Machine Learning Exercise Sheet 5

Prof. Dr. Lars Schmidt-Thieme, Martin Wistuba Information Systems and Machine Learning Lab University of Hildesheim

November 18th, 2014 Submission until November 25th, 13.00 to wistuba@ismll.de

Exercise 17: Naive Bayes (5 Points)

Given is the following training data:

	Color	Туре	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Blue	Sports	Domestic	Yes
4	Blue	Sports	Domestic	No
5	Blue	Sports	Imported	Yes
6	Blue	Grand tourer	Imported	No
7	Blue	Grand tourer	Imported	Yes
8	Blue	Grand tourer	Domestic	No
9	Red	Grand tourer	Imported	No
10	Red	Sports	Imported	Yes

Calculate the probabilities

P(Red|Yes), P(Grand tourer|Yes), P(Domestic|Yes), P(Red|No), P(Grand tourer|No), P(Domestic|No)

Calculate q_{yes} and q_{no} . Estimate whether a car with $X_1 = \text{Red}, X_2 = \text{Grand tourer}, X_3 = \text{Domestic will}$ be stolen.

Exercise 18: Bayesian Network, D-Separation (5 Points)

a) Construct the graph representing the Bayesian Network for the variables $\{X_1, X_2, X_3, X_4\}$ with $X_1 \perp X_4 | X_2 X_3 \perp X_4 | X_1, X_2$

x	y	x	y
-4	4	3	3
-4	3	4	3
-3	3	4	4
-4	-3	3	-3
-4	-4	4	-4
-3	-3	4	-3
3	3	4	-4

Tabelle 1: Data to be clustered in Exercise 19.

b) Given is the following DAG and $A = \{1\}$ and $B = \{7\}$.



- Are A and B d-separated by $Z = \{2\}$?
- Are A and B d-separated by $Z = \{3\}$?

Exercise 19: (5 Points)

- a) Explain the optimization function of the K-Means clustering in your own words.
- b) What is the primary difference between K-Means and K-Medoids?

c) Apply the K-Means Clustering for two iterations on the data in Table 1 for k = 4. The first cluster center was randomly chosen to be (3,3).

Exercise 20: (5 Points)

- a) What do Gaussian Mixture Models and K-Means have in common? What are the differences.
- b) In the lecture it was mentioned that K-Means are a special case of Gaussian Mixture Models. Explain.