Machine Learning
Exercise Sheet 10

Prof. Dr. Dr. Lars Schmidt-Thieme, Rafael Rego Drumond
HiWi's: Khouloud Sallami, Maurício Camargo, Shayan Jawed, Sami Diaf
Information Systems and Machine Learning Lab
University of Hildesheim

December 22th, 2017
Submission until January 11th, 2018(noon) by dropping at MACHINE LEARNING postbox
(please indicate in which tutorial are you participating!)
The postboxes is located inside the Samelsonplatz building to the right.

Exercise 19: Basic Network Topologies (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 (Yes), P (Red|Yes), P (Grand tourer|Yes), P (Domestic|Yes), P (No), P (Red|No),

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

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