- KDD process:

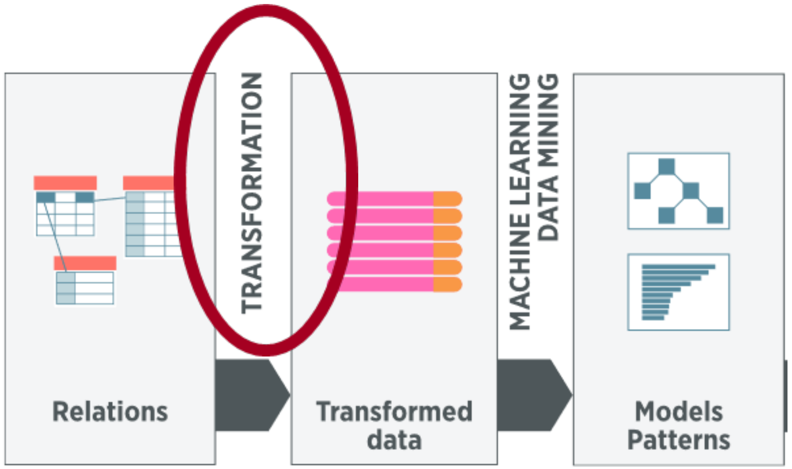


- Representation learning = Automated data transformation, performed on manually preprocessed data

- Transformation requires handling heterogeneous data
  - Data (feature vectors, documents, pictures, data streams, …)
  - Background knowledge (multi-relational data tables, networks, text corpora, …)

Representation Learning in Relation Learning setting

- Representation learning = automated transformation of multi-relational data



- Two main approaches:
  - Propositionalization of relational data, of heterogeneous information networks, …
  - Embeddings of texts, networks, knowledge graphs, entities (features), … is highly popular in the last few years

Propositionalization: Data transformation for Relational Learning

customer

ID	Zip	Sex	So St	Income	Age	Club	Resp
...	...	...	...	...	...	...	...
3478	34677	m	si	60-70	32	me	nr
3479	43666	f	ma	80-90	45	nm	re
...	...	...	...	...	...	...	...

order

Customer ID	Order ID	Store ID	Delivery Mode	Paymt Mode
...	...	...	...	...
3478	2140267	12	regular	cash
3478	3446778	12	express	check
3478	4728386	17	regular	check
3479	3233444	17	express	credit
3479	3475886	12	regular	credit
...	...	...	...	...

store

Store ID	Size	Type	Location
...	...	...	...
12	small	franchise	city
17	large	indep	rural
...	...	...	...

Relational representation of customers, orders and stores.

Step 1

Propositionalization

1. constructing relational features
2. constructing a propositional table

	f1	f2	f3	f4	f5	f6	...				...	fn
g1	1	0	0	1	1	1	0	0	1	0	1	1
g2	0	1	1	0	1	1	0	0	0	1	1	0
g3	0	1	1	1	0	0	1	1	0	0	0	1
g4	1	1	1	0	1	1	0	0	1	1	1	0
g5	1	1	1	0	0	1	0	1	1	0	1	0
g1	0	0	1	1	0	0	0	1	0	0	0	1
g2	1	1	0	0	1	1	0	1	0	1	1	1
g3	0	0	0	0	1	0	0	1	1	1	0	0
g4	1	0	1	1	1	0	1	0	0	1	0	1

Propositionalization: Data transformation for Relational Learning

customer							
ID	Zip	Sex	SoSt	Income	Age	Club	Resp
...	...	...	...	...	...	...	...
3478	34677	m	si	60-70	32	me	nr
3479	43666	f	ma	80-90	45	nm	re
...	...	...	...	...	...	...	...

order				
Customer ID	Order ID	Store ID	Delivery Mode	Paymt Mode
...	...	...	...	...
3478	2140267	12	regular	cash
3478	3446778	12	express	check
3478	4728386	17	regular	check
3479	3233444	17	express	credit
3479	3475886	12	regular	credit
...	...	...	...	...

store			
Store ID	Size	Type	Location
...	...	...	...
12	small	franchise	city
17	large	indep	rural
...	...	...	...

