HW03

Q1:

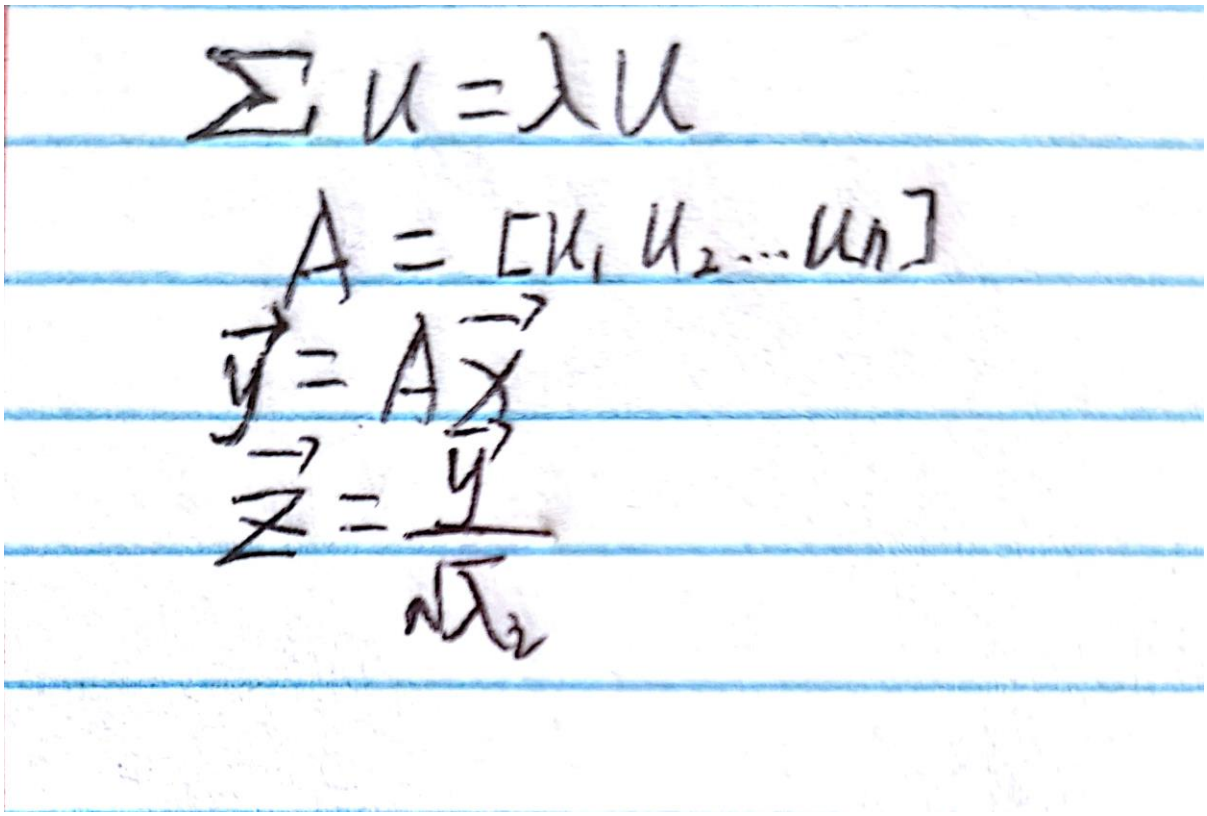
1. For Figure 1, PCA will be effective and keep the two clusters separated in the reduced dimensionality data. Because PCA maintain the dimensions with largest variance. If our data's most useful information (like labels) has the same direction of largest variance, PCA will be effective, if not, PCA will fail because PCA itself does not consider label. LDA will be effective because two data sets are well separated and can be linear separated. It can be separated when projected into orthogonal dimension of separated line.

2. For Figure 2, LDA will be effective and keep the two clusters separated in the reduced dimensionality data. Because two data sets are well separated and can be linear separated. It can be separated when projected into orthogonal dimension of separated line. If two data sets are linearly separable, LDA works well, if not, LDA will fail. PCA will fail because PCA maintain the dimensions with largest variance. Here our two cluster will not be separated.

3. For Figure 3, PCA will fail because our data cannot be separated perfectly by maintaining dimension with largest variance. LDA will fail too because two data sets are not linearly separable.

Q2:

1.


$\Sigma u = \lambda u$
$A = [u_1\ u_2 \dots u_n]$
$\vec{y} = A\vec{x}$
$\vec{z} = \frac{\vec{y}}{\sqrt{\lambda_2}}$

2.

