# Leren: Programming Assignment 4

Week 4 - Semester 1 2015/16

**May be done in teams of 2**
Please indicate which version of Python you are using.
Include the data in a zipfile so that everything runs using only the files in the zip.

1. (6 points) *K-nearest neighbour*
   (a) Implement the k-nearest neighbour algorithm. Use Euclidean distance. Make k a parameter the program so that it can be run with different values of k. Apply it by using the data on digits-123-1 as training data and the data on digits-123-2 as test data. Calculate the proportion correct classifications, the accuracy.

   (b) Implement a new version that weighs each neighbor with the distance to the test example. Evaluate if this is better than the basic version using the same evaluation as above..

   (c) Evaluate the predictive value of each variable by using it to predict the class of an example and measuring how well it predicts. Make a version of k-nearest neighbour that weighs each variable with its predictive value and evaluate this in the same way as above.

   *Submit: code, accuracies on digits-123-2 for each version.*

2. (4 points) *Comparing algorithms*
   Use the digits data of digits123-1 and digits123-2 in the previous assignment.

   (a) Use the programmes that you developed in this course (e.g. Logistic Regression, backpropagation, K-nearest neighbour) to create the best possible handwritten digit recognizer. Be as efficient as possible, minimize the number of optimization steps (a step is running an algorithm on(part of) the data).

   (b) *Submit: report of your optimization steps, the reasons for these steps and the effect on the accuracy op digits123-2. We do not so much evaluate the performance that you obtained as if you took smart steps to optimize it.*

   (c) Include one systematic comparison, using cross validation, to show that your best solution is better than another solution.

   *Submit: code, list of steps that you took with reasons for taking this step and result of the step, results systematic comparison*