Principal Component Analysis

- **Objective**

  To implement Dimensionality reduction using PCA. Analyse the reconstruction error for various values of k. Display the EigenVectors.

- **Procedure**

  1. Choose an appropriate dataset.
  2. Construct the covariance matrix with the smallest dimension possible.
  3. If the number of examples are greater than the dimension of the feature set then
     a. Find the n (reduced dimension of the feature set) greatest eigenvalues and corresponding eigenvectors.
     b. Construct the projection matrix transpose(U) from the obtained eigenvectors and the reconstruction matrix U.
  4. If the number of examples are less than the dimension of the feature set then
     a. Find the n (reduced dimension of the feature set) greatest eigenvalues and corresponding eigenvectors.
     b. Divide all eigenvectors by their norm.
     c. Construct the projection matrix transpose(U) from the obtained eigenvectors and the reconstruction matrix U.
  5. Obtain the images for the projected matrix and calculate the reconstruction error.
  6. Repeat the above steps for different values of n.

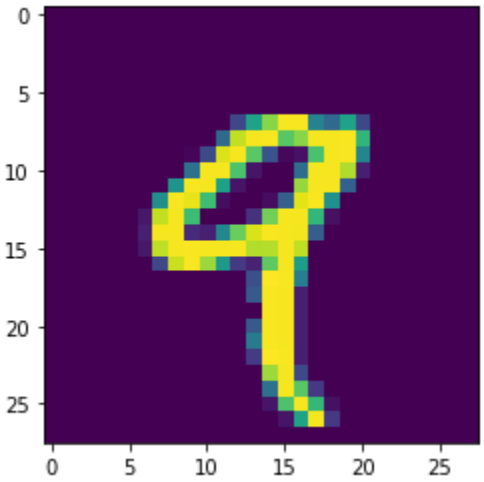
- **Observations**

  ➢ **Dataset Parameters**

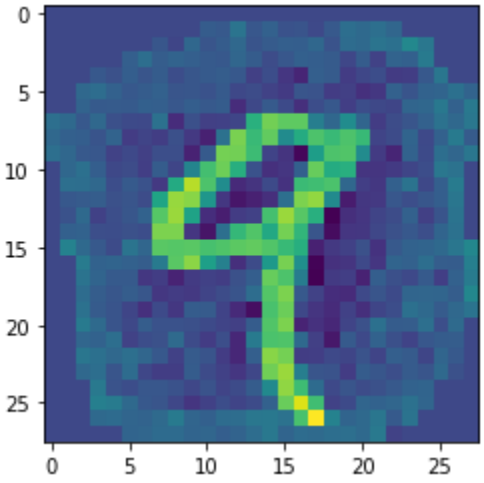
  - The dataset contains the images of handwritten digits from 0 to 9.
  - There are a total 4,000 images in the dataset.
  - Each example is represented by 784 features.

➢ **Comparison of images for different values of n**

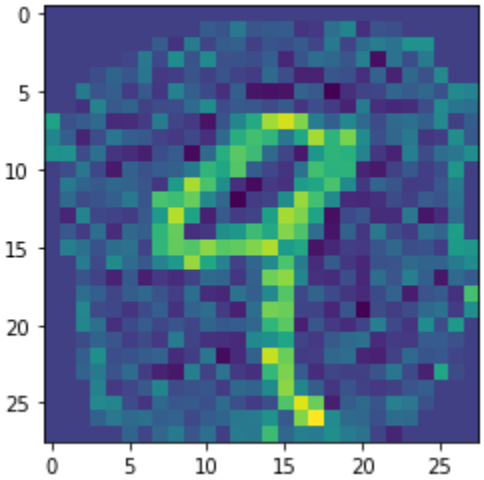
- The Original Image is shown below:



