Decision trees

A decision tree is a predictive model which maps observations about an item to conclusions about the item's target value. Another name for such tree models is classification trees. In these tree structures, leaves represent classifications and branches represent conjunctions of attribute-values that lead to those classifications. In decision trees, each interior node corresponds to an attribute; an arc to a child represents a possible value of that attribute. A leaf represents a possible value of target variable given the values of the variables represented by the path from the root.

A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner. The recursion is completed when splitting is either non-feasible, or a singular classification can be applied to each element of the derived subset. In advanced algorithms like C4.5 (J48), other stopping criteria are also used.

Decision tree induction - Algorithm ID3

Given: training set S

1. Compute the entropy E(S) of the set S
2. IF *E(S)* = 0
3. The current set is "clean" and therefore a leaf in our tree
4. IF *E(S)* > 0
5. Compute the information gain of each attribute *Gain(S, A)*
6. The attribute A with the highest information gain becomes the root
7. Divide the set S into subsets S_i according to the values of A
8. Repeat steps 1-7 on each S_i

The information gain of an attribute *Gain(S ,A)* is computed as follows:

number of examples in the subset S_v

(probability of the branch)

set S

atribut A

$$Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot E(S_v)$$

number of examples in set S

The entropy of a set *E(S)* is computed as follows, where p_c are probabilities of each class:

$$E(S) = -\sum_{c=1}^{N} p_c . \log_2 p_c$$

Exercise

Given: Attribute-value data with nominal target variable Lenses.
Induce a decision tree and estimate its performance on new data.

The data:

Person	Age	Prescription	Astigmatic	Tear_Rate	**Lenses**
P1	young	myope	no	normal	**YES**
P2	young	myope	no	reduced	**NO**
P3	young	hypermetrope	no	normal	**YES**
P4	young	hypermetrope	no	reduced	**NO**
P5	young	myope	yes	normal	**YES**
P6	young	myope	yes	reduced	**NO**
P7	young	hypermetrope	yes	normal	**YES**
P8	young	hypermetrope	yes	reduced	**NO**
P9	pre-presbyopic	myope	no	normal	**YES**
P10	pre-presbyopic	myope	no	reduced	**NO**
P11	pre-presbyopic	hypermetrope	no	normal	**YES**
P12	pre-presbyopic	hypermetrope	no	reduced	**NO**
P13	pre-presbyopic	myope	yes	normal	**YES**
P14	pre-presbyopic	myope	yes	reduced	**NO**
P15	pre-presbyopic	hypermetrope	yes	normal	**NO**
P16	pre-presbyopic	hypermetrope	yes	reduced	**NO**
P17	presbyopic	myope	no	normal	**NO**
P18	presbyopic	myope	no	reduced	**NO**
P19	presbyopic	hypermetrope	no	normal	**YES**
P20	presbyopic	hypermetrope	no	reduced	**NO**
P21	presbyopic	myope	yes	normal	**YES**
P22	presbyopic	myope	yes	reduced	**NO**
P23	presbyopic	hypermetrope	yes	normal	**NO**
P24	presbyopic	hypermetrope	yes	reduced	**NO**

We split the data into two parts: one for training and one for testing.

Training set

Person	Age	Prescription	Astigmatic	Tear_Rate	**Lenses**
P1	young	myope	no	normal	**YES**
P2	young	myope	no	reduced	**NO**
P4	young	hypermetrope	no	reduced	**NO**
P5	young	myope	yes	normal	**YES**
P6	young	myope	yes	reduced	**NO**
P7	young	hypermetrope	yes	normal	**YES**
P8	young	hypermetrope	yes	reduced	**NO**
P10	pre-presbyopic	myope	no	reduced	**NO**
P11	pre-presbyopic	hypermetrope	no	normal	**YES**
P14	pre-presbyopic	myope	yes	reduced	**NO**
P17	presbyopic	myope	no	normal	**NO**
P18	presbyopic	myope	no	reduced	**NO**
P19	presbyopic	hypermetrope	no	normal	**YES**
P20	presbyopic	hypermetrope	no	reduced	**NO**
P21	presbyopic	myope	yes	normal	**YES**
P22	presbyopic	myope	yes	reduced	**NO**
P24	presbyopic	hypermetrope	yes	reduced	**NO**

