

Assignment 3

Write jupyter notebook scripts for all of the questions. Print out the required results.

1. Linear Discriminant Analysis Classifier (50 points)

Apply LDA in `sklearn` to virginica and versicolor in the iris dataset. Use the `train_test_split` method to split the data for training and testing. Use 10% of the samples for testing. For each testing, compute the average error rate. Repeat the process of training, testing, and computing average error rate 20 times. Plot the average error rate vs index ([1, 2, ..., 20]) of the training-testing cycle.

Use the following python code to import the LDA module:

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
```

Note that setosa flowers do not need to be considered for this problem.

You can find information about how to use LDA in sklearn here.

2. Naive Bayes Classifier (50 points)

(1) Repeat the process in Problem 1, but using the sklearn Naive Bayes classifier.

Use the following python code to import the Naive Bayes classification module:

```
from sklearn.naive_bayes import GaussianNB
```

You can find information about how to use Naive Bayes Classifier in sklearn here.

(2) Use the `train_test_split` method to split the data for training and testing. Use 10% of the samples for testing. For the testing, use

```
predict_proba()
```

to obtain the probabilities of the predicted class labels. Print out the probabilities. You only need to do this training-testing cycle ONCE.