İpek Öztaş
22003250
CS464-Sec2

CS464 Introduction to Machine Learning

Fall 2023

Homework 2

İpek Öztaş

22003250

Section-2

PCA

Q1.1

```
proportion of variance explained (PVE)
PVE for k=1: 0.09704664368151002
PVE for k=2: 0.07095924066613628
PVE for k=3: 0.061690887728496284
PVE for k=4: 0.053894194943091854
PVE for k=5: 0.04868797012682379
PVE for k=6: 0.04312231323349435
PVE for k=7: 0.03271929951323684
PVE for k=8: 0.02883895448304446
PVE for k=9: 0.02762029396742109
PVE for k=10: 0.0235700055115963
```

Figure 1 PVE for First 10 Principal Components

İpek Öztaş
22003250
CS464-Sec2

Q1.2

```
PC for k=1 0.09704664368151002
PC for k=2 0.1680058843476463
PC for k=3 0.22969677207614259
PC for k=4 0.2835909670192344
PC for k=5 0.3322789371460582
PC for k=6 0.3754012503795526
PC for k=7 0.4081205498927894
PC for k=8 0.43695950437583386
PC for k=9 0.46457979834325497
PC for k=10 0.4881498038548513
PC for k=11 0.5092416999945412
PC for k=12 0.5294716108091951
PC for k=13 0.5466297928004837
PC for k=14 0.5635509071169658
PC for k=15 0.5793373199278713
PC for k=16 0.5941668460731245
PC for k=17 0.6074124555500645
PC for k=18 0.6201814292345892
PC for k=19 0.632054058620294
PC for k=20 0.6435808956319572
PC for k=21 0.6542425577029631
PC for k=22 0.6643096914142529
PC for k=23 0.6738454246873816
PC for k=24 0.6829708621325816
PC for k=25 0.69180490903893
PC for k=26 0.7001981004236029
PC for k=26 is the min number of principal components to explain the 70% of the data.
```

Figure 2 Cumulative Proportion of Variance

Q1.3

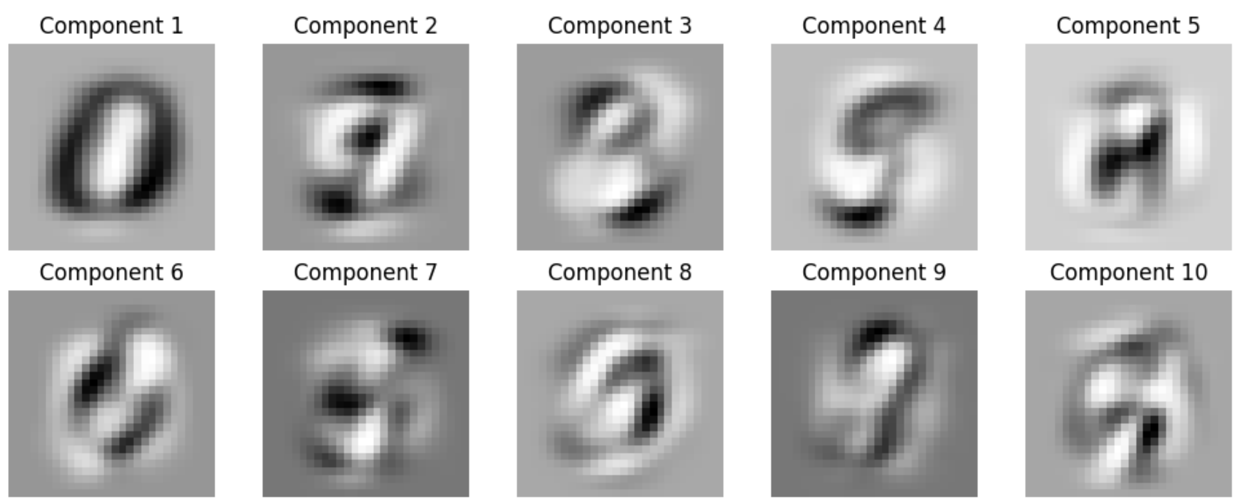


Figure 3 Grayscale Principal Component Images of Size 28 × 28

İpek Öztaş
22003250
CS464-Sec2
In Figure 3, first ten principal components are shown. We can observe some patterns like circular shapes or curves as well as intersections of different digits.

