Task1, student number: s1643102

Report for k_NN classification

my_knn_system:	Elapsed time is 29.95538seconds.		
Number of Neighbours	Number of test samples	Number of errors	Accuracy
K = 1	7800	1103	85.86%
K = 3	7800	1045	86.60%
K = 5	7800	1070	86.28%
K = 10	7800	1123	85.60%
K = 20	7800	1219	84.37%

In order to reduce the runtime, I created a function “myDistance.m” which takes as arguments two matrices, the first containing testing samples and the second – training samples, and calculates the squared Euclidean distance between each pair of testing and training samples. This improves the runtime significantly because it avoids using nested for loops which take a lot of time to implement due to the large dataset.

For the calculation of the distance matrix DI I used the formula:

$$DI = \underbrace{(XX,\ldots,XX)}_{M} - 2 * X * Y^T + \underbrace{(YY,\ldots,YY)}_{N}{}^T$$

where

$$XX = \begin{pmatrix} X_1 * X_1^T \\ \vdots \\ X_N * X_N^T \end{pmatrix}, \quad YY = \begin{pmatrix} Y_1 * Y_1^T \\ \vdots \\ Y_M * Y_M^T \end{pmatrix}$$

(X_i – test sample; Y_i – train sample).