Istanbul Technical University –Fall 2018

BLG527E – Machine Learning

Homework-2



504161354 – Merve Turhan

504161354\_HOMEWORK-2



There are four main file in this homework.

1) Q1-Multivariate Analysis

2) Q2 - Dimensionality Reduction – PCA

3) Q2 - Dimensionality Reduction – LDA

4) Q3 - Clustering

Solutions Summary :

Q1) : Training\_error = 0.03740517917865549

Test\_error = 0.06121313299944353

Q2) : Test\_error\_PCA 0.4752365052865888

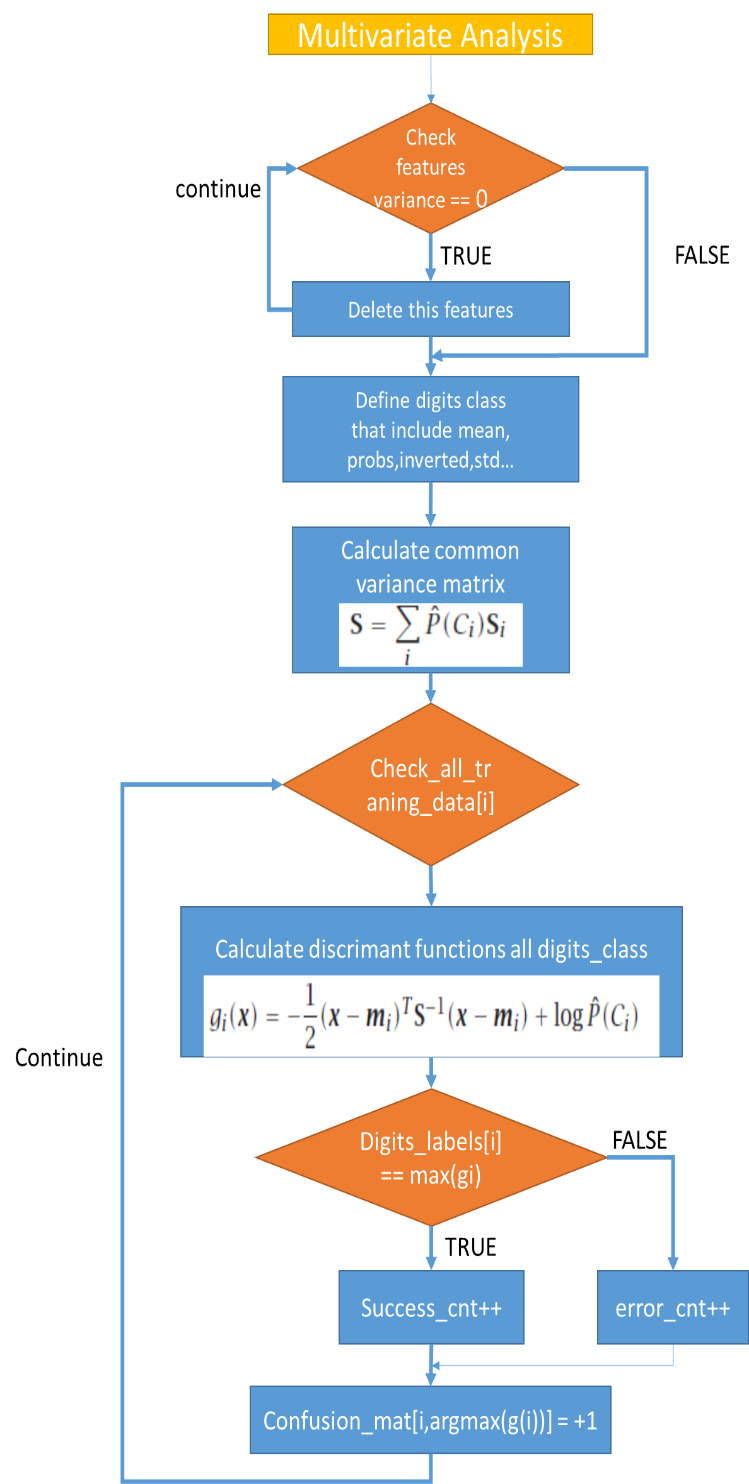
Test\_error\_LDA = 0.4062326099053979

Q3) : L1 = 0.0986136541982736

L2 = 0.08396547214229663

**Q1 ) Multivariate Analysis :**

The fact that the variance is zero shows us that it has no information in these features. So features have zero variance should be deleted. Multivariate Method software diagram is as follows.





