

Machine Learning

Exercise Sheet 7

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

December 8th, 2014
Submission until December 15th, 13.00 to wistuba@ismll.de

Exercise 13: Naive Bayes (6 Points)

a) Given is an imbalanced data set for binary classification with 100,000 instances, only 10 labeled as positive, all remaining as negative. Our learned classifier is able to classify 99.99% of the instances correctly. What is the probability that an instance that is classified as positive is actually negative?

b) Given is the following training data:

	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

Calculate the probabilities

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

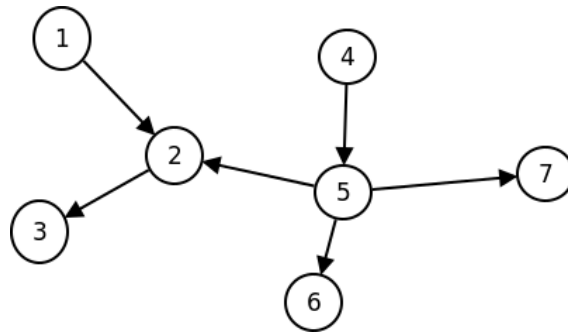
Predict the probability that a car with properties $X_1 = \text{Red}$, $X_2 = \text{Grand tourer}$, $X_3 = \text{Domestic}$ will be stolen.

Exercise 14: Bayesian Network, D-Separation (4 Points)

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

Hint: Try to apply the chain rule such that you can use the conditional independence between variables.

c) 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 $Z = \{3\}$?

Explain your decision.