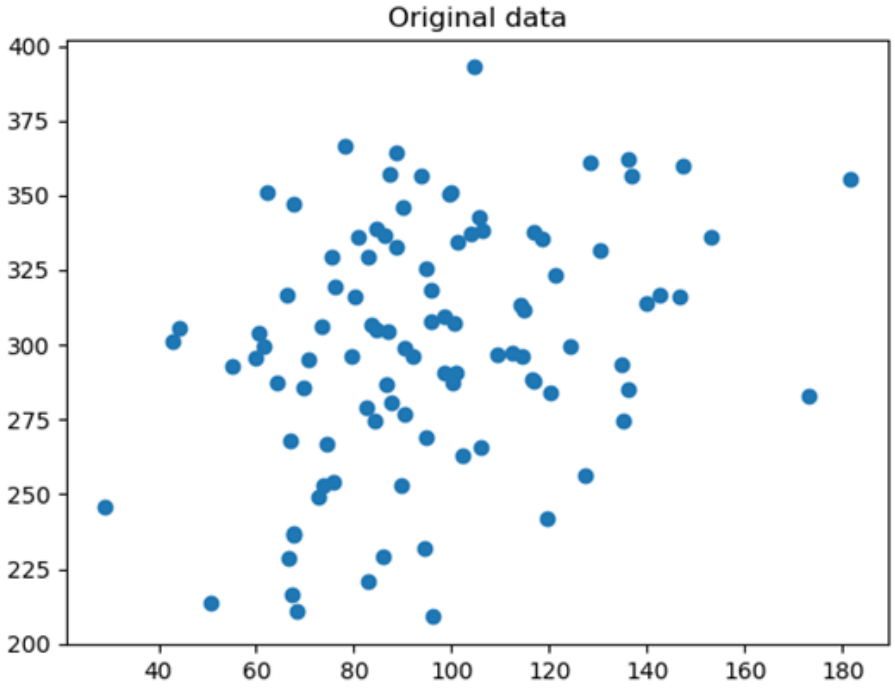


Figure 1

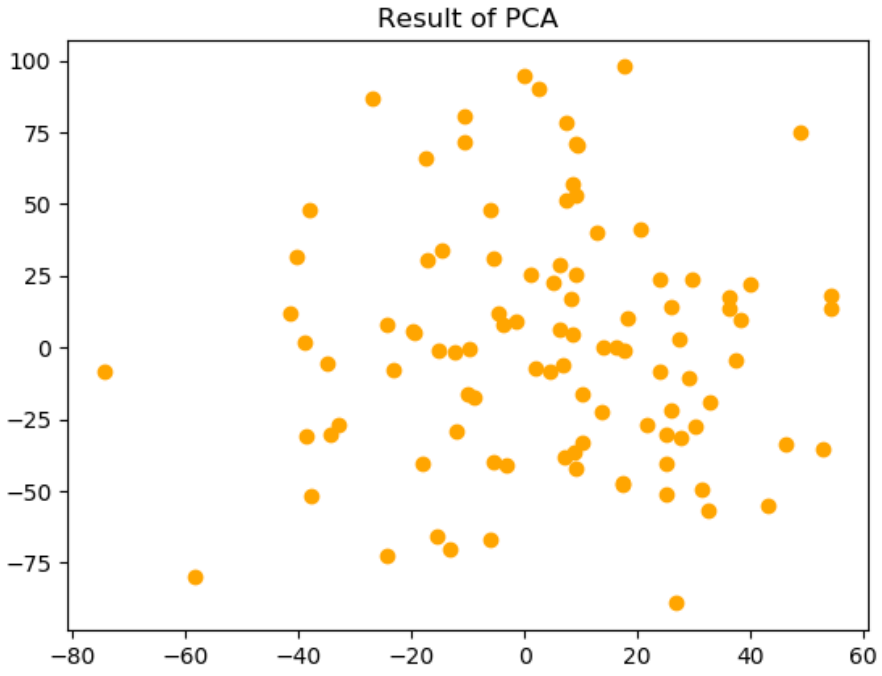


Figure 2

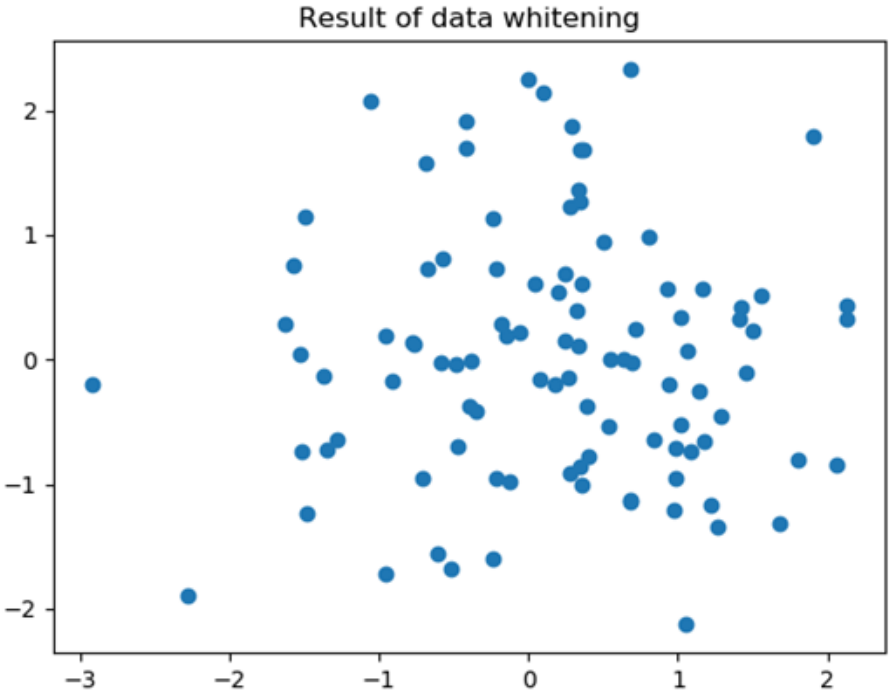


Figure 3

For my result, when original data set, X, has features with widly varying range and variance, I do not want to perform whitening on data. Figure1, Figure 2, Figure 3 plot original data, data after PCA and data after whitening. Figure 4, Figure 5, Figure 6 shows the covariance of the original data, after PCA and after data whitening. Covariance after PCA is a diagonal matrix and covariance after data whitening is almost an identity matrix. I do not want to perform data whitening because my data does not perform well after whitening, in other word it almost has no difference with original data. Also , it still has very small covariance and not a perfect identity matrix.

```
In[6]: print(cov)
[[1000, 500], [500, 2000]]
```

Figure 4

```
In[3]: print(covpca)
[[6.53011049e+02 1.03633768e-13]
 [1.03633768e-13 1.76491439e+03]]
```

Figure 5

```
In[4]: print(covwhitendata)
[[1.00000000e+00 6.43321909e-17]
 [6.43321909e-17 1.00000000e+00]]
```

Figure 6