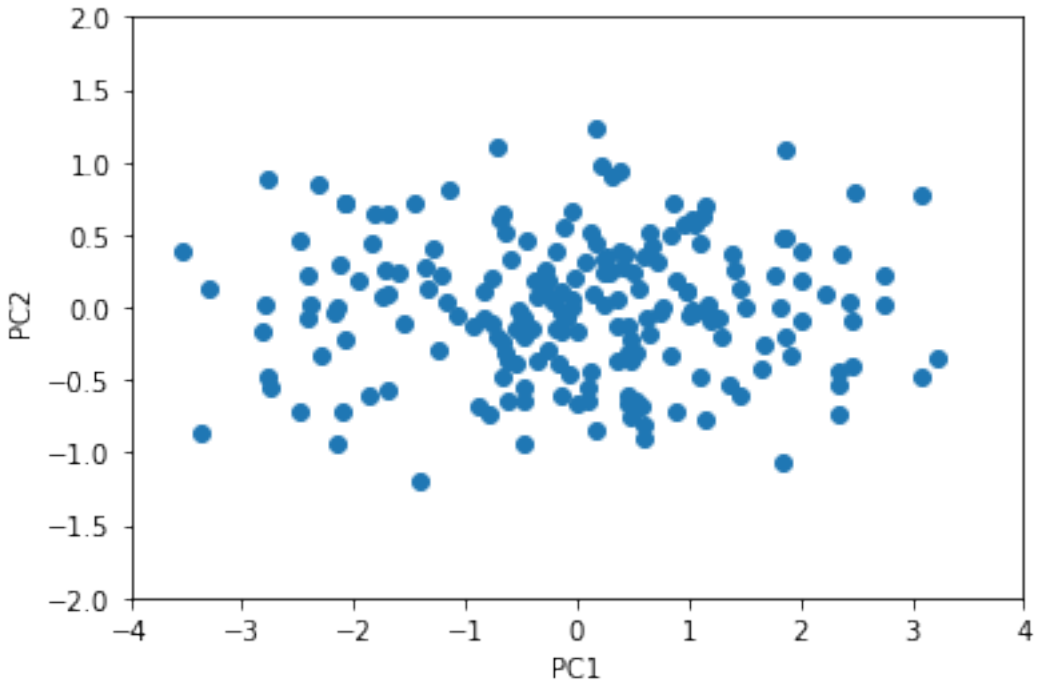
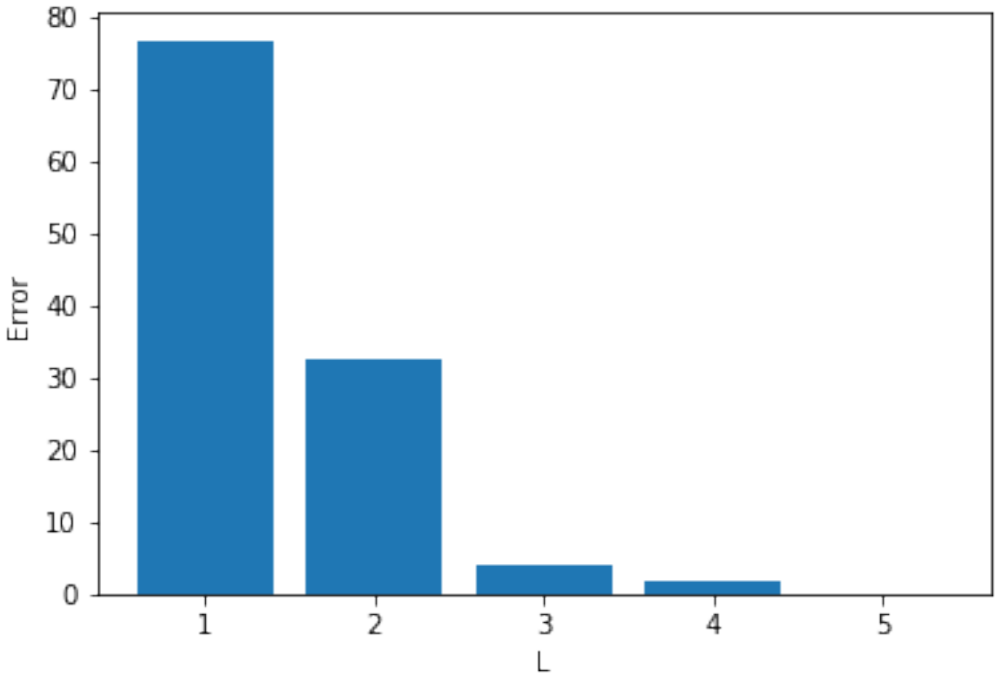
Computer

March 21, 2019

```
In [91]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         from scipy.linalg import eigh

In [80]: data = np.loadtxt('data.csv', delimiter=',')

In [81]: print(data.shape, ' Are D and N respectively')

(200, 5)  Are D and N respectively

In [82]: covariance = np.cov(data.T)

In [121]: evals, evecs = np.linalg.eig(covariance)

In [158]: print(len(evals))
          print(evals)

5
[2.01264957 0.22286198 0.14290211 0.01042867 0.00925912]

In [137]: xHat = np.matmul(data, evecs[:,:2])

In [174]: plt.scatter(xHat[:,0], xHat[:,1])
          plt.xlim(-4, 4)
          plt.ylim(-2, 2)
          plt.xlabel('PC1')
          plt.ylabel('PC2')
          plt.plot()

Out[174]: []
```




```
In [171]: ww = list(map(lambda x: np.matmul(evecs[:,:x +1], evecs[:,:x + 1].T), range(len(evals))))
          wwx = list(map(lambda x: np.matmul(data, ww[x]), range(len(evals))))
          xsum = list(map(lambda x: sum(sum((data - wwx[x])**2)), range(len(evals))))
          xsum

Out[171]: [76.70497237116591,
           32.35543842531078,
           3.917904165714545,
           1.8425906022674978,
           2.1995217574872173e-28]

In [172]: plt.bar(range(1, 6), xsum)
          plt.xlabel('L')
          plt.ylabel('Error')
          plt.show()
```



The first two eigenvalues account for nearly all of the variance that exists in the data.