Machine Learning
Exercise Sheet 10

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

January 16th, 2017
Submission until January 23rd, 13.00 via learnweb!

Exercise 19: Naive Bayes - Theory (10 Points)

Naive Bayes estimates the conditional probability of a label Y given its features $X = (X_1, ..., X_n)$ as:

$$P(Y|X_1, ..., X_n) = \frac{1}{Z} P(Y) \prod_{i=1}^{n} P(X_i|Y)$$

where Z is a normalization constant that does not depend on Y. Using the independencies

$$P(X_i|Y, X_{\neg i}) = P(X_i|Y) \quad \forall\, i$$

and the chain-rule of probabilities, show how the Naive Bayes Classifier can be derived. What are the benefits, what are the downsides of using Naive Bayes models?

Exercise 20: Naive Bayes - Application (10 Points)

Given is the following training data:

Car	Color	Type	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	Yes
10	Red	Sports	Imported	Yes

Estimate the following probabilities:

$$P\,(\text{Yes})\,, \quad P\,(\text{Red}|\text{Yes})\,, \quad P\,(\text{Grand tourer}|\text{Yes})\,, \quad P\,(\text{Domestic}|\text{Yes})\,,$$
$$P\,(\text{No})\,, \quad P\,(\text{Red}|\text{No})\,, \quad P\,(\text{Grand tourer}|\text{No})\,, \quad P\,(\text{Domestic}|\text{No})$$

from the data set, by computing the relative frequencies. Then, predict the probability that a car with properties $X_1 = \text{Red}$, $X_2 = \text{Grand tourer}$, $X_3 = \text{Domestic}$ will be stolen.