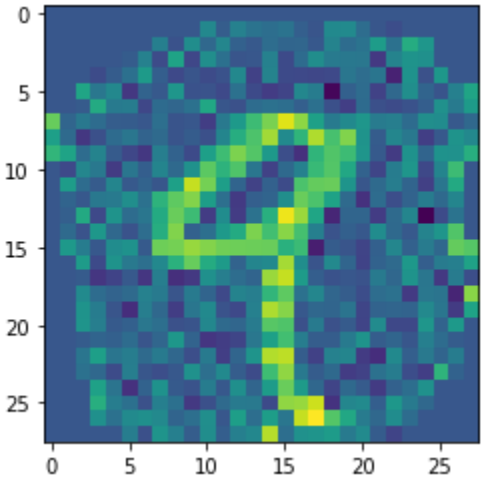
- The image after reducing the dimension to `750`



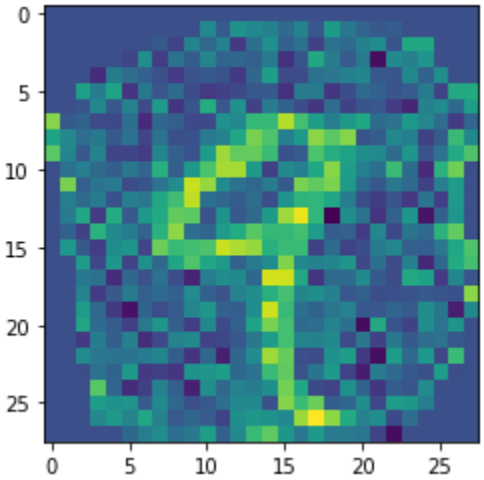
- The image after reducing the dimension to `600`



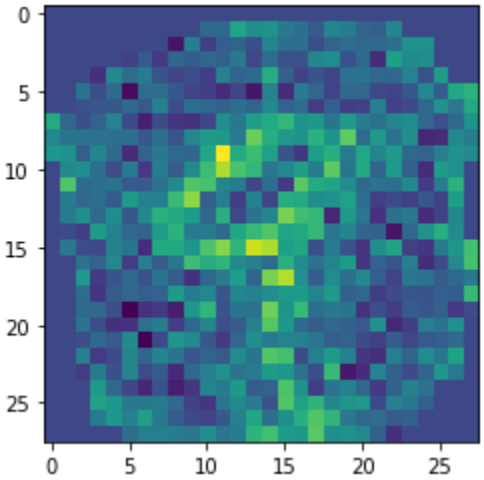
- The image after reducing the dimension to `500`



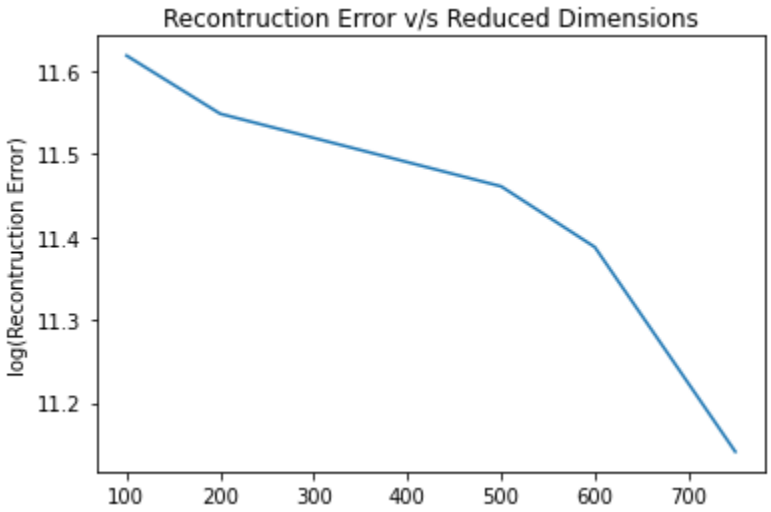
- The image after reducing the dimension to `200`



- The image after reducing the dimension to `100`



➢ **Variation of reconstruction error with n**



Conclusion

The reconstruction error increases with decreasing the value of the final required dimension of the feature set as some information is always lost per dimension reduction.