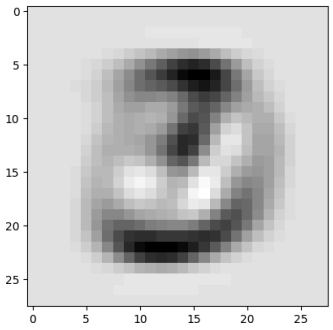
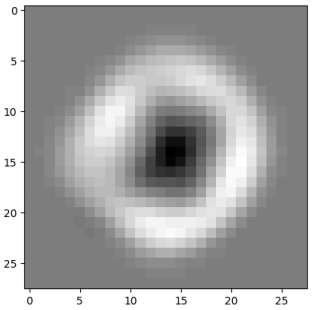
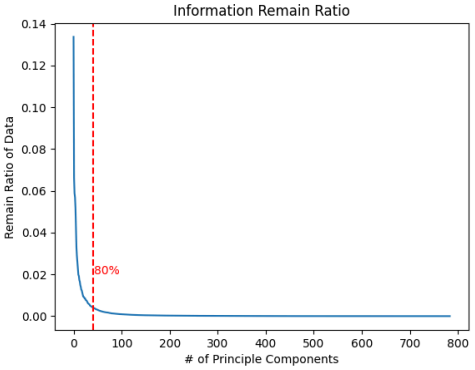
Name: Nguyen Dang Hoang Khang
Student ID: 109006269

1. Describe what problems you encountered and how did you solve them when implementing the basic and advanced functions
   The problem I encountered was choosing the appropriate mean for data centralization. I initially thought that we had to choose the mean of the training data before splitting, but this was not the case because the results were not correct. So I tried to use all the means and I found that only the mean of the training data after splitting produced the correct answer.
2. Briefly describe how you implement the basic PCA
   In step 1, I centralize the data by subtracting the mean. In step 2, I calculate the covariance matrix, which represents the relationships between different features. In step 3, I compute the eigenvectors and eigenvalues of the covariance matrix using np.linalg.eigh(), then I sort the eigenvectors in descending order based on their corresponding eigenvalues for the selection of top k eigenvectors to form the projection matrix W in step 4 and 5. Finally in step 6, I project the original data onto the new feature subspace using the multiplication between projection matrix W and the input matrix
3. Plot eigenvectors (right image) and the reconstruct image (left image)





4. Plot the distribution of eigenvalues



5. Introduce the preprocessing/PCA methods you implemented in the advanced part
   After some testing, for the advanced ranking part, I decided to use the basic PCA rather than sparse PCA, with 400 components and no further data preprocessing.