Testing set					
Person	Age	Prescription	Astigmatic	Tear_Rate	**Lenses**
P3	young	hypermetrope	no	normal	**YES**
P9	pre-presbyopic	myope	no	normal	**YES**
P12	pre-presbyopic	hypermetrope	no	reduced	**NO**
P13	pre-presbyopic	myope	yes	normal	**YES**
P15	pre-presbyopic	hypermetrope	yes	normal	**NO**
P16	pre-presbyopic	hypermetrope	yes	reduced	**NO**
P23	presbyopic	hypermetrope	yes	normal	**NO**

We induce a decision tree on the training set S according to the algorithm ID3.
Compute the entropy E(S) of the set S:

Training set					
Person	Age	Prescription	Astigmatic	Tear_Rate	**Lenses**
P1	young	myope	no	normal	**YES**
P2	young	myope	no	reduced	**NO**
P4	young	hypermetrope	no	reduced	**NO**
P5	young	myope	yes	normal	**YES**
P6	young	myope	yes	reduced	**NO**
P7	young	hypermetrope	yes	normal	**YES**
P8	young	hypermetrope	yes	reduced	**NO**
P10	pre-presbyopic	myope	no	reduced	**NO**
P11	pre-presbyopic	hypermetrope	no	normal	**YES**
P14	pre-presbyopic	myope	yes	reduced	**NO**
P17	presbyopic	myope	no	normal	**NO**
P18	presbyopic	myope	no	reduced	**NO**
P19	presbyopic	hypermetrope	no	normal	**YES**
P20	presbyopic	hypermetrope	no	reduced	**NO**
P21	presbyopic	myope	yes	normal	**YES**
P22	presbyopic	myope	yes	reduced	**NO**
P24	presbyopic	hypermetrope	yes	reduced	**NO**

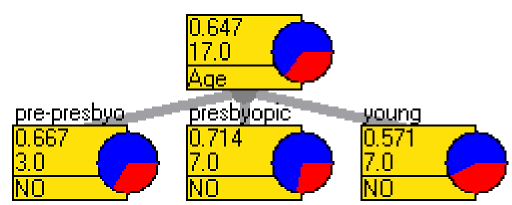
There are 17 examples in our training set. 6 of them have value Lenses=YES and 11 of them have the value Lenses=NO.

E(S) = E(6/17, 11/17) = 0.94

Since the entropy E(S) is not zero, we compute the information gain of each attribute: Gain(S, A).

Information gain of the attribute Age on set S:

Training set					
Person	Age	Prescription	Astigmatic	Tear_Rate	**Lenses**
P1	young	myope	no	normal	**YES**
P2	young	myope	no	reduced	**NO**
P4	young	hypermetrope	no	reduced	**NO**
P5	young	myope	yes	normal	**YES**
P6	young	myope	yes	reduced	**NO**
P7	young	hypermetrope	yes	normal	**YES**
P8	young	hypermetrope	yes	reduced	**NO**
P10	pre-presbyopic	myope	no	reduced	**NO**
P11	pre-presbyopic	hypermetrope	no	normal	**YES**
P14	pre-presbyopic	myope	yes	reduced	**NO**
P17	presbyopic	myope	no	normal	**NO**
P18	presbyopic	myope	no	reduced	**NO**
P19	presbyopic	hypermetrope	no	normal	**YES**
P20	presbyopic	hypermetrope	no	reduced	**NO**
P21	presbyopic	myope	yes	normal	**YES**
P22	presbyopic	myope	yes	reduced	**NO**
P24	presbyopic	hypermetrope	yes	reduced	**NO**

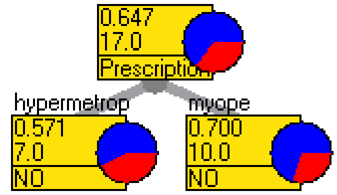


The attribute Age splits the set S into three subsets: Age=young, Age=pre-presbyopic and Age=presbyopic with 7, 3 and 7 instances respectively.
In the subset Age = young, there are 3 items with Lenses=YES and 4 with Lenses=NO.
E(Age=young) = E(3/7, 4/7) = 0.99.
Similar for the other two sets:
E(Age=pre-presbyopic) = E(1/3, 2/3) = 0.92
E(Age=presbyopic) = E(2/7, 5/7) = 0.86

