Machine Learning (Spring 2020) **(Due: 16.04.2020 23:55)**

Assignment 3

Instructor: Dr. Sergey Kosov *TA:* Yuhou Zhou

Course Policy: Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- The homework assignments are for practice purpose. The grade from your homework will not affect your final grade of the course.
- Please submit your answer sheet, either by a scanned copy or a typeset PDF file, to Moodle before the deadline.
- No late submission is accepted.
- You can do this assignment in groups of 2. Please submit no more than one submission per group.

Problem 1: K-nearest Neighbors Application

(10+5+5=20 points)

From Moodle, download the digits data set and the python script for basic visualization (If you choose to use other languages, you have to translate it). This data set includes 2000 handwritten numerals ('0'-'9') image vectors, extracted from a collection of Dutch utility maps. 200 patterns per class have been digitalized in binary images.

(a) Implement your own K-nearest Neighbors algorithm, using C++ (DGM library, etc) or Python (NumPy, Pandas, etc). Carry out a classification task using your KNN, with a k of your choice. Divide the data set into training set and test set. Show the accuracy of classification.

(b) Implement 10-fold cross validation, and evaluation the accuracy again. Discuss the change, and explain how the cross validation works and why it is useful, in your own words.

(c) Change the value of k, and try to find the k with the highest classification accuracy. Plot the accuracy vs. k. Discuss what you find out, and briefly explain why we get such a result.

Deliverable: A typeset report including instructive graphics, and in a separate text file the code that you produced. Package your two files in a zipfile named "partner1name_partner2name_hw3". The code file must be minimally documented inline such that the TA can quickly grasp what you are computing where in your code. The TA will do no code-checking or code reviewing, but he may want to inspect your code to resolve possible ambiguities in your report.