The initial principal components are comprehensive representations of dominant features within the dataset. The components captures features such as curves and loops which define the overall structure. The oval shape of Component 1 indicates ‘0’ whereas the third component highlights areas significant for '3' or '8'.

The interpretability diminishes as we move to higher components, reflecting variations less common in the dataset. The visualizations are valuable for understanding significant features, aiding in informed decisions regarding feature selection and model complexity.

The first images, corresponding to the first principal components, exhibit clear and recognizable patterns, while the interpretability diminishes for higher-order components.

İpek Öztaş
22003250
CS464-Sec2

Q1.4

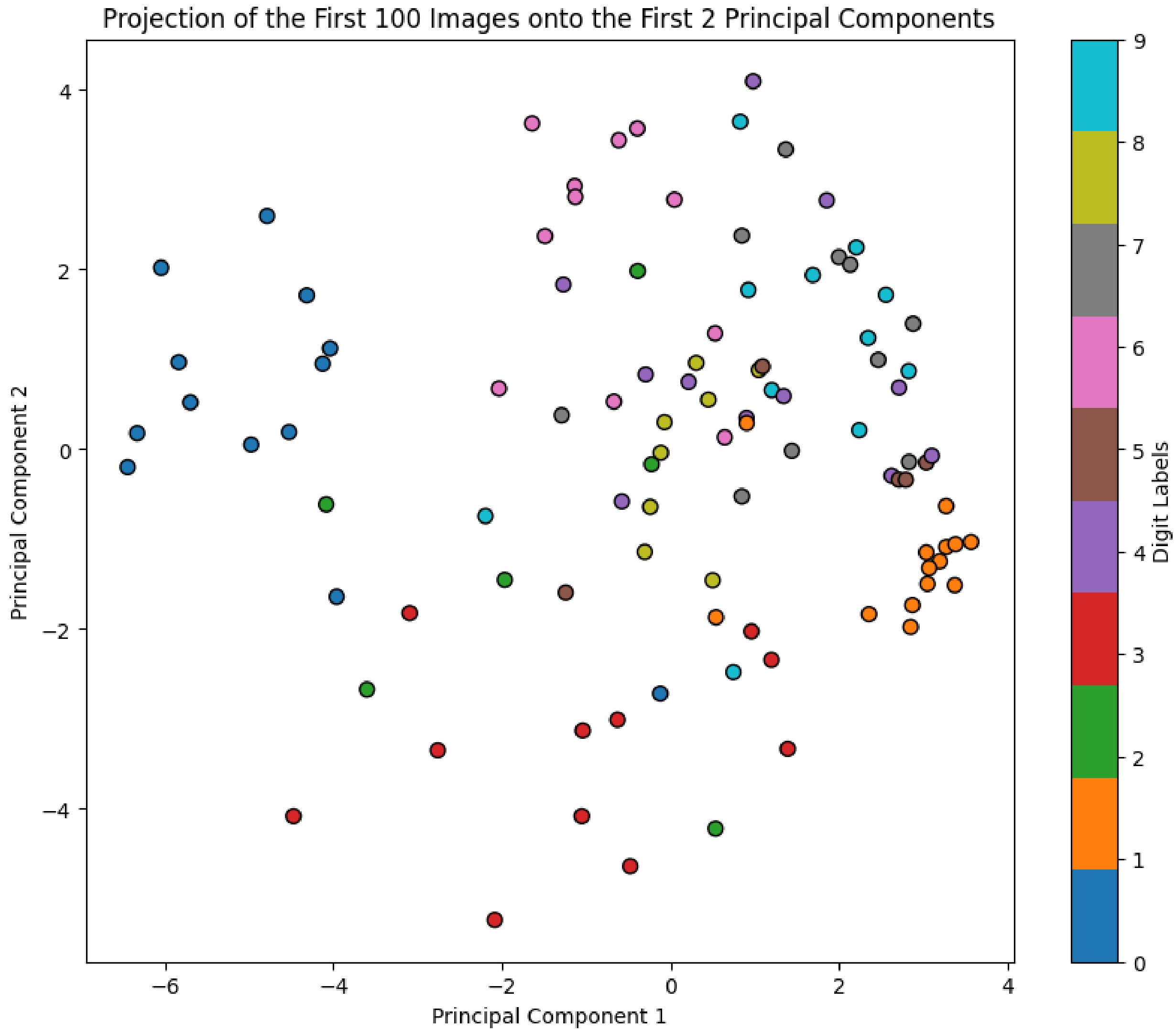


Figure 4 Projected Data Points

In Figure 4, the scatter plot indicates that the first two main components capture crucial information for differentiating between digits and shows that there is natural clustering depending on digit classifications. Effective classification is made possible by closely clustered points of the same hue, which show shared properties among the digits.

