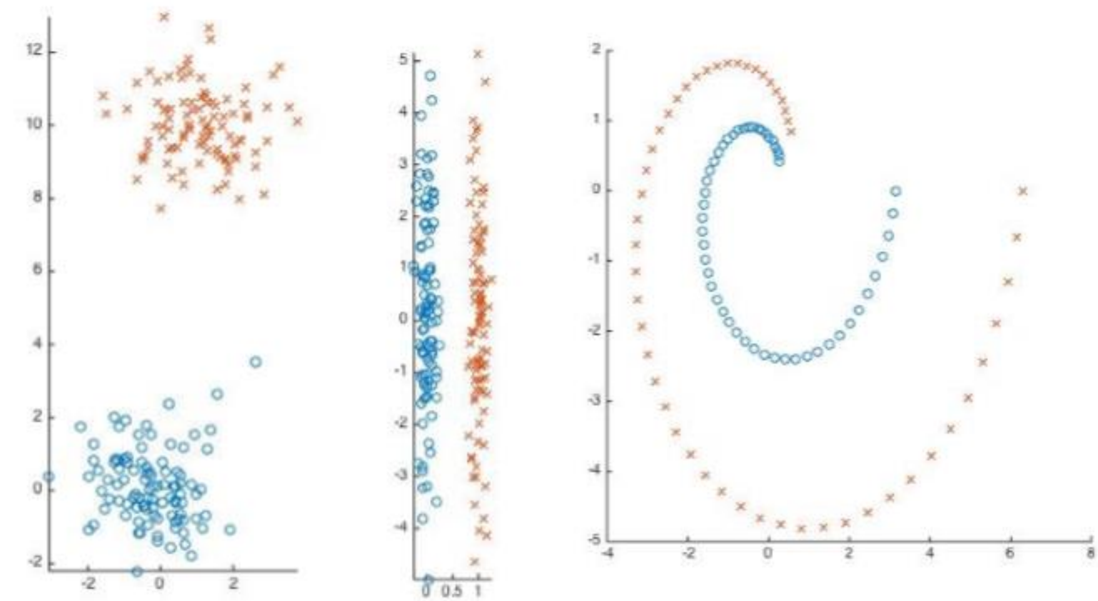
Diandra Prioleau
23 October 2018
EEL 5840 – Fundamentals of Machine Learning
HW03

Question 1



1. For the second and third data sets above, PCA would not be effective in keeping the two clusters separated in the reduced dimensionality data. When applying PCA to the second data set, the data will collapse on the y-axis, since the maximum variance between the two dimensions is along the y-axis. As a result, the data in each cluster in the data set will be on top of each other, losing separation between the clusters. When applying PCA to the third data set, the data will collapse on the x-axis, since the maximum variance between the two dimension is on the x-axis. The third data set is not linearly separable. In order for PCA to be effective, the dataset must be linearly separable. Therefore, for the third dataset, another method would need to be used to keep the clusters separated in a reduced dimensionality data.
   In contrast, for the first data set, PCA would be effective in keeping the two clusters separated in the reduced dimensionality data because after applying PCA to the data set, the data would collapse on the y-axis, which has the maximum variance in comparison to the x-axis and would not overlap between the clusters. This means the dimension maintained (y-axis) also contains the most important information in discriminating the data. However, examining the data more closely, it is possible for maximum variance to also be along the x-axis. If so, the data of each cluster will overlap causing PCA to not be effective in keeping the clusters separated in the reduced dimensionality data.
2. For the first and second data sets above, LDA would be effective in keeping the two clusters separated in reduced dimensionality data because the data is linearly separable. In addition, when the data of each class is projected onto the orthogonal line, the reduced dimensionality data maintains separation between classes. However, for the

third data set, LDA would not be effective because the data set is not linearly separable; as a result, when the data of each classes is projected onto the orthogonal line, the data from each class would overlap.

Question 2

1. The formula for computing z, which is the *whitened* version of x where ∑ is covariance for the original data set X and μ is the mean (x is one data point in X), is below.

$$z = L^{-1/2}U^T(x-\mu)$$

   where L is a DxD diagonal matrix with elements λ_i and U is a DxD orthogonal matrix with columns given by u_i.

   After whitening the data, the covariance of the transformed data set will be the identity matrix, which is shown below.

$$\frac{1}{N}(z_n z_n)^T = \frac{1}{N}\sum_{n=1}^{N} L^{-1/2}U^T(x-\mu)(x-\mu)^T U L^{-1/2}$$

$$= L^{-1/2}U^T S U L^{-1/2} = L^{-1/2} L L^{-1/2} = I$$

   Since the covariance is an identity matrix, it indicates that the data has been decorrelated and normalized.

