ASSIGNMENT 4

Training Iris/Fisher's Data Set using Gradient Descent.

You have to implement the gradient descent algorithm and use it to train the Irsi/Fisher's Flower data set. This data set has three classes. You can assign a class label of 1, 2 and 3 to these classes.

The purpose of this assignment is to see how well does a perceptron (minus step/activation function ;)) trained using gradient descent algorithm perform on a three class data set.

Step 1: Data set. Presently, the data set is sorted class-wise. It is best to randomize the data set. You will use all 4 features during training. Split the dataset into 2 parts. One part is for training and other testing.

Step 2: Training: Use first 75 (randomized) samples to train the perceptron (minus step function) using gradient descent. You need to document the training error. Ideally, we would want it to be 0, but if it isn't and your algorithm has exhausted its **max_iterations** limit then there would be some cost/error which you need to **document**. You must also **note** the cost after every 100 iterations and show how its improves gradually. A **graph** would be nice with x-axis showing iterations: 100, 200 etc. and y-axis is the cost.

Step 3: Testing: Once the training is complete and you have a trained model which would be the weight vector [wo, w1, w2, w3, w4], test the classifier using the other 75 samples. Take each sample, get it classified by the trained classifier. Count the number of mismatches. Is there a problem? Can you see the problem of not using a step/activation function? **Write/document this**.

Try rounding off the output for every test sample. Suppose the classes are 1, 2 and 3. Suppose the output of the classifier is yHat. Then round off the value of yHat using the following scheme:

class is 1 if yHat<= 1.5
class is 2 if 1.5 < yHat <= 2.5
class is 3 if yHat > 2.5

Calculate the mismatches again. Did the results improve? **Document** the results. The rounding off must only be done during the testing phase.

Repeat step 2 and 3 by splitting the data 100:50. Did the results improve? **Document**.

You need to submit the code and a report that contains all the highlighted details in the problem statement.