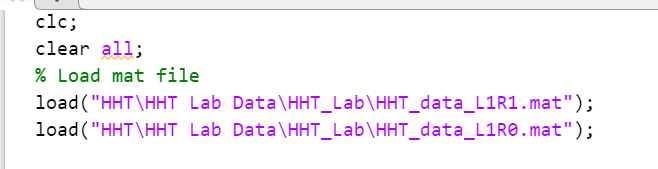
110522130 資工碩一 李信鋌

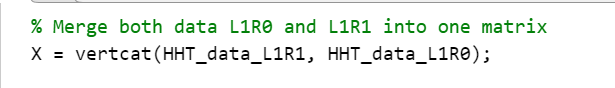
110526005 資工碩一 林季陽

資料科學實務 – Lab Activity PCA

Code We have written

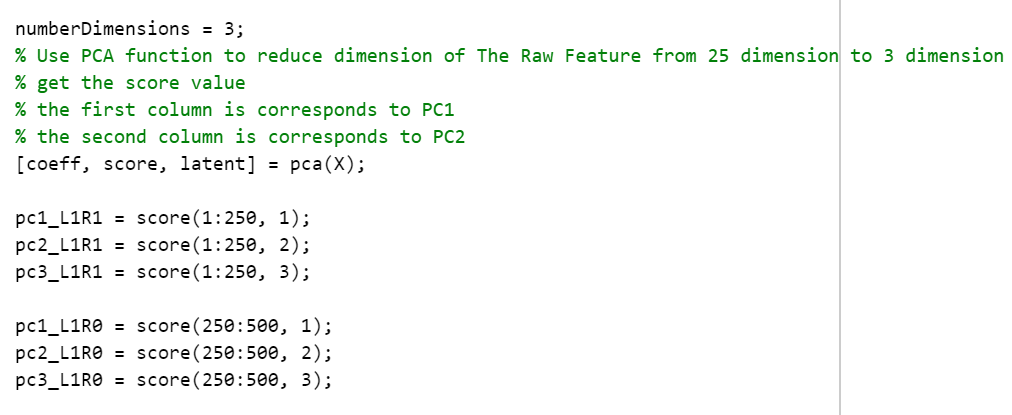


Load the data of L1R1 and L1R0. Dimension of those data are 250 \* 27. (Data size is 250, and dimension of data is 27)



Merge both data into one matrix, dimension of X is 500 \* 27 after concatenation.

(Data size: 500, dimension: 27)

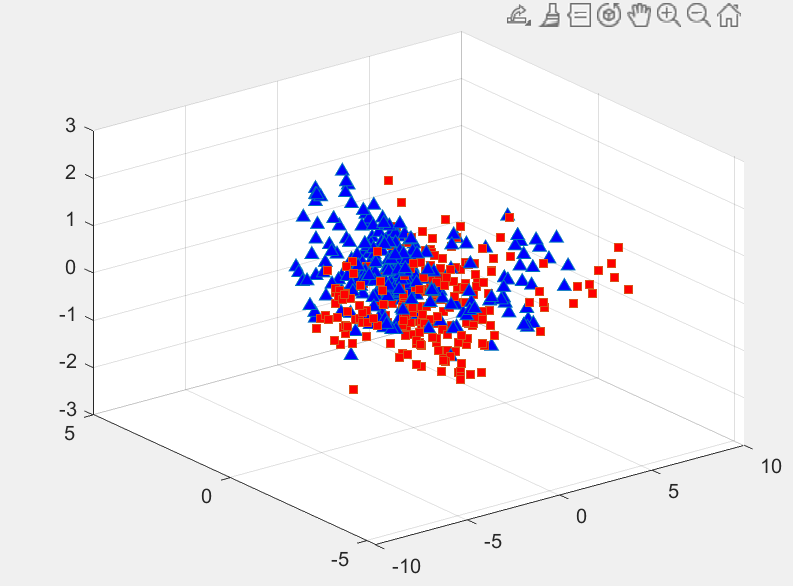


Implement Dimension Reduction with PCA. Reduce dimension from 27 to 3.

Coeff: the principal component coefficients, loadings for the n-by-p data matrix.

Latent: principal component variances (eigenvalues of the covariance matrix of X)

Score: Principal Component Scores



The upper chart is the result.

It seems that the most important features of data are captured after implement PCA as those data are very close to each other.

It can prove that PCA can reduce dimension successfully because the variance is not reduce much.

The basic assumption of principal component analysis is that the data are expected to find a projection axis (vector) in the feature space that will give the maximum variation of the data set.

Therefore, the PCA program is trying to find a line to do projection within the data.

After the projection is done, PCA is trying to find the projection vector that maximizes the variance of the projected data.

In other word, it means that we need to solve and find out the eigenvalue and eigenvector in covariance matrix.