RWTH Aachen
Lehrstuhl für Informatik 9
Prof. Dr. T. Seidl

January 11th, 2016

Exercise 9 for the lecture
Data Mining Algorithms
WS 2015/2016

Hand in your solutions on January 18th <u>before</u> the lecture. The tutorial for this exercise will be held on January 22nd.

Exercise 9.1) Decision Tree Pruning **2 points**

Consider the decision tree from the lecture (slide 83). Perform "Reduced-Error Pruning" using the following "test set":

no.	forecast	temperature	humidity	wind	class
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	Yes
3	overcast	hot	high	weak	Yes
4	rain	mild	high	weak	Yes
5	rain	cool	normal	weak	Yes
6	rain	cool	normal	strong	No
7	overcast	cool	normal	strong	Yes
8	sunny	mild	high	weak	Yes
9	sunny	cool	normal	weak	Yes
10	rain	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rain	mild	high	strong	No

Exercise 9.4) NN-Classification **3 points**

Consider the following data set:

P1=(0,4), P2=(0,6), P3=(1,5), P4=(2,1), P5=(2,4), P6=(3,0), P7=(4,5),
P8=(1,3), P9=(3,3), P10=(3,6), P11=(4,4), P12=(5,2)
Two classes: red = {P1, …, P7}, black = {P8, …, P12}.

Euclidean distance and the majority voting criteria is employed to determine the class.
Use the leave-one-out strategy to determine the classification accuracy of the 3-NN classifier.

Exercise 9.3) Bayes Classification **5=4+1 points**

Use the naïve Bayes classifier approach to classify the following data. The dataset given contains two categorical attributes X and Y. The X attribute consists of three possible values i, k and l while the Y attribute has a domain of four different values i, k, l and m. Two classes are distinguished (C_1 and C_2).

a) Use the following dataset {class, X,Y} to train your naïve Bayesian classifier:

   (C_1, i, i); (C_1, k, k); (C_1, i, k); (C_1, k, l); (C_2, i, l);
   (C_2, k, m); (C_2, k, k); (C_2, l, m); (C_2, l, l);

Illustrate the conditional probability distributions for each class by specifying the corresponding histograms.

b) Use the classifier from above to predict a class label for the following two data points:
(?, i, k), (?, k, m)

Exercise 9.4) Conditional Probabilities **5=2+1+2 points**

Let the sample space Ω be a finite set together with a probability measure P. Furthermore consider the event $A \subseteq \Omega$ and the events $B_1, B_2, B_3 \subset \Omega$ that form a partition of Ω (i.e. they are disjoint and their union $B_1 \cup B_2 \cup B_3 = \Omega$ is equal to the set of all possible outcomes).
Give detailed **proofs** or **counterexamples** for the following statements:

a) $P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$.
b) $P(A|B_1) + P(A|B_2) + P(A|B_3) = 1$.
c) $P(B_1|A) + P(B_2|A) + P(B_3|A) = 1$.

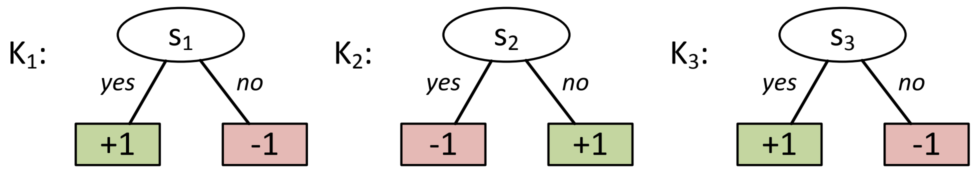
Exercise 9.5) Ensemble Classification and Decision Stumps **5=2+3 points**

Consider the following data set having the attributes s_1, s_2, s_3 and s_4:

ID	s_1	s_2	s_3	s_4
x_1	Yes	No	Yes	No
x_2	No	Yes	Yes	No
x_3	Yes	No	No	Yes
x_4	No	No	No	Yes
x_5	Yes	Yes	Yes	Yes
x_6	No	Yes	No	Yes

A decision stump is a decision tree consisting only of a root node and leaf nodes and thus makes a decision based on just a single attribute.

a) Given the following decision stumps K_1, K_2, K_3 and the weights $\alpha_1 = 0.5, \alpha_2 = 0.3, \alpha_3 = 0.2$, classify the data points $x_1, \dots, x_6$ into the classes $\{+1, -1\}$ using the boosting decision rule.



b) Given the data set above and the class labels $y_1 = y_3 = y_4 = y_5 = +1$ and $y_2 = y_6 = -1$ (and no classifiers yet), perform M=3 steps of the AdaBoost algorithm. In each step, choose as classifier K_m a decision stump that minimizes the error function J_m. (If there are several possibilities minimizing the error function, always choose the one using the attribute with the smallest ID).
After learning the classifiers and weights with AdaBoost, classify the points $x_1, \dots, x_6$ using the final model.

Note: You don't have to normalize the weights, thus the constraint $\sum_{i=1}^{M} \alpha_i = 1$ does not have to be fulfilled here.