Gain (S,Age) =
E(S) – 7/17 E(Age= young) – 3/17 E(Age=pre-presbyopic) – 7/17 E(Age=presbyopic) =
= 0.94 – 7/17 * 0.99 – 3/17 *0.92 – 7/17 *0.86 = 0.02

Information gain of the attribute Prescription on set S:

Training set					
Person	Age	Prescription	Astigmatic	Tear_Rate	Lenses
P1	young	myope	no	normal	YES
P2	young	myope	no	reduced	NO
P4	young	hypermetrope	no	reduced	NO
P5	young	myope	yes	normal	YES
P6	young	myope	yes	reduced	NO
P7	young	hypermetrope	yes	normal	YES
P8	young	hypermetrope	yes	reduced	NO
P10	pre-presbyopic	myope	no	reduced	NO
P11	pre-presbyopic	hypermetrope	no	normal	YES
P14	pre-presbyopic	myope	yes	reduced	NO
P17	presbyopic	myope	no	normal	NO
P18	presbyopic	myope	no	reduced	NO
P19	presbyopic	hypermetrope	no	normal	YES
P20	presbyopic	hypermetrope	no	reduced	NO
P21	presbyopic	myope	yes	normal	YES
P22	presbyopic	myope	yes	reduced	NO
P24	presbyopic	hypermetrope	yes	reduced	NO



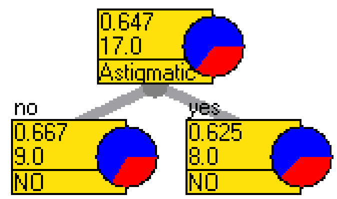
E(Prescription=hypermetrope) =
= E(3/7, 4/7) = 0.99

E(Prescription=myope) =
= E(3/10, 7/10) = 0.88

Gain (S, Prescription) =
= E(S) –
– 7/17 E(Prescription=hypermetrope)
– 10/17 E(Prescription=myope) =
= 0.94 – 7/17* 0.99 – 10/17 * 0.88 = 0.02

Information gain of the attribute Astigmatic on set S:

Training set					
Person	Age	Prescription	Astigmatic	Tear_Rate	Lenses
P1	young	myope	no	normal	YES
P2	young	myope	no	reduced	NO
P4	young	hypermetrope	no	reduced	NO
P5	young	myope	yes	normal	YES
P6	young	myope	yes	reduced	NO
P7	young	hypermetrope	yes	normal	YES
P8	young	hypermetrope	yes	reduced	NO
P10	pre-presbyopic	myope	no	reduced	NO
P11	pre-presbyopic	hypermetrope	no	normal	YES
P14	pre-presbyopic	myope	yes	reduced	NO
P17	presbyopic	myope	no	normal	NO
P18	presbyopic	myope	no	reduced	NO
P19	presbyopic	hypermetrope	no	normal	YES
P20	presbyopic	hypermetrope	no	reduced	NO
P21	presbyopic	myope	yes	normal	YES
P22	presbyopic	myope	yes	reduced	NO
P24	presbyopic	hypermetrope	yes	reduced	NO



E(Astigmatic=no) = E(3/9, 6/9) = 0.92

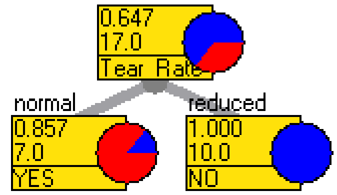
E(Astigmatic =yes) = E(3/8, 5/8) = 0.95

Gain (S, Astigmatic) =
= E(S) –
– 9/17 E(Astigmatic=no)
– 8/17 E(Astigmatic=yes) =
= 0.94 – 9/17* 0.92 – 8/17 * 0.95 = 0.006

Information gain of the attribute Tear_Rate on set S:

Training set

Person	Age	Prescription	Astigmatic	Tear_Rate	Lenses
P1	young	myope	no	normal	YES
P2	young	myope	no	reduced	NO
P4	young	hypermetrope	no	reduced	NO
P5	young	myope	yes	normal	YES
P6	young	myope	yes	reduced	NO
P7	young	hypermetrope	yes	normal	YES
P8	young	hypermetrope	yes	reduced	NO
P10	pre-presbyopic	myope	no	reduced	NO
P11	pre-presbyopic	hypermetrope	no	normal	YES
P14	pre-presbyopic	myope	yes	reduced	NO
P17	presbyopic	myope	no	normal	NO
P18	presbyopic	myope	no	reduced	NO
P19	presbyopic	hypermetrope	no	normal	YES
P20	presbyopic	hypermetrope	no	reduced	NO
P21	presbyopic	myope	yes	normal	YES
P22	presbyopic	myope	yes	reduced	NO
P24	presbyopic	hypermetrope	yes	reduced	NO