Relational representation of customers, orders and stores.

Step 1

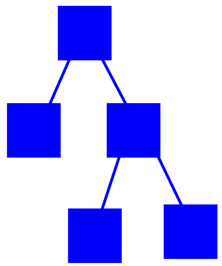
Propositionalization

	f1	f2	f3	f4	f5	f6	...				...	fn
g1	1	0	0	1	1	1	0	0	1	0	1	1
g2	0	1	1	0	1	1	0	0	0	1	1	0
g3	0	1	1	1	0	0	1	1	0	0	0	1
g4	1	1	1	0	1	1	0	0	1	1	1	0
g5	1	1	1	0	0	1	0	1	1	0	1	0
g1	0	0	1	1	0	0	0	1	0	0	0	1
g2	1	1	0	0	1	1	0	1	0	1	1	1
g3	0	0	0	0	1	0	0	1	1	1	0	0
g4	1	0	1	1	1	0	1	0	0	1	0	1

	f1	f2	f3	f4	f5	f6	...				...	fn
g1	1	0	0	1	1	1	0	0	1	0	1	1
g2	0	1	1	0	1	1	0	0	0	1	1	0
g3	0	1	1	1	0	0	1	1	0	0	0	1
g4	1	1	1	0	1	1	0	0	1	1	1	0
g5	1	1	1	0	0	1	0	1	1	0	1	0
g1	0	0	1	1	0	0	0	1	0	0	0	1
g2	1	1	0	0	1	1	0	1	0	1	1	1
g3	0	0	0	0	1	0	0	1	1	1	0	0
g4	1	0	1	1	1	0	1	0	0	1	0	1

Step 2

Machine Learning



model, patterns, …

Propositionalization: Data transformation for Relational Learning

customer

ID	Zip	Sex	SoSt	Income	Age	Club	Resp
...	...	...	...	...	...	...	...
3478	34677	m	si	60-70	32	me	nr
3479	43666	f	ma	80-90	45	nm	re
...	...	...	...	...	...	...	...

order

Customer ID	Order ID	Store ID	Delivery Mode	Paymt Mode
...	...	...	...	...
3478	2140267	12	regular	cash
3478	3446778	12	express	check
3478	4728386	17	regular	check
3479	3233444	17	express	credit
3479	3475886	12	regular	credit
...	...	...	...	...

store

Store ID	Size	Type	Location
...	...	...	...
12	small	franchise	city
17	large	indep	rural
...	...	...	...

Relational representation of customers, orders and stores.

Step 1

Propositionalization

1. construct relational features
2. construct a propositional table

	f1	f2	f3	f4	f5	f6	...				...	fn
g1	1	0	0	1	1	1	0	0	1	0	1	1
g2	0	1	1	0	1	1	0	0	0	1	1	0
g3	0	1	1	1	0	0	1	1	0	0	0	1
g4	1	1	1	0	1	1	0	0	1	1	1	0
g5	1	1	1	0	0	1	0	1	1	0	1	0
g1	0	0	1	1	0	0	0	1	0	0	0	1
g2	1	1	0	0	1	1	0	1	0	1	1	1
g3	0	0	0	0	1	0	0	1	1	1	0	0
g4	1	0	1	1	1	0	1	0	0	1	0	1

	f1	f2	f3	f4	f5	f6	...				...	fn
g1	1	0	0	1	1	1	0	0	1	0	1	1
g2	0	1	1	0	1	1	0	0	0	1	1	0
g3	0	1	1	1	0	0	1	1	0	0	0	1
g4	1	1	1	0	1	1	0	0	1	1	1	0
g5	1	1	1	0	0	1	0	1	1	0	1	0
g1	0	0	1	1	0	0	0	1	0	0	0	1
g2	1	1	0	0	1	1	0	1	0	1	1	1
g3	0	0	0	0	1	0	0	1	1	1	0	0
g4	1	0	1	1	1	0	1	0	0	1	0	1

