Introduction to Machine Learning

許木羽 / 111000177

Problems Encountered

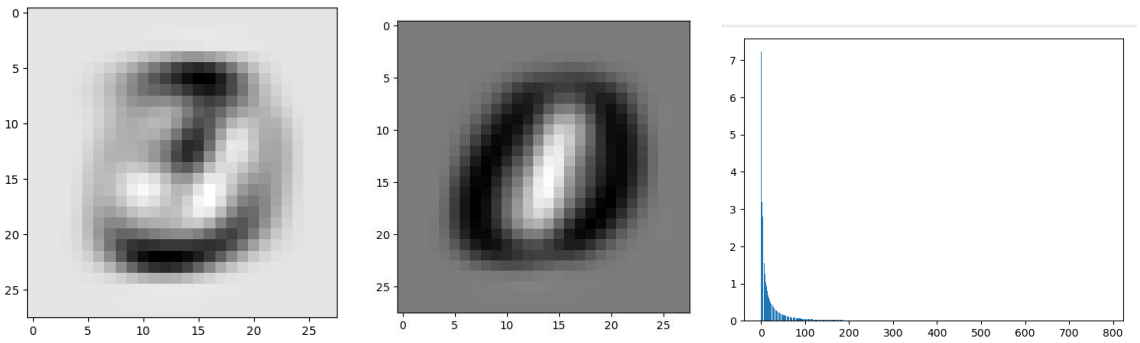
Have so many issues with the dimension's matrix, as we need to be careful with the inverse and know how the NumPy eigenvector does, etc. works. Also, we need to be careful as eigenvector is a column here.

Implement PCA

By using the given formula, we find the centralized data, then find the covariance matrix by the given formula = 1/n * C.T @ T, as C centered matrix. Then we find the eigenvector and sort it by eigenvalue. For reconstruct data, there were so many trials and errors as my lack of understanding. To transform, we need to find the center based on average x train, then multiply by the components, which top k eigenvectors.

Image of Reconstruct Image, Eigenvectors, and Eigenvalues

Reconstructed Image (left), Eigenvectors (middle) and Eigenvalue (right).



Data Preprocessing in Advance Part

My thought is I will not use PCA or SPCA since it removes the data input and may have some lost information. But one thing that I'm sure of is that augmentation data plays a very important role here. So, I did median filter for the data, then I used threshold, so the input normalizes become either 0 or 1. Then to make the data larger and balanced, I augmented the data, and less data for specific label is, the more I will augment them.