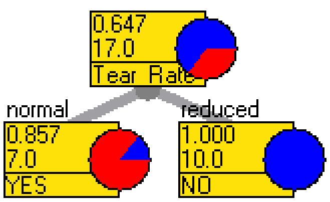


E(Tear_Rate=normal) = E(6/7, 1/7) = 0.59

E(Tear_Rate=reduced) = E(0/10, 10/10) = 0

Gain (S, Tear_Rate) =
= E(S) –
– 7/17 E(Tear_Rate=normal)
– 10/17 E(Tear_Rate=reduced) =
= 0.94 – 7/17* 0.59 – 10/17 * 0 = 0.70

The attribute with the highest information gain is Tear_Rate with information gain of 0.70. This attribute is chosen to become the root of our tree. We recursively continue to build the tree on subsets of set S according to values of the attribute Tear_Rate.

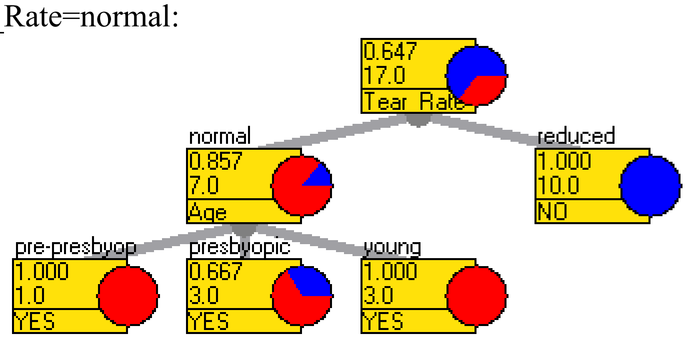


On the one hand, the entropy of the subset with Tear_Rate=normal is not zero, therefore we continue with the built. On the other hand, the entropy of the set Tear_Rate=reduced is zero, which means that the algorithm has reached the end and this node is a leaf of the tree. It classifies into class Lenses=NO.

Information gain of the attribute Age on set Tear_Rate=normal:

Training set

Person	Age	Prescription	Astigmatic	Lenses
P1	young	myope	no	YES
P5	young	myope	yes	YES
P7	young	hypermetrope	yes	YES
P11	pre-presbyopic	hypermetrope	no	YES
P17	presbyopic	myope	no	NO
P19	presbyopic	hypermetrope	no	YES
P21	presbyopic	myope	yes	YES



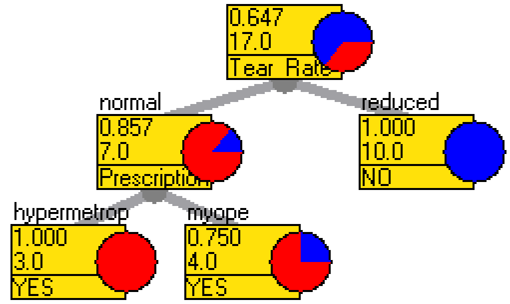
E(Age=young | Tear_Rate=normal) = E(3/3, 0/3) = 0
E(Age=pre-presbyopic | Tear_Rate=normal) = E(1/1, 0/1) = 0
E(Age=presbyopic | Tear_Rate=normal) = E(2/3, 1/3) = 0.92

Gain (S Tear_Rate=normal, Age) =
E(S Tear_Rate=normal) – 3/7 E(Age=young | Tear_Rate=normal)
– 1/7 E(Age=pre-presbyopic | Tear_Rate=normal)
– 3/7 E(Age=presbyopic | Tear_Rate=normal) =
= 0.59 – 3/7 * 0 - 1/7 * 0 - 3/7 *0.92 = 0.20

Information gain of the attribute Prescription on set Tear_Rate=normal:

Training set

Person	Age	Prescription	Astigmatic	**Lenses**
P1	young	myope	no	**YES**
P5	young	myope	yes	**YES**
P7	young	hypermetrope	yes	**YES**
P11	pre-presbyopic	hypermetrope	no	**YES**
P17	presbyopic	myope	no	**NO**
P19	presbyopic	hypermetrope	no	**YES**
P21	presbyopic	myope	yes	**YES**