Step 2

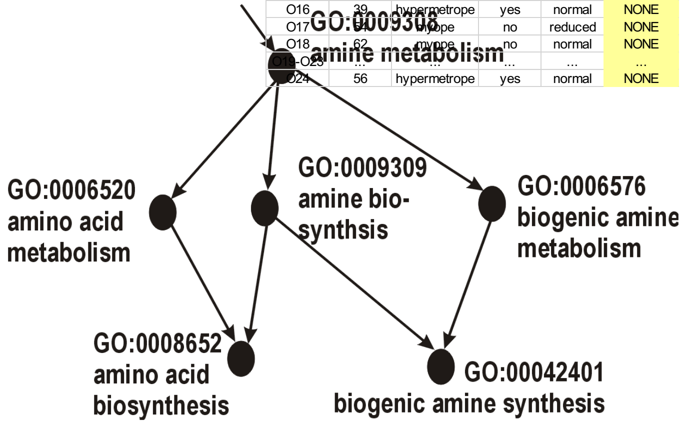
Subgroup discovery

```
target(A) :-
    'Doctor'(A), 'Italy'(A).

target(A) :-
    'Public'(A), 'Gold'(A).

target(A) :-
    'Poland'(A), 'Deposit'(A), 'Gold'(A).

target(A) :-
    'Germany'(A), 'Insurance'(A).

target(A) :-
    'Service'(A), 'Germany'(A).
```

patterns (set of rules)

Propositionalization: Data transformation for Semantic Data Mining

Person	Age	Spect. presc.	Astigm.	Tear prod.	Lenses
O1	17	myope	no	reduced	NONE
O2	23	myope	no	normal	SOFT
O3	22	myope	yes	reduced	NONE
O4	27	myope	yes	normal	HARD
O5	19	hypermetrope	no	reduced	NONE
O6-O13	...	...	...	...	...
O14	35	hypermetrope	no	normal	SOFT
O15	43	hypermetrope	yes	reduced	NONE
O16	39	hypermetrope	yes	normal	NONE
O17	54	myope	no	reduced	NONE
O18	62	myope	no	normal	NONE
O19-O23	...	...	...	...	...
O24	56	hypermetrope	yes	normal	NONE



Step 1

Propositionalization

1. constructing relational features
2. constructing a propositional table

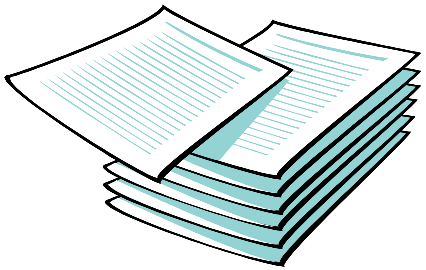
	f1	f2	f3	f4	f5	f6	...				...	fn
g1	1	0	0	1	1	1	0	0	1	0	1	1
g2	0	1	1	0	1	1	0	0	0	1	1	0
g3	0	1	1	1	0	0	1	1	0	0	0	1
g4	1	1	1	0	1	1	0	0	1	1	1	0
g5	1	1	1	0	0	1	0	1	1	0	1	0
g1	0	0	1	1	0	0	0	1	0	0	0	1
g2	1	1	0	0	1	1	0	1	0	1	1	1
g3	0	0	0	0	1	0	0	1	1	1	0	0
g4	1	0	1	1	1	0	1	0	0	1	0	1

The approach: Using relational subgroup discovery in the SDM context

- General purpose system **RSD** for **Relational Subgroup Discovery**, using a propositionalization approach to relational data mining
- Applied to semantic data mining in a biomedical application by using the Gene Ontology as background knowledge in analyzing microarray data

