Tasks to do:

1. Install the numpy, matplotlib, and scikit-learn libraries.

2. Familiarize yourself with the functionalities available in the sklearn.datasets submodule https://scikit-learn.org/stable/datasets.html#datasets. The tasks in this lab will be performed on the Iris dataset.

3. Familiarize yourself with the documentation of the SVC class and the description of the SVM method (SVC). https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

4. For the breast cancer dataset, do the following:

Data set description:

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html

(a) Load the dataset.

(b) Split the data into a training set and a testing set using the train_test_split function with a 0.7 training data and 0.3 testing data ratio:

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html.

(c) Train the SVC classifiers on the training data.

(d) Calculate the classification accuracy on the testing data.

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html.

(e) Repeat steps 4c and 4d to find the optimal parameter C and to examine the influence of the kernel argument, namely the kernel='linear' and kernel='rbf' options.

(f) Train two classifiers: one with kernel='linear' and the other with kernel='rbf', as well as their respective values of C, and then for each of them do the following:

i. Create a confusion matrix showing the classification accuracy on the testing data (https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html#confusion-matrix).

ii. Create a scatter plot showing the actual class division, and then create the same plot using the predicted results of the built classifier instead of the actual class vector. The points on the plot must be colored according to their respective classes!

Note: Due to the large number of attributes in this dataset, it is necessary to use dimensionality reduction methods, such as PCA

(https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html).