Exercise 1
The goal of this exercise is to implement a KNN for some experiments. You can use the dataset (letter-recognition.data) on our course website.

(4P) Subtask 1.1
Implement the $k$-nearest neighbour classifier. Test your implementation with a $d_1, d_2, d_{1/2}, d_4$ and $d_\infty$ metric. Use the following definition:

$$\|x\|_p := \begin{cases} \left(\sum_i |x_i|^p\right)^{\frac{1}{p}} & 0 < p < \infty \\ \max_i x_i & p = \infty \end{cases}$$

with $d_p(x, y) = \|x - y\|$. 

(2P) Subtask 1.2
The quality of classification depends on the amount of trainingsdata. Use the first 100, 200, 500, 1000, 2000, 5000 and 7500 examples as trainigs data and the last 1000 as test data. Plot the number of correct classified examples against the amount of trainigs data. You can use the $L_2$ metric.

1Matlab function: importdata
Exercise 2

First, you will implement a PCA and second, you will make some experiments with it. Use the subset (alo.txt) of the letter dataset, which you will find on our course website.

(3P) Subtask 2.1
Implement the following function:

\[ [P, V] = \text{PCA}(D); \]

\( D \in \mathbb{R}^{M \times N} \) is a matrix with \( N \) data of dimension \( M \). \( P \in \mathbb{R}^{M \times M} \) is the projection and \( V \in \mathbb{R}^{M} \) is a vector with Eigenvalues.

Implement the function in the following steps:

- substract the mean \( \mu \)
- Compute the covariance matrix
- Compute the Eigenvectors and Eigenvalues
- Sort the Eigenvalues and modify the Projection accordingly

(3P) Subtask 2.2
Determine the effect of dimension reduction on the dataset. Compute the means square error \( \epsilon_l \) between original data and the projection.

\[ \epsilon_l := \frac{1}{N} \sum_{i=1}^{N} ||x_i - (P_l y_i + \mu)||^2 \]

\( x_i \in \mathbb{R}^M \) is the \( i \)-th vector out of \( D \in \mathbb{R}^{M \times N} \). The dimension reduction can be calculated via

\[ y_i = P_l^T \cdot x_i \in \mathbb{R}^{(M-l)} \]

In \( P_l \in \mathbb{R}^{M \times (M-l)} \) are the \( M-l \) Eigenvectors corresponding to the biggest Eigenvalues. Plot the mean square error \( \epsilon_l \) for \( 1 \leq l \leq 16 \) as function of \( l \).

---

\(^2\)Matlab function: \texttt{mean}, \texttt{repmat}

\(^3\)Matlab function: \texttt{eig}, \texttt{diag}

\(^4\)Matlab function: \texttt{sort}

Questions: Munir.Georges@lsv.uni-saarland.de 2 www.lsv.uni-saarland.de