**Aman Bassi(1001393217)**

**Problems Encountered**: -

1. At First, it was difficult to understand the numpy arrays.
2. Secondly, it was difficult to visualize what’s happening in background when we are finding the optimized weights or parameters.
3. Moreover, it was also very difficult to analyse the iris- dataset on the graph which I was trying in octave by plotting it and trying to analyse it
4. Lastly, it was really difficult to choose K value for K-fold cross validation

**Data:**

Our data is of Iris flower where we are having 3 classes or 3 types of it: Setosa, Versicolour, Virginica and our data has 150 examples and four features : Sepal length, sepal width, petal length and petal width

**Results**:

Since I am generating random indices for test data and train data so I am getting different accuracies every time but for most of the time my accuracies are coming around **96 percent**

Which is the mean accuracy for accuracies of all folds.

**Description**:

My code is in file Bassi\_00\_01.py which is a code in python and I have used numpy and random

Packages of python. You can directly run the file and u can see the 10 accuracies for 10 different folds and Atlast, you will be seeing the mean of all those accuracies.

It’s in python3 so make sure your interpreter supports python3.

I have changed the classes label as 1 ,2 ,3 instead of : Setosa, Versicolour, Virginica which I have changed in my text file which I have added in the code folder as irisdata.txt so you need keep the irisdata.txt and my python file in the same folder and then you can simply run it.

I have used K=10 because:

1. It’s a small dataset and moreover the data is quite uniform.
2. I believe that k=10 gives the best trade-off between the running speed and the accuracy estimation.
3. Moreover, I tested my dataset for different values of k and I was getting the best results in case of k=10
4. Atlast, for k=10 in one folds there will be 135 training examples and 15 testing examples which shows you are getting some data to train your model and giving those optimized parameters whereas ,if it would have been k=5 then there will be 120 training examples and 30 testing that was also predicting good but not better than k=10 as training was just 120 so I think for such small dataset u need more training as well as some testing. If you now look for k=15 then there will be just 10 testing examples which are too small. Though training is good but I don’t think it would be good for generalization as based on just 10 examples you can not say that it is predicting good on unseen data.