

Homework – 1

(Deadline: 8 April 2020, 23:59)

The aim of this homework is to

- i. Understand the importance of feature scaling
- ii. Implement a KNN classifier on a binary classification problem

Data Set:

The dataset for this homework is provided in an accompanying excel file. The name of the excel file is “dataset.xlsx”.

There are 61 rows and 3 columns in the dataset. First two columns are for features, “feat_1” and “feat_2”. The name of the final column is “target” and it contains the class labels.

First row contains the row names. Rows from 2 to 26 are samples from “class 1” and rows from 27 to 51 are samples from “class 0”. In other words, there are 25 samples from each of the classes.

Finally, rows from 52 to 61 are the test samples.

Distribution of the data are given in scatter plot in Figure 1.

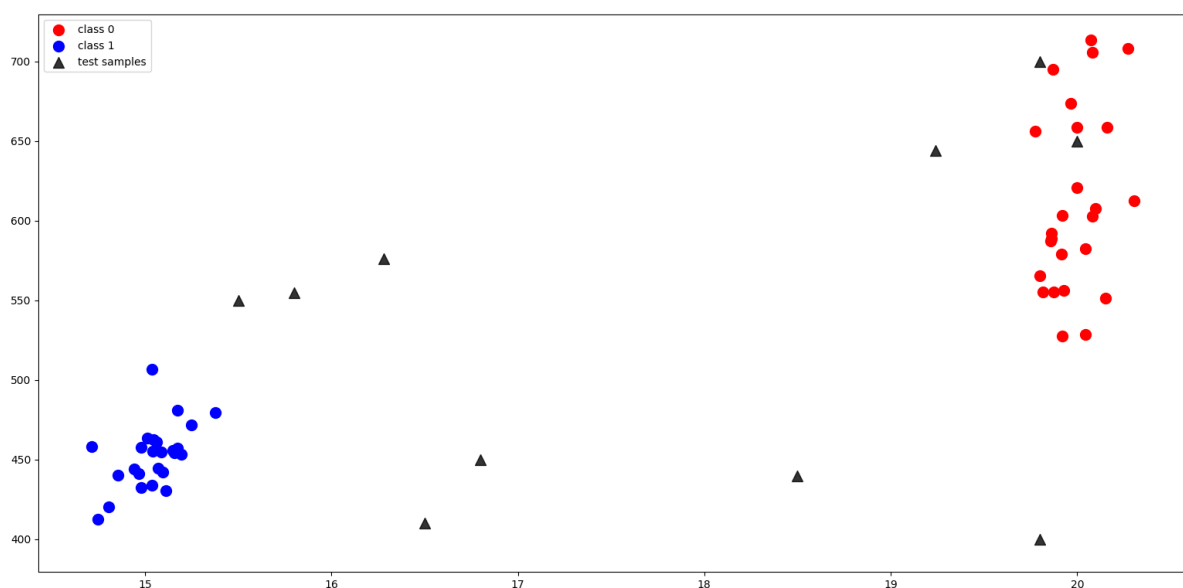


Figure 1. Distribution of the data in the dataset

Assignments:

You are expected to complete the following tasks:

1. Read the excel file. Split training and test data and store them in different variables.
2. Write a function that normalizes the feature vectors. The name of the function should be “scale_feature”. The method to be implemented within the function is “z-score normalization”.
 - i. Input of the function: one single feature vector

- ii. Output of the function: scaled version of the feature vector
3. Write a function that classifies a given data using KNN classification method. The name of the function should be “knn_classify”.
 - i. Input of the function: training data, test data (note that these are to be obtained in task 1) and K value of KNN classifier
 - ii. Output of the function: predicted class labels for the test data
4. Write a function that generates the “confusion matrix” for evaluation of your classifier built in task 3. The name of your function should be “confusion_matrix”.
 - i. Input of the function: predicted class labels and actual class labels
 - ii. Output of the function: the confusion matrix (it may be a 2D array), and the classification accuracy.
5. Write a function that plots training data together with your prediction results. The name of your function should be “plot_data”.
 - i. Input of the function: test samples with predicted class labels, training samples
 - ii. Output of the function: A scatter plot displayed on screen. (no value is returned from the function). The scatter plot should be something similar to Figure 1. Only the black rectangles should be replaced by your prediction results.
6. After you complete the first five tasks, you will call the “knn_classify” function twice.
 - i. In the first call, you should input the data without performing scaling (i.e. without calling the function “scale_feature”.) In other words, input the data you read from the excel file directly to the “knn_classify” function. Then calculate the performance of the classification by calling “confusion_matrix” function and print the classification accuracy on screen.
 - ii. In the second call, you should input the data after performing scaling (i.e. first you should call the function “scale_feature”.) In other words, input the data you read from the excel file first to the “scale_feature” function, then that scaled data should be input to the “knn_classify” function. Then calculate the performance of the classification by calling “confusion_matrix” function and print the classification accuracy on screen.

Notes:

1. You are free to use any programming language you like.
2. Your code should be well-commented. (i.e. you are expected to insert explanations in your code)
3. You should write the algorithm for KNN (task 3) by yourself. In other words, you are not allowed to use ready-to-use machine learning packages for task 3. For other tasks, you may use such packages in case you need.

Submission:

In order to submit your homework, you need to send the file containing your source code to ercanavsar@gmail.com.

EE-0010 - IoT applications of Machine Learning Methods

The name of the file should contain your name and surname. For instance, for a source code written in Python, and for a student whose name is ERCAN and surname is AVŞAR, the name of the submitted file should be “hw1_ERCAN_AVSAR.py”.