Homework 1

Submit your solution as one pdf file with Moodle before Wednesday April 29. Include all code.

1. Suppose $Y \sim Bern(0{,}5)$, $X|Y = 0 \sim \mathcal{N}((0,0), \mathbb{A})$ and $X|Y = 1 \sim \mathcal{N}((1,2), \mathbb{A})$ with

$$\mathbb{A} = \begin{pmatrix} 1 & 0{,}4 \\ 0{,}4 & 1 \end{pmatrix}$$

   a) Derive the optimal Bayes classifier for a symmetric loss function.

   b) Generate samples of (X, Y) of size $N = 50, 100, 500$.
      Use the Python function `numpy.random.multivariate_normal` .

   c) Visualize each sample, mark the border of the k-NN classifier with $k = 1, 3, 5$ and compare it with the optimal Bayes classifier.

2. In this excercise we work with the MNIST dataset (as you used in the first part of the course). Split it into a training and test set. Estimate the misclassification error on traning and test set of the k-NN classifier for different values of k. What is the optimal k?