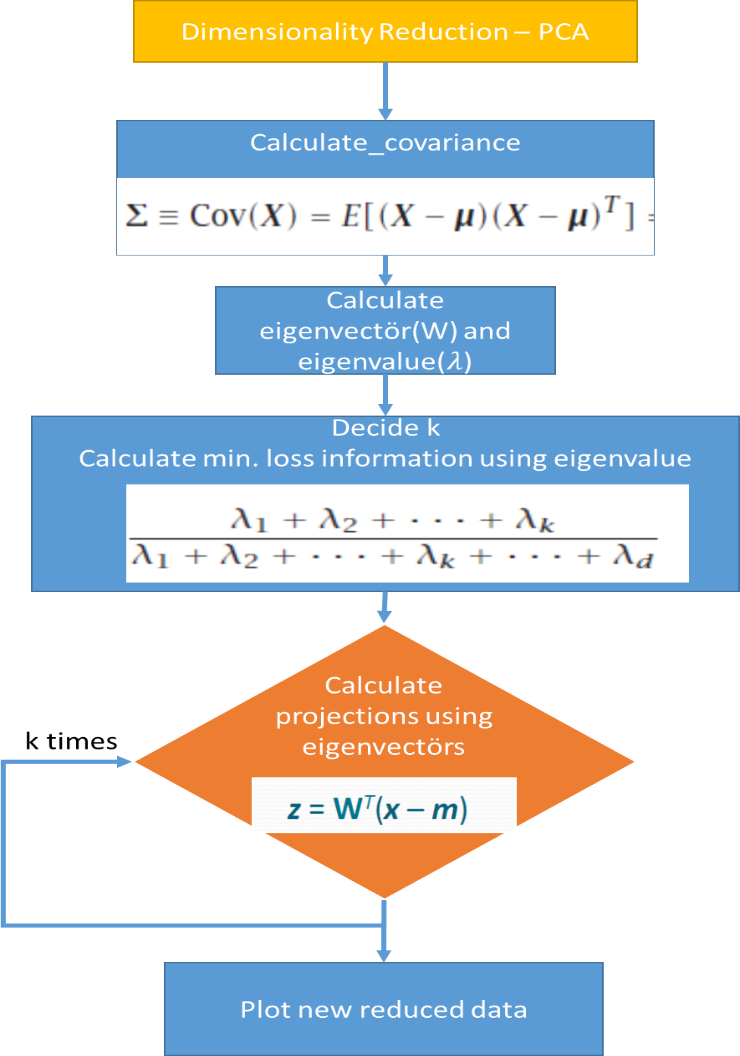
**Table 1 Traning Confusion Matrix**

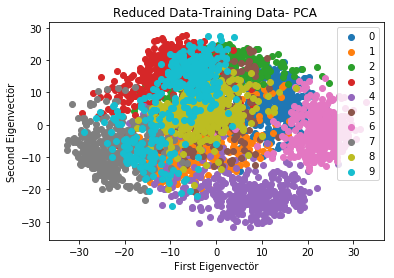


Table 2 Test Confusion Matrix

**Q2.a) Dimensionality Reduction – PCA**

Principial compenent analisis software diagram is as follows. k = 2 is fixed in this question.





