**CS5590/490 Python-DeepLearning**

**Python**

**LAB**

**Assignment-3**

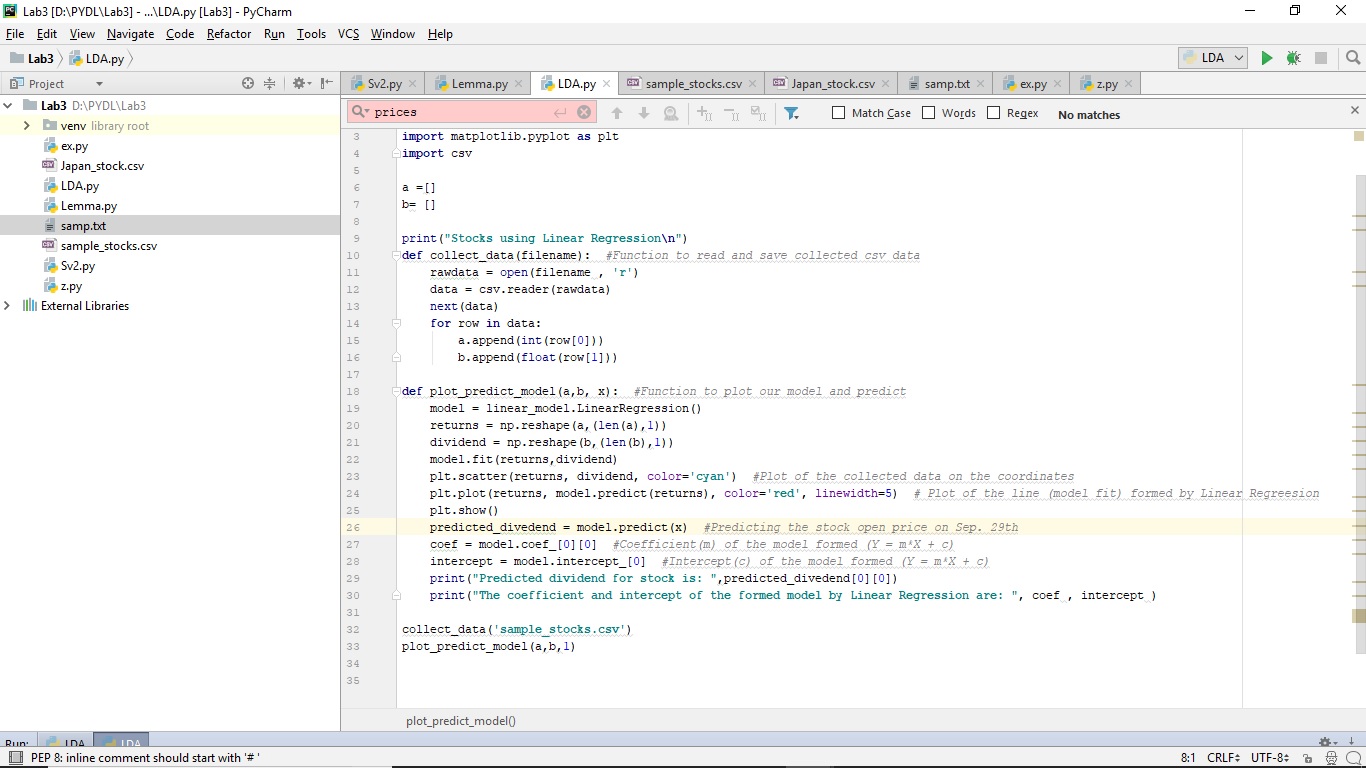
**Submitted by: Shajee Uddin Zain Mohammed**

**Student ID: 16230179**

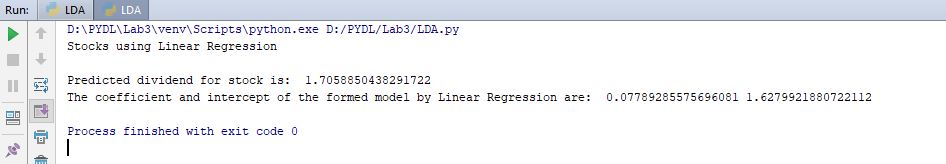
**Class ID: 28**

1. Pick any dataset from the dataset sheet in class sheet and make one prediction model using your imagination with Linear Discriminant Analysis.

Source Code:



Output:



Outcome:

In linear regression the outcome graph is continuous. It has large number of possible values.

In logistic regression the outcome is a curve so there are limited number of possible values.

Dependent Variable:

In linear regression the response variable is continuous such as age, height.

In logistic regression the response variable is categorical.

Error minimization Techniques:

Linear regression uses least squares method.

Logistic regression uses maximum likelihood.

