DATA.ML.100 Introduction to Pattern Recognition and Machine Learning

Exam 1, October 14th 2020

J.-K. Kämäräinen / TAU Computing Sciences

page 1/1

You must complete this programming exam yourself using a computer with Internet connection. You are allowed to search and use all available material. You are not allowed to discuss in any form (chat, email, etc.) during the exam (your instructor is the only exception - use Slack). All code and text must be written by yourself. Return *only* the requested files and nothing else!

1. One-Nearest Neighbor (1-NN) Classifier (30p)

Download the 3-dimensional training data (X_train.txt) and labels (Y_train.txt) and test data and labels (X_test.txt and Y_test.txt).

Write Python code that classifies the data points in X_test using the one nearest neighbor classifier (1-NN) and then computes the classification accuracy (correct labels in Y_test).

You must write the classifier code from scratch and you are not allowed to use existing codes or libraries for nearest neighbor classifiers.

Return your code and a screenshot of your desktop that includes the window where your code runs and prints out the classification accuracy:

- <surname>_<student_number>_1nn.py
- <surname>_<student_number>_1nn.png

2. k-Nearest Neighbor (k-NN) Classifier (10p)

Extend the previous code to calculate distances to k best matches and then returns the most frequent class (mode) of them.

You must write the classifier from scratch and not using available examples or ready-made Python libraries for nearest neighbor classification.

Return your code and a graph of the classification accuracy for k=1,2,3,5,10,20:

- <surname>_<student_number>_knn.py
- <surname>_<student_number>_knn.png