University of Stuttgart | ANALYTIC COMPUTING

Machine Learning Exercise (SS 22)

Assignment 1: k-nearest neighbors (Solution)

Mojtaba Nayyeri
Mojtaba.Nayyeri@ipvs.uni-stuttgart.de

Alexandra Baier
Alexandra.Baier@ipvs.uni-stuttgart.de

Akram Hosseini
Akram.Hosseini@ipvs.uni-stuttgart.de

Tim Schneider
timphillip.schneider@ipvs.uni-stuttgart.de

This assignment sheet consists of theoretical and programming tasks.

Submit your solution in ILIAS as a single PDF file.[1] Make sure to list full names of all participants, immatriculation number, study program and B.Sc. or M.Sc on the first page. Optionally, you can *additionally* upload source files (e.g. PPTX files). If you have any questions, feel free to ask them in the exercise forum in ILIAS.

Submission is open until Monday, 02 May 2022, 11:59 AM.

[1] Your drawing software probably allows to export as PDF. An alternative option is to use a PDF printer. If you create multiple PDF files, use a merging tool (like pdfarranger) to combine the PDFs into a single file.

K-Nearest Neighbors (Theoretical)

Task: The data points $x_1, \ldots, x_6$ are given as follows:

$$x_1 = \begin{bmatrix} 4 \\ 6 \end{bmatrix}, x_2 = \begin{bmatrix} 7 \\ 2 \end{bmatrix}, x_3 = \begin{bmatrix} 9 \\ 3 \end{bmatrix}, x_4 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, x_5 = \begin{bmatrix} 10 \\ 7 \end{bmatrix}, x_6 = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$$

The data points x_1, x_2 belong to class c_1, the data points x_3, x_4, x_5 belong to class c_2. The data point x_6 has not been classified, yet. Show step-by-step how the data point x_6 can be classified by the k-Nearest Neighbors algorithm once with k=1 and once with k=3. Use the squared Euclidean distance as distance metric.

Step
$d_{squared-euclidian}(x, y) = (x_1 - y_1)^2 + (x_2 - y_2)^2$

	Distance to X_6
X_1	9
X_2	16
X_3	13
X_4	17
X_5	10

In the case of $k = 1$, X_6 gets classified as C_1, because the closest data point to X_6 is X_1 and it belongs to C_1.
In the case of $k = 3$, X_6 gets classified as C_2, because in top 3 closest ones X_3 and X_5 are from C_2. So, the most probable cluster is C_2.

K-Nearest Neighbors (Programming)

Task: Please download the Jupyter notebook assignment1.ipynb. Follow the instructions in the Jupyter notebook.