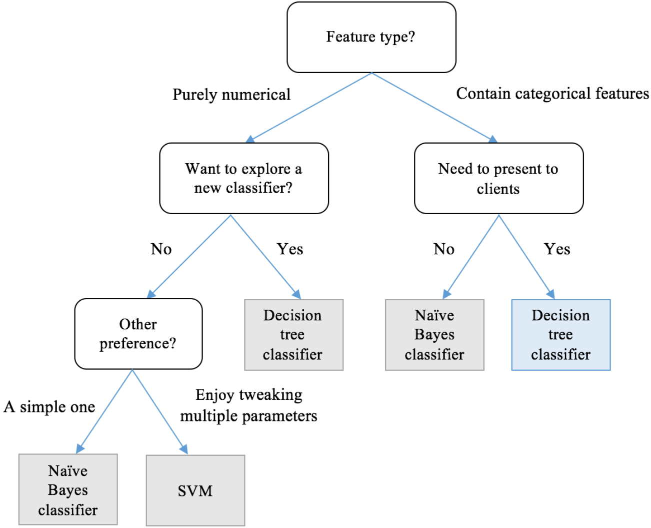
Rozpoznávanie obrazcov - 10. cvičenie
Decision trees

Viktor Kocur
viktor.kocur@fmph.uniba.sk

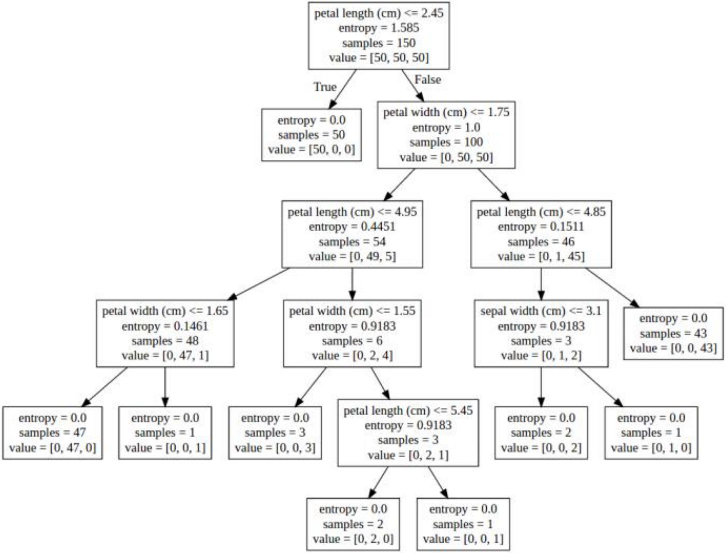
DAI FMFI UK

27.4.2020

Decision tree



Decision tree



petal length (cm) <= 2.45
entropy = 1.585
samples = 150
value = [50, 50, 50]

True

False

entropy = 0.0
samples = 50
value = [50, 0, 0]

petal width (cm) <= 1.75
entropy = 1.0
samples = 100
value = [0, 50, 50]

petal length (cm) <= 4.95
entropy = 0.4451
samples = 54
value = [0, 49, 5]

petal length (cm) <= 4.85
entropy = 0.1511
samples = 46
value = [0, 1, 45]

petal width (cm) <= 1.65
entropy = 0.1461
samples = 48
value = [0, 47, 1]

petal width (cm) <= 1.55
entropy = 0.9183
samples = 6
value = [0, 2, 4]

sepal width (cm) <= 3.1
entropy = 0.9183
samples = 3
value = [0, 1, 2]

entropy = 0.0
samples = 43
value = [0, 0, 43]

entropy = 0.0
samples = 47
value = [0, 47, 0]

entropy = 0.0
samples = 1
value = [0, 0, 1]

entropy = 0.0
samples = 3
value = [0, 0, 3]

petal length (cm) <= 5.45
entropy = 0.9183
samples = 3
value = [0, 2, 1]

entropy = 0.0
samples = 2
value = [0, 0, 2]

entropy = 0.0
samples = 1
value = [0, 1, 0]

entropy = 0.0
samples = 2
value = [0, 2, 0]

entropy = 0.0
samples = 1
value = [0, 0, 1]

Constructing trees

Splitting rules

The tree is constructed by selecting a feature and a value based on which we split the set of elements into two parts. This process is repeated with both subsets until some stopping criterion is fulfilled.

Stopping criterion

Examples: each subset contains only one class, the tree reach a certain depth, fewer misclassifications than a certain thresholds, next best feature for selection is worse than some threshold.

á

ID3

We choose a feature with lowest entropy, e.g. a feature for which the information gain is the highest (mutual information with classes is the highest).

C4.5

Similar to ID3, but this time we optimize for highest normalized information gain. C4.5 can also work with numerical data.

Splitting rules - 4th lab theory

Entropy

$$H(Y) = \sum_{y \in \omega} -P(Y = y) \cdot log_2(P(Y = y))$$

Specific conditional entropy

$$H(Y|X = v) = H(Y), \text{len pre hodnoty} Y, \text{kde } X = x$$

Splitting rules - 4th lab theory

Mutual information, information gain

$$I(Y;X) = H(Y) - H(Y|X) = H(Y) - \sum_{x \in \omega} P(X = x) \cdot H(Y|X = x)$$

Normalized information gain

$$nI(Y;X) = \frac{I(Y;X)}{H(X)}$$

Examples

ID3

https://sefiks.com/2017/11/20/a-step-by-step-id3-decision-tree-example/

C4.5

https://sefiks.com/2018/05/13/a-step-by-step-c4-5-decision-tree-example/

Matlab

fitctree

Mdl = fitctree(X,y) - returns a tree classifier.

fitctree

Mdl = fitctree(T,property) - returns a tree classifier for table T and classification target in the property column of the table.

CART

Matlab uses the CART algorithm which is similar to ID3, but slightly different. It is not a part of the lecture so we will not deal with it now.

Matlab

predict

Mdl.predict(x) - returns model prediction

view

Mdl.view('Mode','graph') - displays the tree

Exercise

Create and display a tree for the fisheriris and census1994 database.

Pruning the trees

Pruning

The tree can be too complex which leads to overfitting. It is possible to prune the tree so that its subtrees which only provide marginal benefits are converted to leafs.

prune

MdlP = prune(Mdl,'Property', value) - returns a pruned tree based on the selected property.

Exercise

Prune the tree for the data in fisheriris and census 1994. Test various properties. Check if pruning helps the accuracy on the test set of census1994.