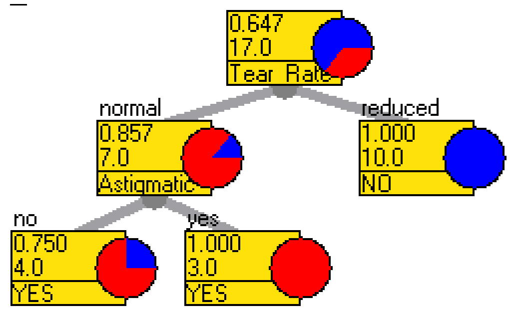
E(Prescription=myope | Tear_Rate=normal) = E(3/4, 1/4) = 0.81
E(Prescription=hypermetropy | Tear_Rate=normal) = E(3/3, 0/3) = 0

Gain (S Tear_Rate=normal, Prescription) =
E(S Tear_Rate=normal) – 4/7 E(Prescription=myope | Tear_Rate=normal)
– 3/7 E(Prescription=hypermetropy | Tear_Rate=normal) =
= 0.59 – 4/7 * 0.81 - 3/7 * 0 = 0.13

Information gain of the attribute Astigmatic on set Tear_Rate=normal:

Training set

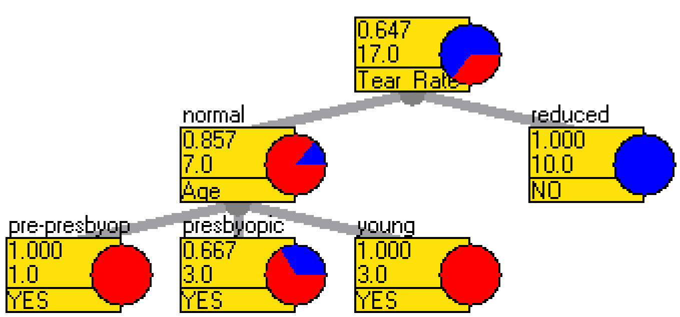
Person	Age	Prescription	Astigmatic	**Lenses**
P1	young	myope	no	**YES**
P5	young	myope	yes	**YES**
P7	young	hypermetrope	yes	**YES**
P11	pre-presbyopic	hypermetrope	no	**YES**
P17	presbyopic	myope	no	**NO**
P19	presbyopic	hypermetrope	no	**YES**
P21	presbyopic	myope	yes	**YES**



E(Astigmatic=no | Tear_Rate=normal) = E(3/4, 1/4) = 0.81
E(Astigmatic=yes | Tear_Rate=normal) = E(3/3, 0/3) = 0

Gain (S Tear_Rate=normal, Astigmatic) =
E(S Tear_Rate=normal) – 4/7 E(Astigmatic=no | Tear_Rate=normal)
– 3/7 E(Astigmatic=yes | Tear_Rate=normal) =
= 0.59 – 4/7 * 0.81 - 3/7 * 0 = 0.13

The attribute with the highest information gain on set Tear_Rate=normal is Age with information gain of 0.20. This attribute is chosen to become the next node of our tree. We recursively continue to build the tree on subsets of this set according to values of the attribute Age.

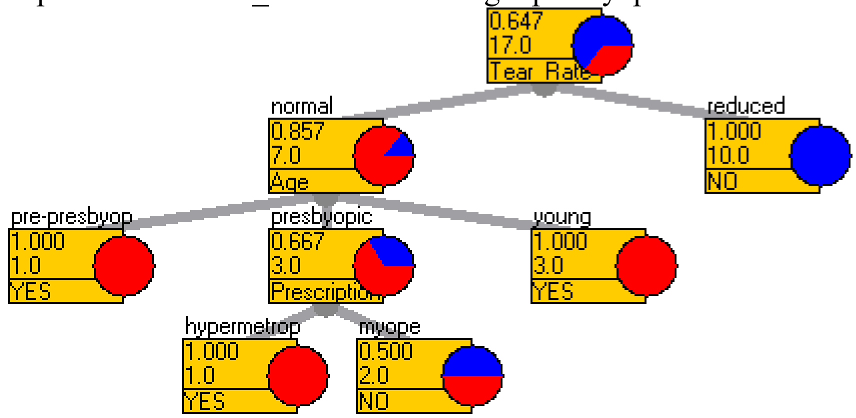


On the one hand, the entropies of the subsets with Age=pre-presbyopic and Age=young are zero, therefore we reached the end of the tree. On the other hand, the entropy of the set Age=presbyopic is not zero, which means that the algorithm continues with the built.