2) Implement Support Vector Machine classification,

1) Choose one of the dataset using the datasets features in the scikit-learn

2) Load the dataset

3) According to your dataset, split the data to 20% testing data, 80% training data(you can also use any other number)

4) Apply SVC with Linear kernel

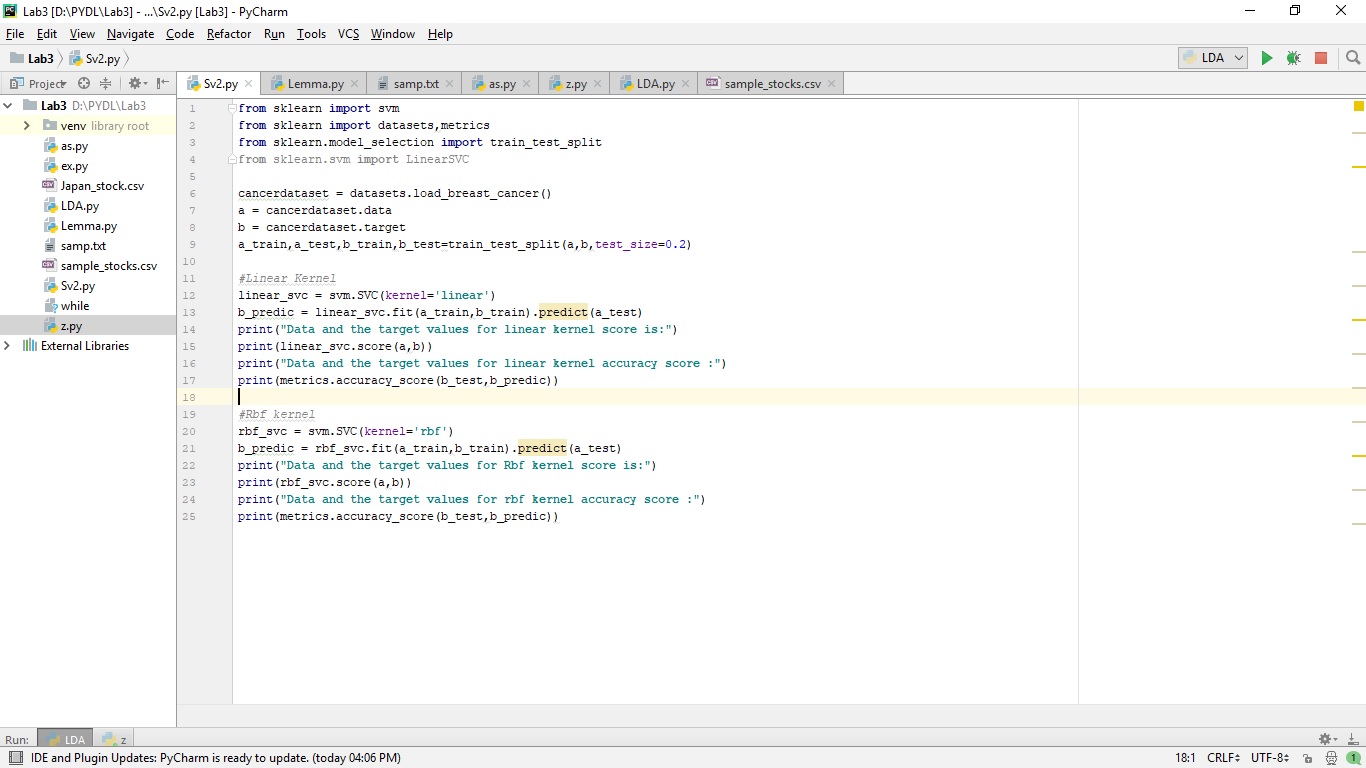
5) Apply SVC with RBF kernel

6) Report the accuracy of the model on both models separately and report their differences if there is

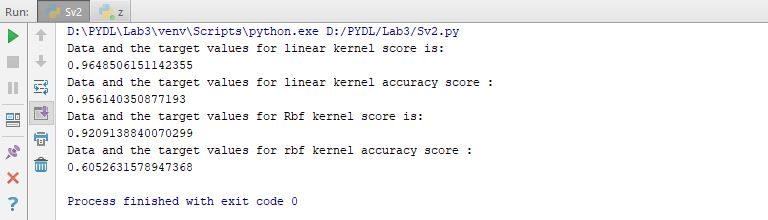
7) Report your view how can you increase the accuracy

and which kernel is the best for your dataset and why

Source Code:



Output:



Accuracy of the Linear kernel and rbf kernel:

Linear: 95%

Rbf: 60%

The accuracy of the linear kernel is more than the Rbf because linear kernel is easy to train the model as I have used 80% data as the train dataset and 20% as the test dataset the accuracy of the Rbf does not improves. Since the dataset I’m using is more suitable for linear as it does not have multi-dimensions. Rbf will be more accurate when we have more training data for observation than the number of features.