Železny and Lavrac, MLJ 2006

Text mining: Viewed in propositionalization context: BoW data transformation



Step 1

BoW vector construction

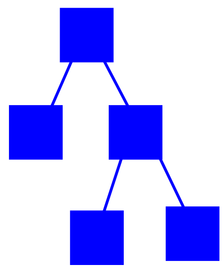
Document	Word1	Word2	…	WordN	Class
d1	1	1	0	1	NO
d2	1	1	0	0	YES
d3	1	1	1	1	NO
d4	1	1	1	0	YES
d5	1	0	0	1	NO
d6-d13	…	…	…	…	…
d14	0	0	0	0	YES
d15	0	0	1	1	NO
d16	0	0	1	0	NO
d17	0	1	0	1	NO
d18	0	1	0	0	NO
d19-d23	…	…	…	…	…
d24	0	0	1	0	NO

1. BoW features construction
2. Table of BoW vectors construction

Document	Word1	Word2	…	WordN	Class
d1	1	1	0	1	NO
d2	1	1	0	0	YES
d3	1	1	1	1	NO
d4	1	1	1	0	YES
d5	1	0	0	1	NO
d6-d13	…	…	…	…	…
d14	0	0	0	0	YES
d15	0	0	1	1	NO
d16	0	0	1	0	NO
d17	0	1	0	1	NO
d18	0	1	0	0	NO
d19-d23	…	…	…	…	…
d24	0	0	1	0	NO

Step 2

Data Mining



model, patterns, clusters, …

BoW construction: Feature weights and Cosine similarity between document vectors

- Each document D is represented as a vector of TF-IDF weights

$$tfidf(w) = tf.\log(\frac{N}{df(w)})$$

- Similarity between two vectors is estimated by the similarity between their vector representations (cosine of the angle between the two vectors):

$$Similarity(D_1, D_2) = \frac{\sum_i x_{1i} x_{2i}}{\sqrt{\sum_j x_j^2}\sqrt{\sum_k x_k^2}}$$

Embeddings-based Data Transformation for Text mining

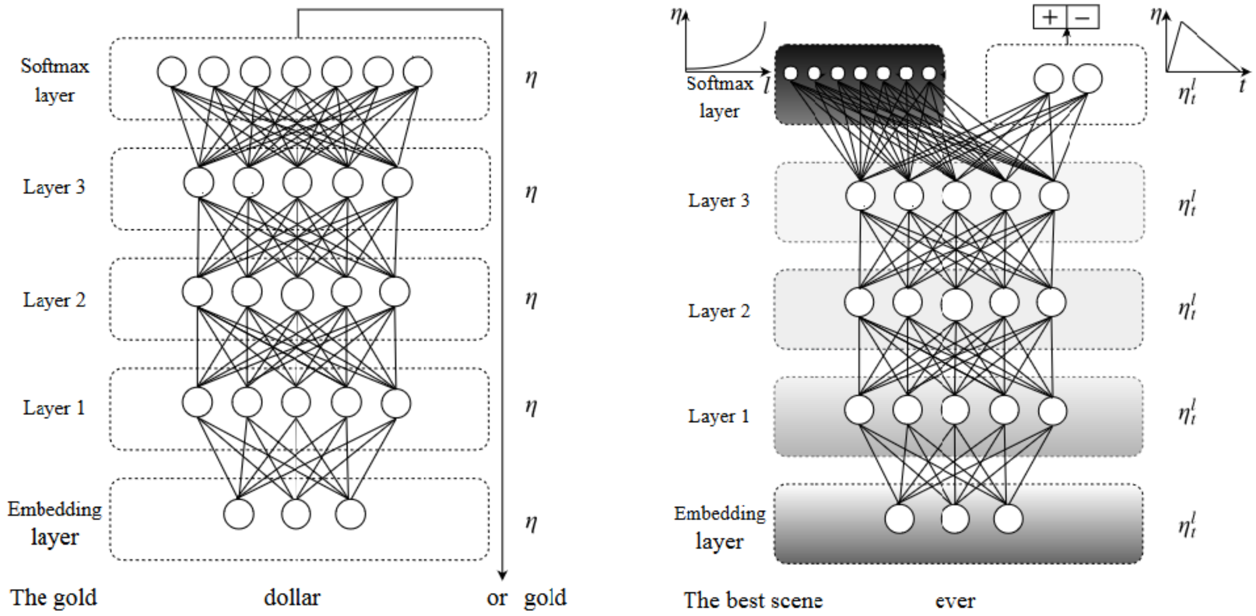
- Corpus embedding, **Document embedding,** Sentence embedding, word embedding (e.g., word2vec)
  - Transforming documents by projecting documents into vectors (rows of a data table)