Information gain of the attribute Prescription on set Tear_Rate=normal&Age=presbyopic:

Training set

Person	Prescription	Astigmatic	Lenses
P17	myope	no	**NO**
P19	hypermetrope	no	**YES**
P21	myope	yes	**YES**



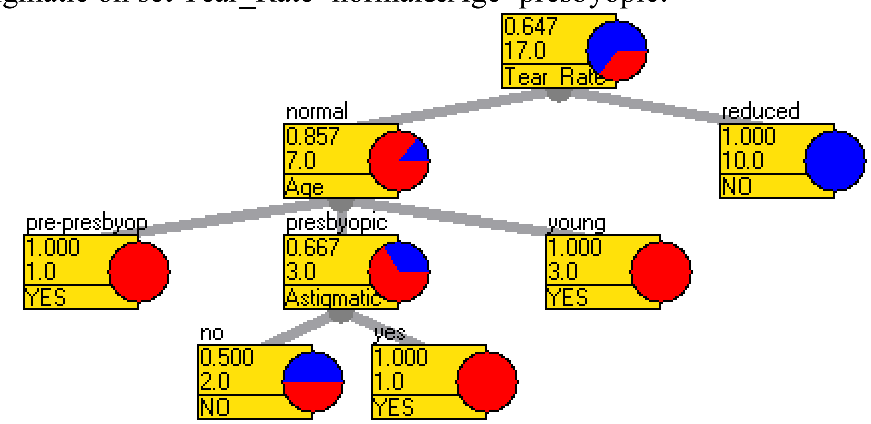
E(Prescription=myope | Tear_Rate=normal&Age=presbyopic) = E(1/2, 1/2) = 1
E(Prescription=hypermetropy | Tear_Rate=normal&Age=presbyopic) = E(1/1, 0/1) = 0

Gain (S Tear_Rate=normal&Age=presbyopic, Prescription) =
E(S Tear_Rate=normal&Age=presbyopic)
– 2/3 E(Prescription=myope | Tear_Rate=normal&Age=presbyopic)
– 1/3 E(Prescription=hypermetropy | Tear_Rate=normal&Age=presbyopic) =
= 0.92 – 2/3 *1 – 1/3 * 0 = 0.25

Information gain of the attribute Astigmatic on set Tear_Rate=normal&Age=presbyopic:

Training set

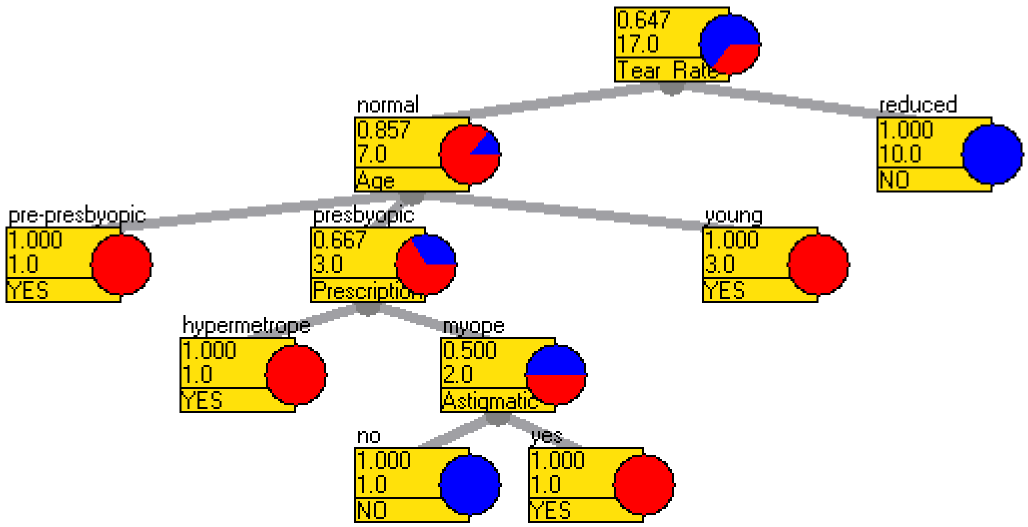
Person	Prescription	Astigmatic	Lenses
P17	myope	no	**NO**
P19	hypermetrope	no	**YES**
P21	myope	yes	**YES**



