Problem Statement:

Scikit-learn's iris dataset (http://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html) only has four features and 150 data points, so PCA and LDA isn't extremely useful. But let's see how they do anyway.

Use PCA (from sklearn, google for details) to reduce the dimensionality to 3, 2, and 1. Then:
Use naïve bayes to classify the reduced datasets (with 3, 2, and 1 features) and also with the unreduced dataset (the original 4 features before PCA was applied).
How does the classification accuracy (the proportion of data points it predicts correctly to the total # of data points) change? Report exact values.

:-

Using PCA & NB

Code File : pca.py

Output:

```
C:\Python\HW3>python pca.py
Flowers:- ['setosa' 'versicolor' 'virginica']
Total No of Data Points 150
**************** With All Features ********************
6 datapoints out of 150 are not correctly predicting with all feature consideration


 Classification Report :-
           precision    recall  f1-score   support

        0       1.00      1.00      1.00        50
        1       0.94      0.94      0.94        50
        2       0.94      0.94      0.94        50

avg / total       0.96      0.96      0.96       150

************** With PCA Component 3 ******************
11 datapoints out of 150 are not correctly predicting using '3 components PCA'


 Classification Report :-
           precision    recall  f1-score   support

        0       1.00      1.00      1.00        50
        1       0.88      0.90      0.89        50
        2       0.90      0.88      0.89        50

avg / total       0.93      0.93      0.93       150

************** With PCA Component 2 ******************
```

```
************** With PCA Component 2 ******************
15 datapoints out of 150 are not correctly predicting using '2 components PCA'.


Classification Report :-
          precision   recall  f1-score   support

      0      1.00      1.00      1.00        50
      1      0.82      0.90      0.86        50
      2      0.89      0.80      0.84        50

avg / total    0.90      0.90      0.90       150

************** With PCA Component 1 ******************
9 datapoints out of 150 are not correctly predicting using '1 components PCA'.


Classification Report
          precision   recall  f1-score   support

      0      1.00      1.00      1.00        50
      1      0.90      0.92      0.91        50
      2      0.92      0.90      0.91        50

avg / total    0.94      0.94      0.94       150


C:\Python\HW3>
```

Using above program result,

Implementing	Total Data Points	Naïve Bayes Result (Wrong Count)	Accuracy (rounded)
All Features (4)	150	6	96%
3 Components PCA implementation	150	11	93%
2 Components PCA implementation	150	15	90%
1 Components PCA implementation	150	9	94%

By looking at above result table Naïve Bayes is have more accurate result without using PCA dimension reduction algorithm.

Using LDA & NB

Code File: lda.py

Output:

```
C:\Python>python lda.py
Flowers:- ['setosa' 'versicolor' 'virginica']
Total No of Data Points 150
**************** With All Features ********************
6 datapoints out of 150 are not correctly predicting with all feature consideration


  Classification Report :-
            precision    recall  f1-score   support

        0       1.00      1.00      1.00        50
        1       0.94      0.94      0.94        50
        2       0.94      0.94      0.94        50

avg / total       0.96      0.96      0.96       150

************** With LDA Component 3 ******************
3 datapoints out of 150 are not correctly predicting using '3 components LDA'


  Classification Report :-
            precision    recall  f1-score   support

        0       1.00      1.00      1.00        50
        1       0.98      0.96      0.97        50
        2       0.96      0.98      0.97        50

avg / total       0.98      0.98      0.98       150

************** With LDA Component 2 ******************
************** With LDA Component 2 ******************
3 datapoints out of 150 are not correctly predicting using '2 components LDA'.


  Classification Report :-
            precision    recall  f1-score   support

        0       1.00      1.00      1.00        50
        1       0.98      0.96      0.97        50
        2       0.96      0.98      0.97        50

avg / total       0.98      0.98      0.98       150

************** With LDA Component 1 ******************
2 datapoints out of 150 are not correctly predicting using '1 components LDA'.


  Classification Report
            precision    recall  f1-score   support

        0       1.00      1.00      1.00        50
        1       1.00      0.96      0.98        50
        2       0.96      1.00      0.98        50

avg / total       0.99      0.99      0.99       150


C:\Python>
```

Using above program result,

Implementing	Total Data Points	Naïve Bayes Result (Wrong Count)	Accuracy (rounded)
All Features (4)	150	6	96%
3 Components LDA implementation	150	3	98%
2 Components LDA implementation	150	3	98%
1 Components LAD implementation	150	2	99%

By looking at above result table Naïve Bayes is have more accurate result with using LDA dimension reduction algorithm.