   ***Reference: Pattern Recognition and Machine Learning Textbook, pg. 568**

2. If an original data set, X has features with widely varying range and variance, it is possible for the data to be whitened. Reasons for performing whitening on data of this type would be to normalize and to decorrelate the data. Figure 1 shows the covariances after conducting PCA and data whitening on the original data, which indicates that the data is decorrelated and normalized. Since the covariances are diagonal matrices, this indicates that data has been decorrelated, meaning the features are not dependent on each other. This also shown in Figure 3, in which the original data has been decorrelated. Decorrelating the data helps to determine how each feature independently affects the data.
   Since the covariance after whitening the data is an identity matrix, this means the data is normalized so that the features are equally weighted. Normalizing the data, shown in Figure 4, is important when some features of a data set have a wider range in comparison to other features. If the data is not whitened, the features with a wider range will be more dominant; therefore, whitening the data will allow all features to be equally weighted, which is an advantage of normalizing the data through data whitening.

```
IPython console
Console 1/A

In [129]: runfile('C:/Users/Diandra/Documents/EEL5840/homework-03-prioleaud/hw03.py', wdir='C:/
Users/Diandra/Documents/EEL5840/homework-03-prioleaud')
Mean of Original Data
 [0.49953287 2.96243978]
Covariance of Original Data
 [[ 2.03361087  4.5107141 ]
 [ 4.5107141  24.95015865]]
[[ 2.03361087  4.5107141 ]
 [ 4.5107141  24.95015865]]
Covariance after PCA
 [[2.58060466e+01 1.92749831e-15]
 [1.92749831e-15 1.17772291e+00]]
Covariance after Data Whitening
 [[1.00000000e+00 3.30733105e-16]
 [3.30733105e-16 1.00000000e+00]]
```

Figure 1 Covariances of Original Data, after PCA, and after Data Whitening

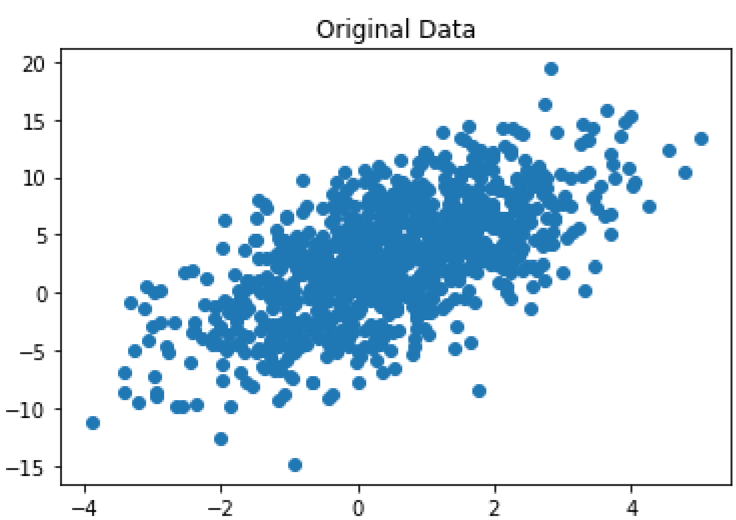


Figure 2 Original 2D Data

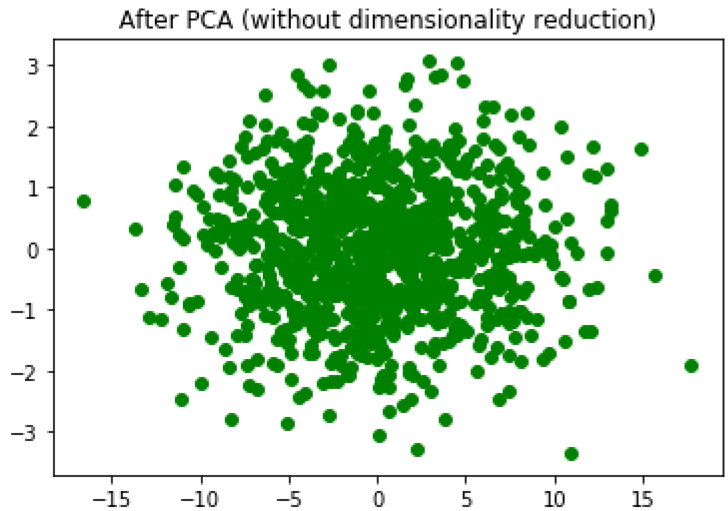


Figure 3 Data after PCA (without dimensionality reduction)

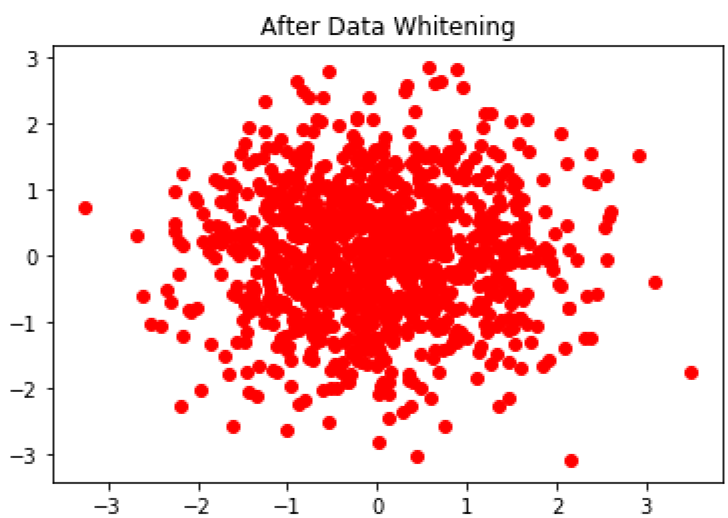


Figure 4 Data after Whitening