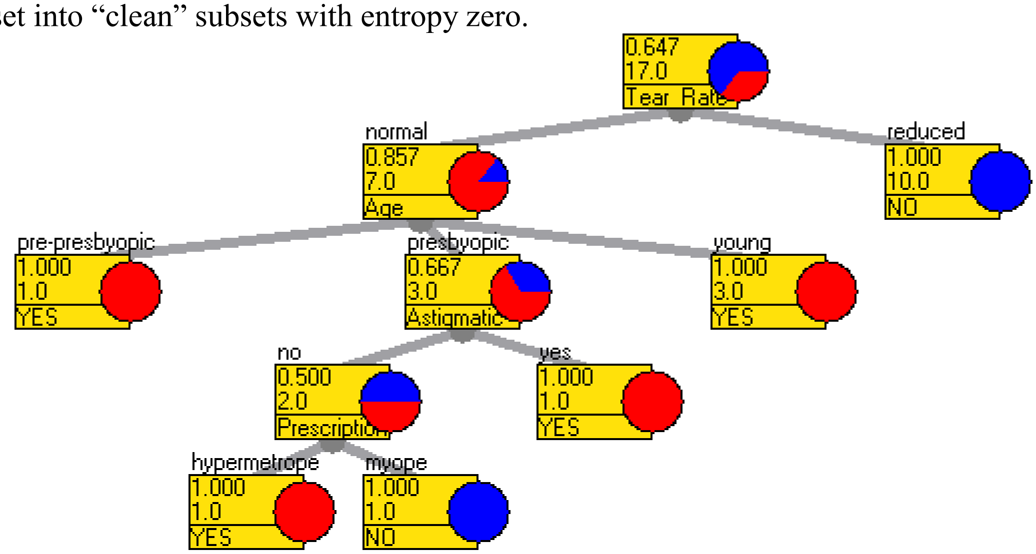
E(Astigmatic=no | Tear_Rate=normal&Age=presbyopic) = E(1/2, 1/2) = 1
E(Astigmatic=yes | Tear_Rate=normal&Age=presbyopic) = E(1/1, 0/1) = 0

Gain (S Tear_Rate=normal&Age=presbyopic, Prescription) =
E(S Tear_Rate=normal&Age=presbyopic)
– 2/3 E(Astigmatic=no | Tear_Rate=normal&Age=presbyopic)
– 1/3 E(Astigmatic=yes | Tear_Rate=normal&Age=presbyopic)=
= 0.92 – 2/3 *1 – 1/3 * 0 = 0.25

Both attributes Prescription and Astigmatic have the same information gain of 0.25. The ID3 algorithm would choose one of them for the next node (implementations usually take the first one). If we choose the attribute Prescription, the only remaining attribute is Astigmatic, which finally splits the dataset into “clean” subsets with entropy zero.



If we choose the attribute Astigmatic, the only remaining attribute is Prescription, which also splits the dataset into "clean" subsets with entropy zero.



We use the former tree and test its performance on the testing set.

Testing set					REAL	PREDICTED
Person	Age	Prescription	Astigmatic	Tear_Rate	**Lenses**	**Lenses**
P3	young	hypermetrope	no	normal	**YES**	**YES**
P9	pre-presbyopic	myope	no	normal	**YES**	**YES**
P12	pre-presbyopic	hypermetrope	no	reduced	**NO**	**NO**
P13	pre-presbyopic	myope	yes	normal	**YES**	**YES**
P15	pre-presbyopic	hypermetrope	yes	normal	**NO**	**YES**
P16	pre-presbyopic	hypermetrope	yes	reduced	**NO**	**NO**
P23	presbyopic	hypermetrope	yes	normal	**NO**	**YES**

Confusion matrix

		predicted	
		Lenses=YES	Lenses=NO
actual	Lenses=YES	TP =3	FN=0
	Lenses=NO	FP=2	TN=2

Classification accuracy is

CA = (TP + TN)/(TP+TN+FP+FN)

= 5 / 7

= 0.71

Petra Kralj Novak
Petra.Kralj.Novak@ijs.si
http://kt.ijs.si/petra_kralj/dmkd.html