1) Program and implement a k nearest neighbours classifier (k-NN). Use this classifier to solve the following problems:
   i. IRIS PLANT DATABASE (Classification of three different kinds of iris plants).
   ii. PIMA INDIANS DIABETES DATABASE (Classification of pregnant Indians of the Pima tribe according to whether they suffer from diabetes or not).

   The relevant data can be found in the file UCIdata-exercise1.rar.

   Report on the percentage of correct classification as a function of the number of nearest neighbours. Use cross-validation to obtain the results.

2) For the second problem, obtain estimates of the probability density functions for each class, under the following assumptions:
   a) Pdfs are gaussian. The covariance matrices are diagonal, with all diagonal elements equal. Mean and variance of the pdfs are estimated using Maximum Likelihood from the available data.
   b) Pdfs are gaussian, with non-diagonal covariance matrices. Means and covariance matrices of the pdfs are estimated using Maximum Likelihood from the available data.
   c) Components of the feature vectors are mutually statistically independent (the usual naïve Bayes approach). Marginal Pdfs are gaussian, with parameters (mean, variance) estimated using Maximum Likelihood from the available data.
   d) Components of the feature vectors are mutually statistically independent (the usual naïve Bayes approach). Marginal pdfs are computed using 1-d Parzen windows with gaussian kernels. Take the width *h* of each window equal to the square root of the number of patterns in the available data.

   For all assumptions compute the following measures of the goodness of your fit for each class: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) (https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118856406.app5). For this question, use the whole data set available to compute the above measures (do not use cross validation).

3) For each of the above assumptions about the pdfs, implement a Bayes classifier and compute its classification accuracy using cross validation (it goes without saying that for each cross validation iteration, probability density functions will have to be calculated using only training set data for this question). For assumptions c) and d), this will obviously be a naïve Bayes classifier. Taking into account your previous findings, investigate whether more accurate estimates for the pdfs (as judged by the model selection criteria in question 2) tend to improve classification accuracy as well. Compare the performance of the Bayes classifiers to the performance of the k-NN classifier.

4) Implement the perceptron algorithm and use it to perform classification on the IRIS PLANT DATABASE data as follows: Examine whether the data of each class are linearly separable from the data of the combined remaining classes (e.g. if the Iris Setosa data are linearly separable from the combined Iris Versicolor and Iris Virginica data).