MTAT.03.227 Machine Learning
Practice session 2
K-nearest neighbors and Naïve Bayes

September 16-20, 2019

Exercise 0. Practice Slides

Go through the practice material provided here and solve the following tasks.

Exercise 1. K-nearest Neighbors

Consider the same classification task as in practice 1, however this time solve it using *K-nearest neighbors algorithm*. As before, we are given data about 5 cats and 5 dogs. The cat training instances are $\{(4,2),(5,1),(5,2),(5,3),(6,2)\}$, and dog training instances are $\{(7,4),(11,8),(11,10),(13,8),(13,10)\}$ where the first element of the pairs correspond to the weight (in unit of *pettygrams*) and second element to the height (in unit of *pettymeters*).

(a) Consider a new instance at (6.0,3.5). Based on KNN, what would be the class with $K = \{1,2,3,5\}$

(b) Between odd and even values for K which one do you think should be preferred and Why?

(c) Visualize the decision boundary for K=1. Optionally, you can also try $K = \{2,3,5\}$ in case you have implemented the decision boundary visualization in Python.

Exercise 2. Naive Bayes (Optional)

Complete the tasks based on table on the left (observations).

row	Tartu's weather	Mood
1	Overcast	Hea[1]
2	Rain	Hea
3	Rain	Halb [2]
4	Sunny	Hea
5	Sunny	Hea
6	Overcast	Halb
7	Overcast	Hea
8	Rain	Halb
9	Sunny	Hea
10	Overcast	Halb
11	Sunny	Halb
12	Rain	Hea
13	Sunny	Hea
14	Sunny	Halb
15	Rain	Halb
16	Rain	Halb

Frequency table

Weather \ Class	Hea	Halb	
Overcast			Marginal
Rain		4	
Sunny			$= \frac{4}{16} = 0.25$
Total	8		
	$\frac{8}{16} = 0.5$		
	Marginal		

Likelihood table – $P(weather \mid mood)$

Weather \ Class	Hea	Halb
Overcast		
Rain		$\frac{4}{8} = 0.5$
Sunny		

[1] Hea in Estonian means good
[2] Halb in Estonian means bad

(a) Complete Frequency and Likelihood tables

(b) What is Naive Bayes Prediction for the mood when weather is Overcast, Rain, and Sunny respectively? Perform the necessary calculations.