姓名:王國倫　　學號:0860077　　Homework4

Problem1:

The detail code in HW4_0860077_王國倫(1).ipynb

(a) I use the testing dataset for training with k-means algorithm, and validate the model with testing dataset, obtain accuracy 59.82%. Before calculate the accuracy, the data required to transform, because the k-means labels are not real target labels, finding relationship with k-means labels and target labels, and transform new labels.

(b) In order to show data distribution, I apply principal component analysis (PCA) method to decrease dimension from 784 to 2, and visualize to 2D space, just like below two pictures, the upper one is origin data distribution, and lower one is clustered data distribution.

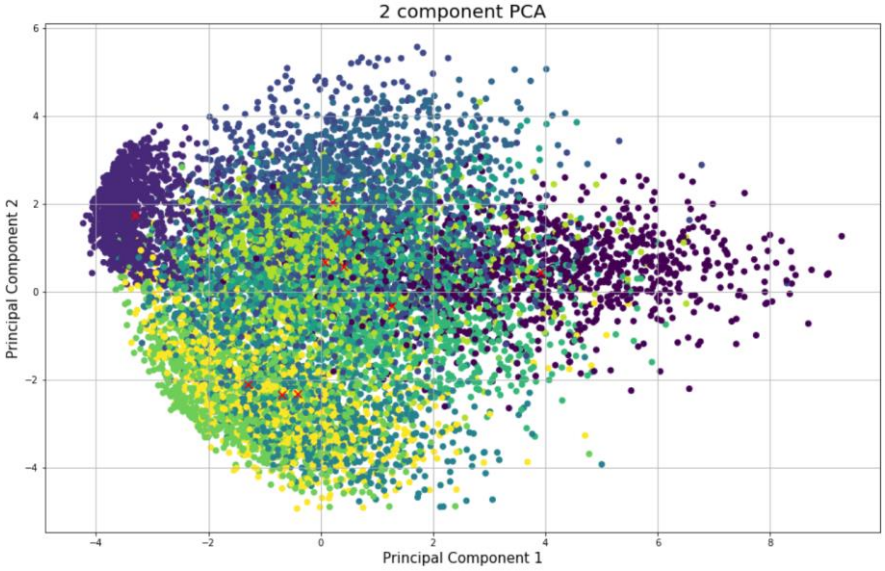


Fig.1 origin data distribution

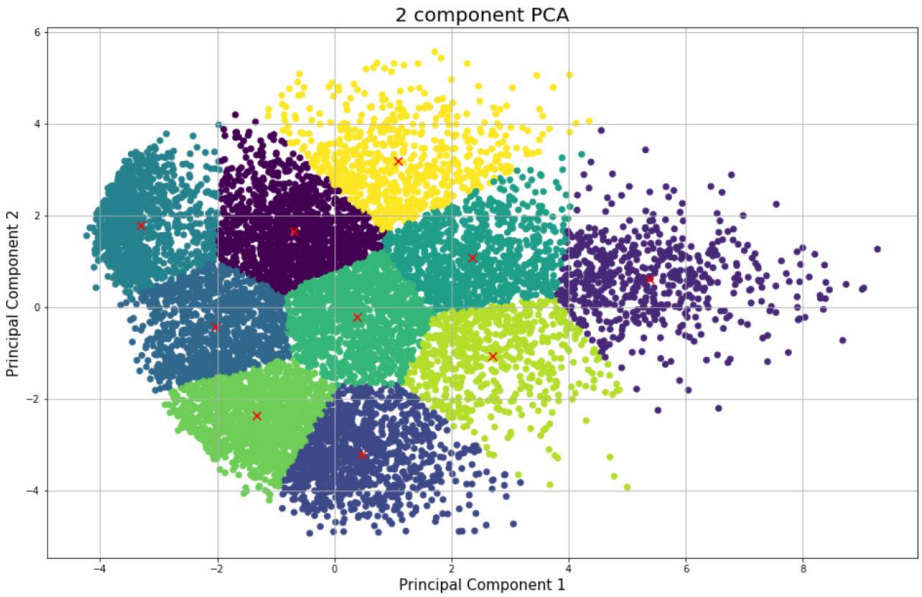


Fig.2 clustered data distribution

Problem2:

The detail code in HW4_0860077_王國倫(2).ipynb

(a) The Fig.4 show that reconstruction results of different dimension. Compare with origin image, the all reconstructed picture is digit 4, the dimension more higher more clearer. The reason is adding the number of dimensions eigen vector with higher eigen value, so the image is so complete with higher dimension. Although higher dimension response clear images, lower dimension also indicate that the fewer parameter and keep important information.