**Q2.b) Dimensionality Reduction – LDA**

1. Between-class scatter matrix and Within-class scatter matrix are calculated using these formulas.



1. The largest eigenvalues and eigenvectors are calculated this equations.



1. And two largest eigenvectors are used

Figure 1 Traning Data after LDA

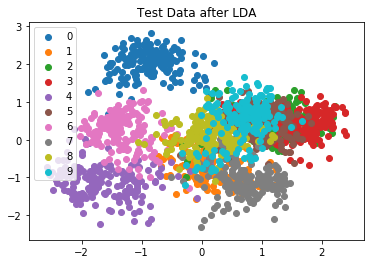
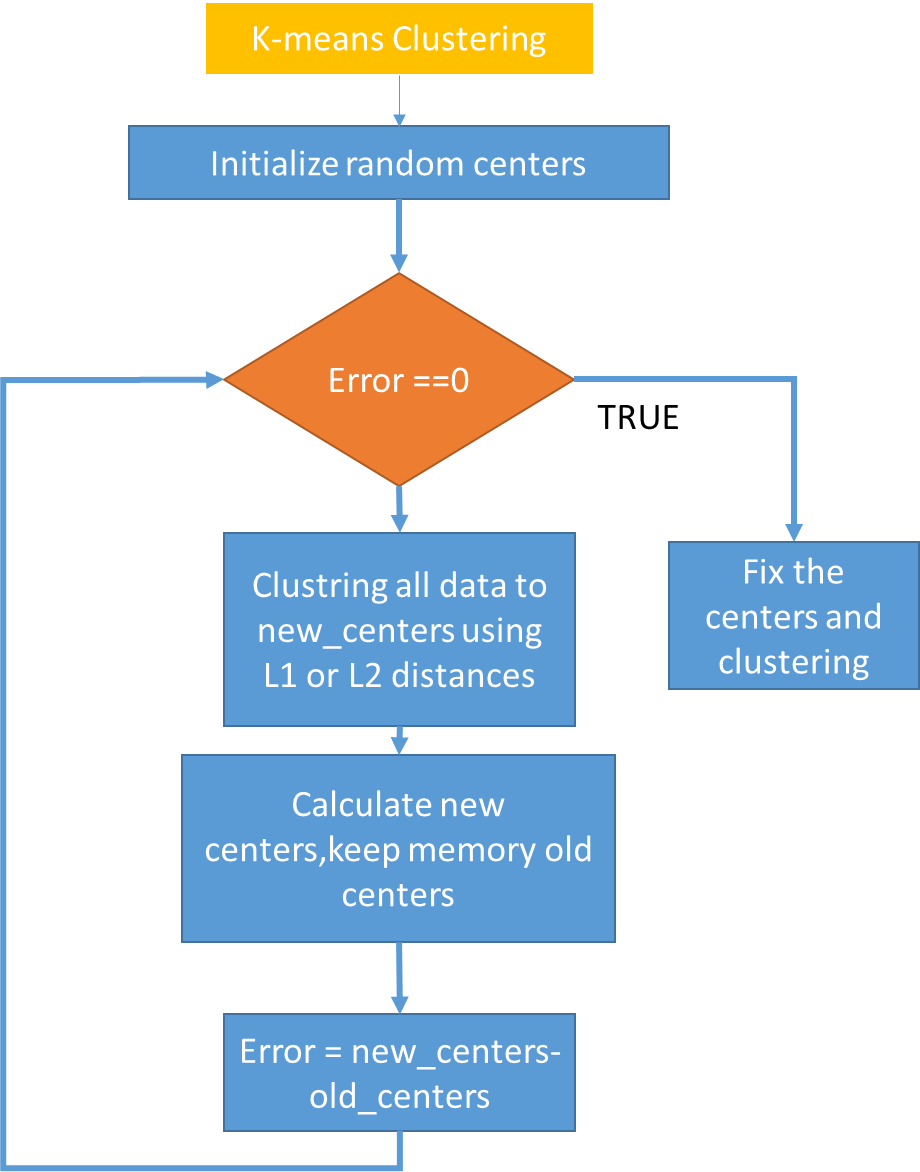


Figure 2 Test Data after LDA

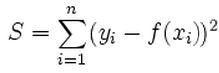
Q3) **[Clustering]**

K-means clustering software diagram is as follows. For this questions , k equals 20.

****

Two diffent measures is used in this questions. Using L2 distances is better than L1 distances.

http://www.chioka.in/wp-content/uploads/2013/12/l1-norm-formula.pngL1\_distances =



L2\_distances