Programming Assignment #1

Question 3: In this question we will provide three files: training_samples.txt, training_labels.txt and testing_samples.txt. Write a program to predict the category of a testing sample with K-NN algorithm given hyperparameter K, training samples and the labels of the training samples. Note that you are expected to implement the K-NN algorithm by yourself. In other words, you are supposed to not directly use the well-implemented K-NN class or functions in packages or libraries, e.g. scikit-learn. But it is ok to use numpy and scipy.

File format:

(1) training_samples.txt: N lines of float values separated by space. Each line is a training sample.
(2) training_labels.txt: N lines where each line contains an integer.
(3) testing_samples.txt: N lines containing the hyperparameter K and a testing sample. The elements in each line are separated by space. The first element is an integer (the hyperparameter K to be used for the testing sample in the line) and the rest elements (float) are the coordinates of the testing sample.

Question 4: In this question we will provide a file containing N lines. Each line includes two values x and y, representing a training point. You are supposed to write and run a program to fit a linear regression model y = wx+b with the given points and then plot the fitted line and training points in a figure. Note that you are expected to implement the linear regression algorithm by yourself. In other words, you are supposed to not directly use the implementation in packages or libraries, e.g. scikit-learn. But it is ok to use numpy and scipy for data processing and calculation, like matrix multiplication.

File format:

(1) 2D_points.txt: N lines of float value pairs separated by space. Each line is a training sample (x,y).