1. Write a program Take an Input file. Use the simple approach below to summarize a text file:- Read the file

-Using Lemmatization, apply lemmatization n the words

-Apply the bigram on the text

-Calculate the word frequency (bi-gram frequency) of the words (bi-grams)

-Choose top five bi-grams that has been repeated most

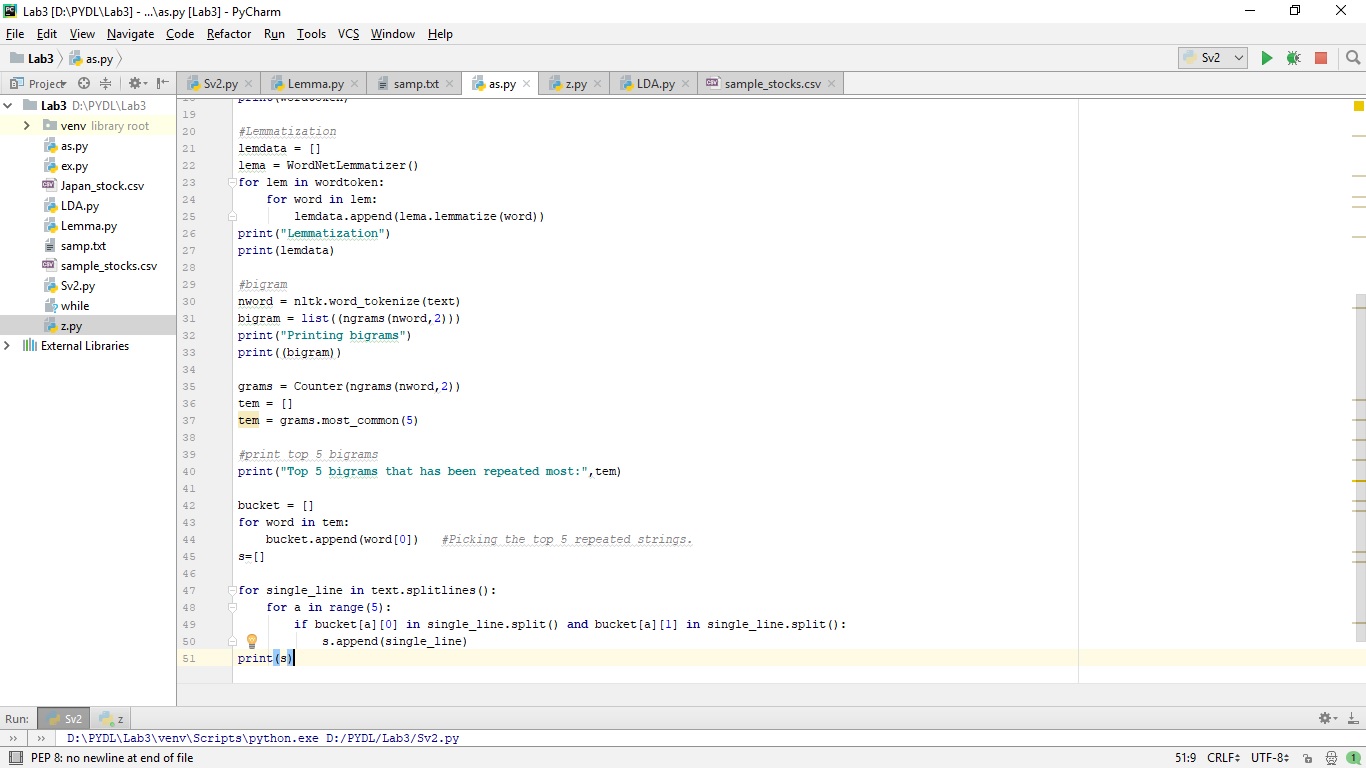
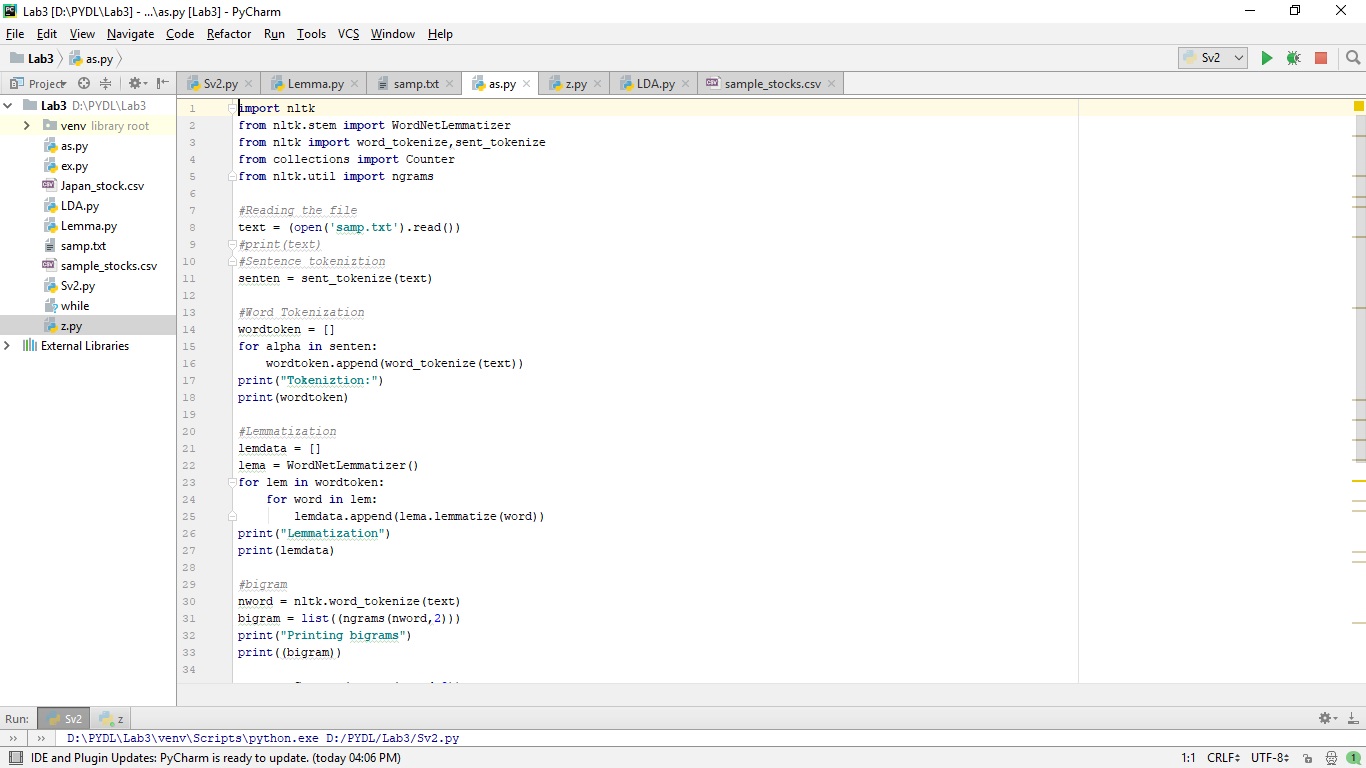
-Go through the original text that you had in the file

