Assignment Week 2

Python

Submission Deadline: 06-11-2018

Assume that you are given a text file named “data.txt”. Each line except the first one of the data file contains information corresponding to a sample. The first value is a number that represents the value of an attribute X of a sample and the second value is either “YES” or “NO”, which represents if the sample belongs to a class or not.

1. Now write a function named `readAllData` that takes as parameter the name of the file. The function then reads the data from the file and returns a list of tuples where each tuple corresponds to the information of a sample.

2. Write a function named `computeAverageForClasses` to the take the data read in function 1 as argument and compute the average value of X for each class. The function then returns a dictionary where the keys are name of the class and the values are average of X for that class.

3. We assume that a sample should be a member of class A, if the value of X for that sample is closer to the average value of X for class A than any other classes. Now using this rule write a function named `countMisclassified` that returns how many samples in this data is misclassified.

4. Using these functions print the average values of X for each class and the number of misclassified samples in the data. Finally, write the misclassified samples in a file named “Misclassified.txt”.