Document	Dim1	Dim2	…	DimN	Class
d1	0.378	0.222	0.333	0.95	NO
d2	...	...	...	...	YES
d3	...	...	...	...	NO
d4	...	...	...	...	YES
d5	...	...	...	...	NO
d6-d13	...	...	...	...	...
d14	...	...	...	...	YES
d15	...	...	...	...	NO
d16	...	...	...	...	NO
d17	...	...	...	...	NO
d18	...	...	...	...	NO
d19-d23	...	...	...	...	...
d24	0.198	0.523	0.715	0.263	NO

Embeddings-based Data Transformation for Text mining

- Corpus embedding, Document embedding, Sentence embedding, **word embedding** (e.g., word2vec)
  - Transforming documents by projecting documents into vectors (rows of a data table)
  - Table values correspond to weights in the embedding layer of a neural network

Document	Dim1	Dim2	…	DimN	Class
d1	0.378	0.222	0.333	0.95	NO
d2	…	…	…	…	YES
d3	…	…	…	…	NO
d4	…	…	…	…	YES
d5	…	…	…	…	NO
d6-d13	…	…	…	…	…
d14	…	…	…	…	YES
d15	…	…	…	…	NO
d16	…	…	…	…	NO
d17	…	…	…	…	NO
d18	…	…	…	…	NO
d19-d23	…	…	…	…	…
d24	0.198	0.523	0.715	0.263	NO