-Find all the sentences with those most repeated bi-grams

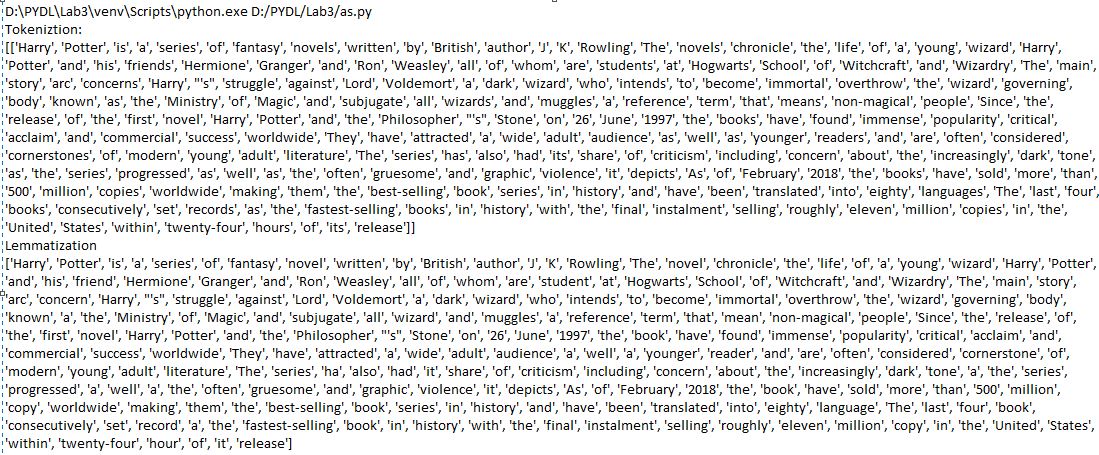
-Extract those sentences and concatenate

-Enjoy the summarization

Source Code:



Output:



4) Report your views on the k nearest neighbor algorithm when we change the K how it will affect the accuracy. Provide a good justification about the changes of the accuracy when we change the amount of K. For example: compare the accuracy when K=1 and K is a big number like 50, why the accuracy will change

Yes, the accuracy will be effected because when you consider k=1 then we are restraining the region and making the classifier more blind to the large set of neighbours.

Since the region is small we get more accurate results. It will have low bias but high variance.

When we use large values of k then we explore the large regions which have smooth decision boundaries which means, it will have low variance and high bias