

Machine Learning

Exercise Sheet 7

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

December 10th, 2013
Submission until December 17th, 13.00 to wistuba@ismll.de

Exercise 16: Distance Metrics (5 Points)

Given are 12 cities with following coordinates:

City i	x_i	y_i
1	11	5
2	6	4
3	4	10
4	4	2
5	2	4
6	7	7
7	8	8
8	9	2
9	5	7
10	7	1
11	1	6
12	11	11

The distance between the city a with coordinates (a_1, a_2) and the city b with coordinates (b_1, b_2) is defined by:

$$d(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

- a) Estimate the distance matrix $\mathbf{D} = (d(a, b))$ for the 12 cities.
- a) Is the distance metric a Minkowski Metric?
- c) A task in bio informatics is to compare DNA sequences. A usual task is to compare two sequences with respect to its edit distance to check if they are similar. Execute this task and estimate the edit distance of following DNA sequences:

AGTCTGTA
GTTCTA

Exercise 17: Programming in R (3 Points)

Read Chapters 9 and 10 in „An Introduction to R“.

- a) Which loop constructions exist in R? Why is `for()` used less in R than in languages like Java or C?
- b) How can you identify custom binary operators in R? When is it not necessary to add arguments when calling a function?
- c) Write two R functions that compute the faculty of a natural number. One of the functions shall compute it iteratively, the other recursively. What happens if you enter the name of a function in the R console?

Exercise 18: k-Nearest-Neighbor in R (2 Points)

Start R and import the library `class`. Apply kNN with different parameters k on the Iris data set and compare the results to LDA/QDA. Use the function `knn.cv()` for evaluation and read the documentation of the function to find out how to adapt the data set.