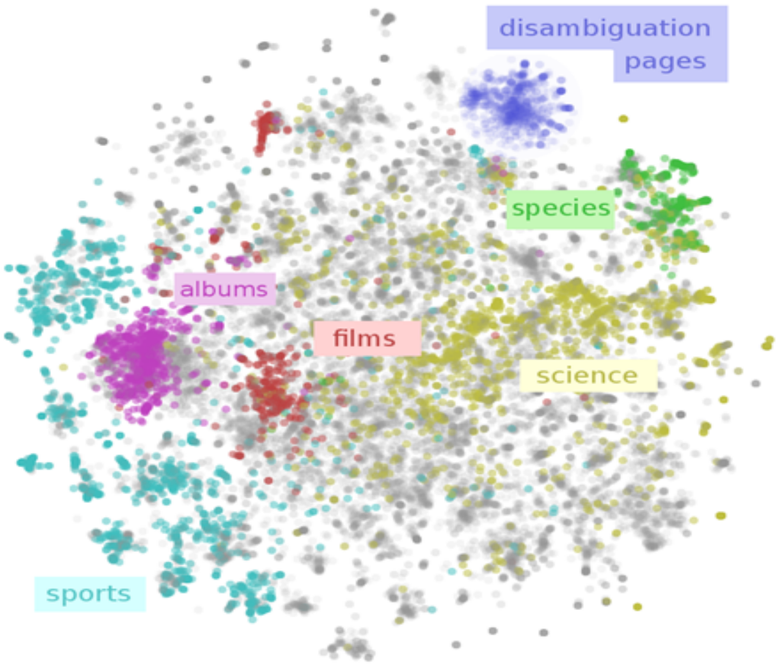


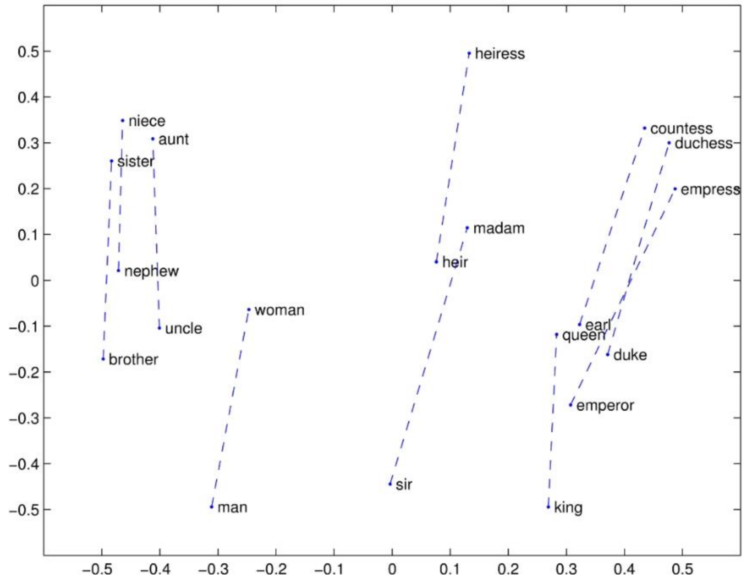
LM pre-training

Classifier fine-tuning

Embedding-based Data Transformation for Text mining

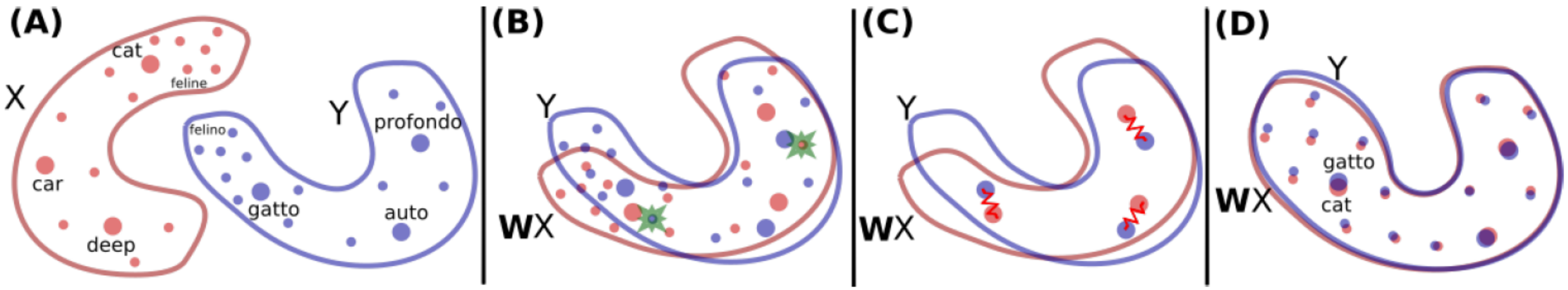
- Corpus embedding, Document embedding, Sentence embedding, **word embedding**, …
  - Representations of word meaning obtained from corpus statistics
  - Spatial relationships correspond to linguistic relationships





Cross-domain or cross-lingual Embeddings-based Data Transformation for Text mining

- Aligning embedding spaces across domains or languages



- **EMBEDDIA** H2020 project (2019-2021) coordinated by Jožef Stefan Institute: **Cross-lingual embeddings for less-represented languages in news media industry**
  - developing new language models for less represented languages
  - using advanced embedding models like GloVe and contextual embedding models like Bert in news analysis applications and in UGC commentary filtering

Course Outline

1. Introduction to machine learning
2. Rule learning
3. Text mining
4. Relational and Semantic machine learning
5. Ensemble learning
6. Support Vector Machines and Kernels
7. Artificial neural networks and deep learning
8. Complex data types and embeddings
9. Autoencoders

Course Outline

1. Introduction to machine learning

2. Rule learning

3. Text mining

4. Relational and Semantic machine learning

5. Ensemble learning

6. Support Vector Machines and Kernels

7. Artificial neural networks and deep learning

8. Complex data types and embeddings

9. Autoencoders