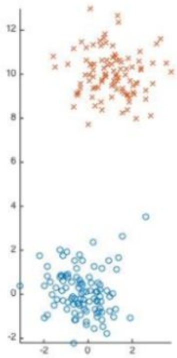
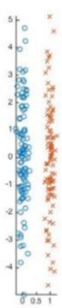
Joshua Siy
HW03
1)
a)
Using PCA on this dataset:. Reducing the dimensions of the data set in the y axis can make a fine line of separation between x and o's, making PCA applicable to this dataset

Using LDA on this dataset is applicable because of the fine line of separation between them where you can draw a horizontal line to classify and keep the clusters separated.



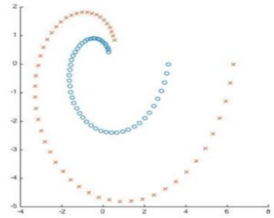
b)
Using PCA on this dataset: I would like to say that PCA works IF the dataset is not confined to selecting the highest covariance as basis to create. Sadly it bases its dimension reduction results of the element with the highest variance. which is around the y axis(vertical) of the dataset therefore making it not applicable.

Using LDA on this dataset is applicable because of the fine line of separation between them where you can draw a vertical line to classify and keep the clusters separated.



c)
Using PCA on this dataset:. Reducing the dimensions of this dataset in the x axis or the y axis does not make any separation between x and o's, making PCA not applicable to this dataset.

Using LDA on this dataset: Cannot draw any kind of line that could separate it. LDA is not applicable to this. However one can use a non-linear approach to separate these datasets.



2)
1))
X=the wholedataset
N=number of data or number of rows in a dataset
U=Eigenvector
x=data point
mu= mean
A=lambda=eigenvalue
E=covariance

mu = X/N
E=summation(X[i]-mu)/(N-1)
E=UA(U.T)
This formula is associated to eigen decomposition and can help us extract eigenvalue and eigenvectors or you can use np.linalg.svd(cov) function to extract, U and A
z=(U.T)*(x-mu)/sqrt(A)

2))

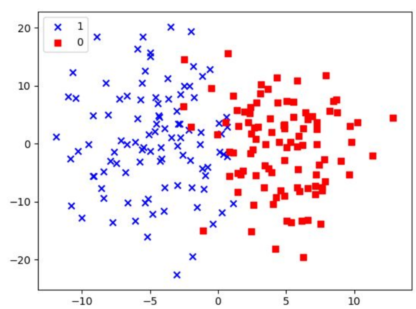
2. Suppose the original data set, $\mathbf{X}$, has features with widly varying range and variance. Could this data be whitened? Why would you or would you not want to perform whitening on data of this type? Generate two dimensional data with these characteristics (i.e., the two features have very different range and variances from each other). Implement python code that applies principal components analysis (without dimensionality reduction, only decorrelation) and data whitening to your generated data set. Scatter plot the original data, the data after application of the PCA transformation and the data after data whitening. What is the covariance of the original data, after PCA and after data whitening? Use these results to motivate your discussion of why you might or might not want to apply data whitening to data with features of widely varying range and variance.

Any data can be whitened as long as the eigenvalue is not equal to zero. It could work well depending on the dataset's structures. Normally whitening is used to reduce redundancy of data. So even if there are no redundant data found, there is probably no reason to not use whitening in the dataset to reduce noise and redundancy

```
original covariance generated from data
[[35.46050535  1.46148765]
 [ 1.46148765 66.58044633]]
covariance of decorrelated data via PCA
[[66.64893155  0.        ]
 [ 0.         35.39202013]]
whitened covariance
[[9.99999850e-01 3.12424570e-17]
 [3.12424570e-17 9.99999717e-01]]
```

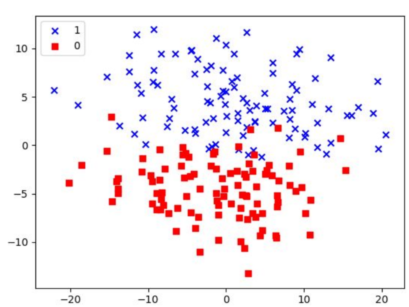
The results show that the whitening the data reduces the covariance into almost making it an identity diagonal matrix. With these results, the assignment taker's perspective sees no reason why whitening shouldn't be applied where it minimizes the data to a comparable amount, reducing redundancy and repositions your data to an observable location. However, if the dataset's shape and size is actually very important where every single detail and position is necessary and need not be reduced, it is sometimes best to leave the data unwhitened. So i guess when we are dealing with data separation of dataset noise reduction, the application of PCA whitening is highly acceptable.

The Original Plot

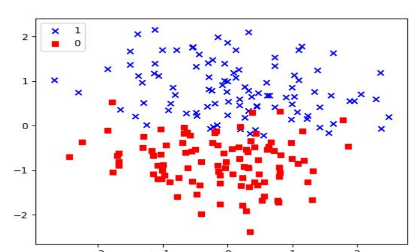


Decorrelated by the algorithm provided in the lectures

[Multiply the Dataset with the set of Eigenvectors that were sorted]



Whitened data:



References:

https://www.kdnuggets.com/2016/03/must-know-tips-deep-learning-part-1.html
http://ufldl.stanford.edu/tutorial/unsupervised/PCAWhitening/
Mr. Wells
Ms. Guo
Mr. McCurley
Gayatri
Lars
Alexandru
Junghoon Woo