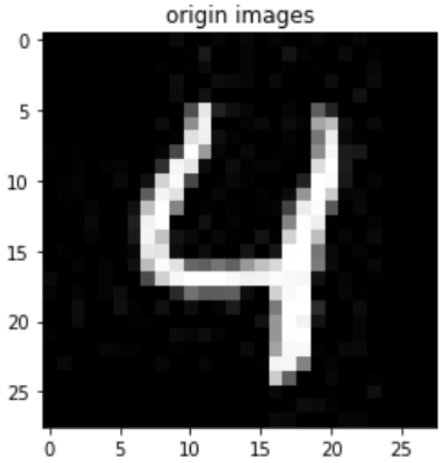


Fig3. Origin image

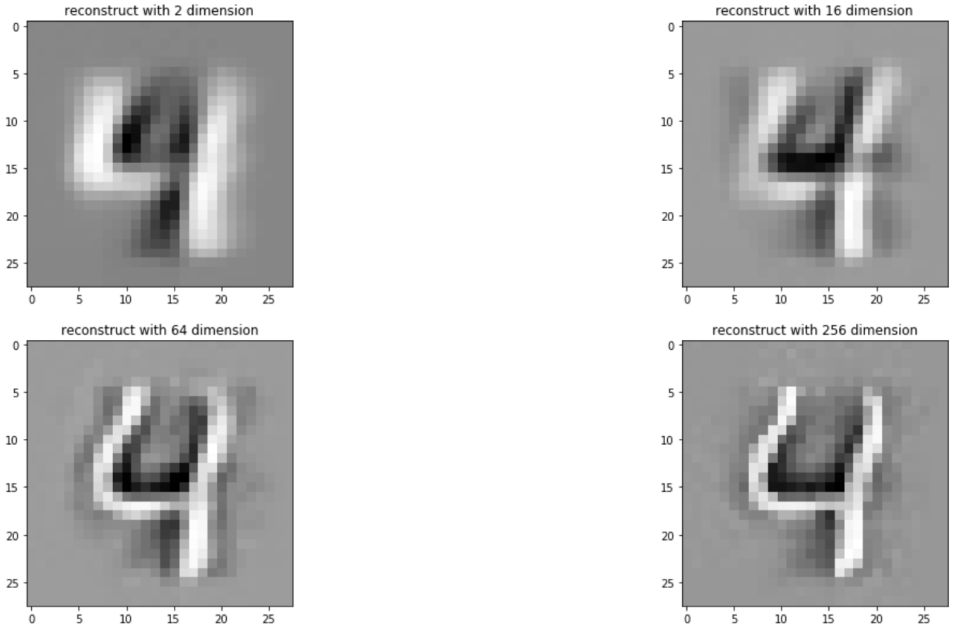


Fig.4 reconstruction with different dimensions

(b) The Fig.5 show that digit 4 projection as a 2D scatter plot.

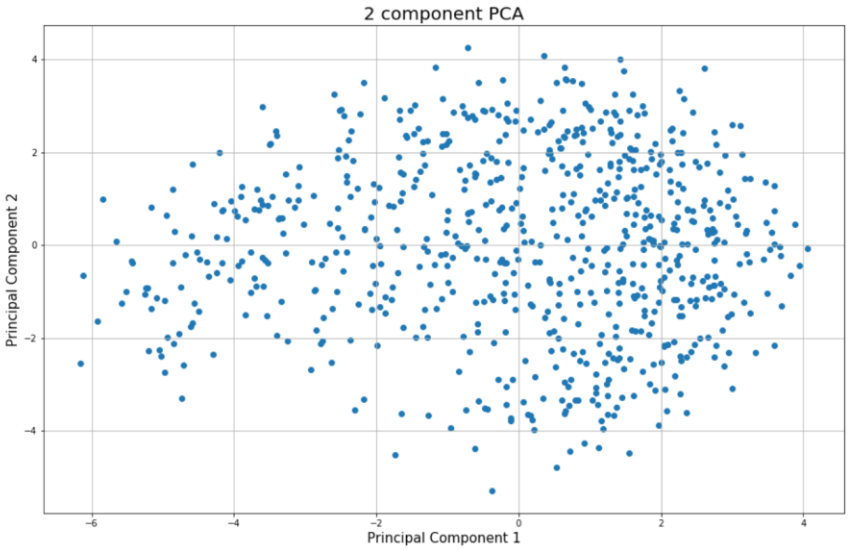


Fig.5 digit 4 distribution

(c) Linear Discriminant Analysis (LDA) is a supervised learning, compare with Principal Component Analysis (PCA) is an unsupervised learning, but these are same aim to reduce dimension. The target of dimension reduction is project origin data to other vectors with higher variance, apart from this, the LDA also take different categoric into consider. As a result, the LDA have better performance than PCA.

(d) The objective function of PCA and LDA show as following formula (1) and (2),

$$v = \arg\max \; v^T C v \qquad (1)$$

$$v = \arg\max \; v^T (S_B {S_w}^{-1}) v \qquad (2)$$

where C is a covariance matrix, $S_B$ is between-class scatter, $S_w$ is within-class scatter. These are same target to find eigenvector, but the matrix is different, the C is all origin data, no matter categoric of data. However, the aim of LDA is maximum the $S_B$, which the scatter between different class, and minimum the $S_w$ , which scatter of the same class. LDA will consider relationship between different classes instead of all data.