PRINCIPAL COMPONENT ANALYSIS - REPORT

Principal component analysis (PCA) is a technique to overcome high dimensional data. It helps to classify the data by using lower dimensional data with minimum loss. Here, PCA is used for digit classification with given principal components. The data has 3000x256 dimension. Here, PCA is used with eigen decomposition. By using d values 50, 100, 200 and 256 as principal components, images are reconstructed. Therefore, the images after and before applying PCA can be compared and can be controlled whether the digits are still classified and distinguished.

As an example, let look at digit 2 after applying PCA.

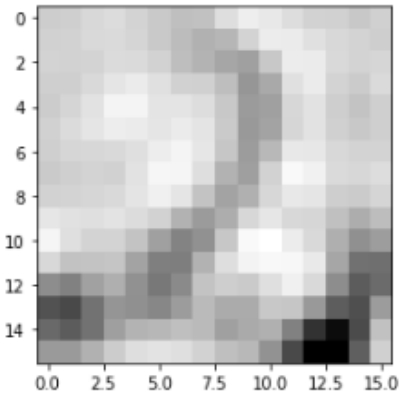

image1: d = 50

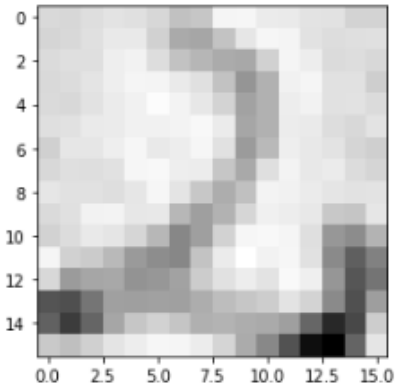

image2: d = 100

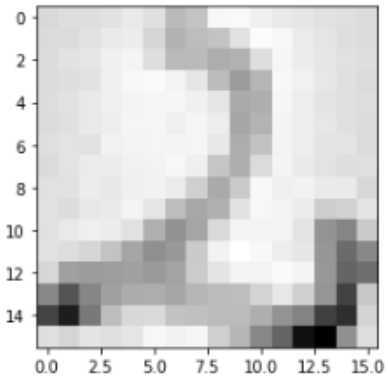

image3: d = 200

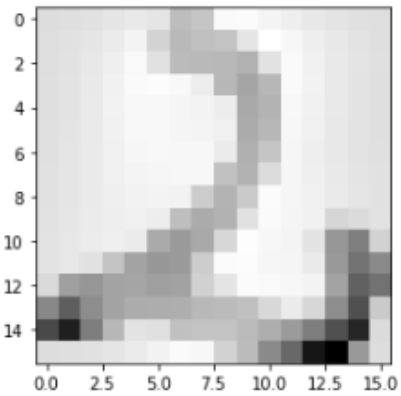

image4: d = 256

According to images above, some image information is lost. When d is equal to 50, it is difficult to classify the digit. However, all images can still be classified as digit 2.

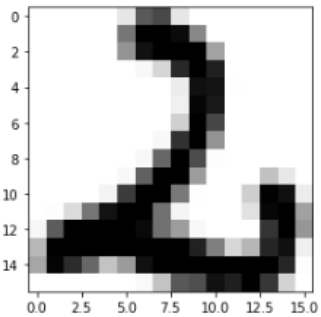

image5: original data

Before applying PCA, digit 2 is seen as in image5. After applying PCA, it is understood that the digits are still classified and distinguished. Since the data has loss of information after applying PCA, it is difficult to classify the digits according to different d values. If minimum loss is intended, then d is chosen as 256. However, for all given d values, the digits are still classified.

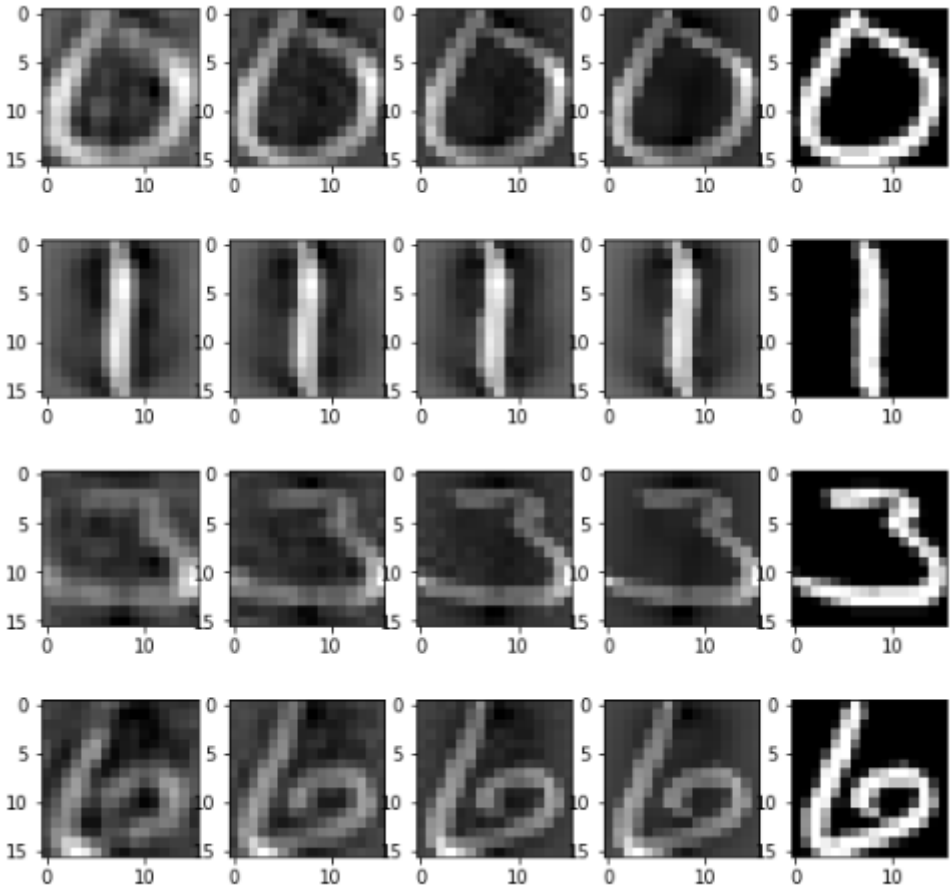

image6: rows show indices i=0,500,1000,2000 and columns show principal components d=50,100,200,300 and last column shows original data, respectively.

In the last part of the task 2, digits are shown according to given different indices and given d values. In image6, digits 0, 1, 3, and 6 can be compared with different d values and original data. As is seen, when d is equal to 50, it is difficult to classify them but PCA still works to distinguish the digits for given d values.