Notably, the first two primary components' unique patterns that resemble the numbers "0" and "9" allow for their accurate classification. Overlapping dots, on the other hand, indicate difficulties in distinguishing specific digits and highlight the necessity of more dimensions for more distinct separation.

The variation that each component captures is reflected in the distribution of points along the primary component axes. Presumably, significant features are represented by the first principal component, and less prominent attributes are captured by the second. All things considered, the clustering supports the characteristics that the first principle components identified, which is consistent with the findings from the earlier question (1.3).

Q1.5

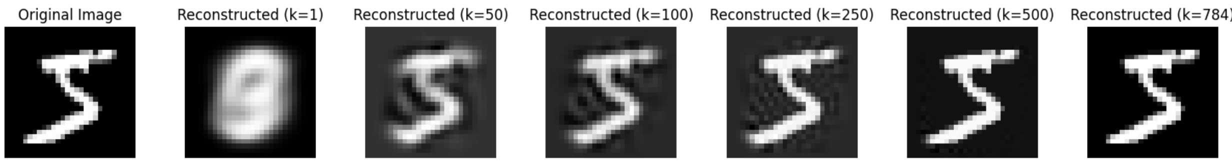


Figure 5 : Reconstructed Image with k Principal Components

Figure 5 demonstrates the reconstructed images. It is done by the following two steps:

1- Projection onto Principal Components:

Subtract the mean image from the original image to center it. Project the centered image onto the first k principal components, capturing the essential features of the digit.

2- Reconstruction from Projected Data:

Multiply the projected data by the transpose of the matrix of selected principal components. Add back the mean of the original data to reconstruct the image.

The amount of detail in the reconstructed image is greatly influenced by the selection of the number of components, k. More k values result in a reconstruction that is more precise and thorough.

It is evident that the image is undisguisable when k = 1 since the reconstruction captures only the most dominant feature. As a result, it can be claimed that a digit cannot be distinguished without the presence of many major components. As k increases, reconstruction quality gets better because it progressively includes more details. The reconstructed image is identical to the original image when all of the primary components are used (k=784).

Logistic Regression

Q2.1

```
Test Accuracy: 0.9063
```

Figure 6 Test Accuracy of Default Model

İpek Öztaş
22003250
CS464-Sec2

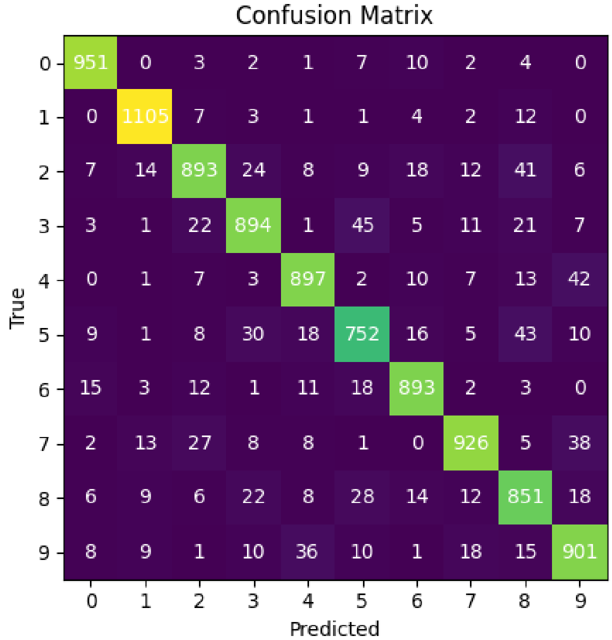


Figure 7 Confusion Matrix of Test Set

,

Q2.2

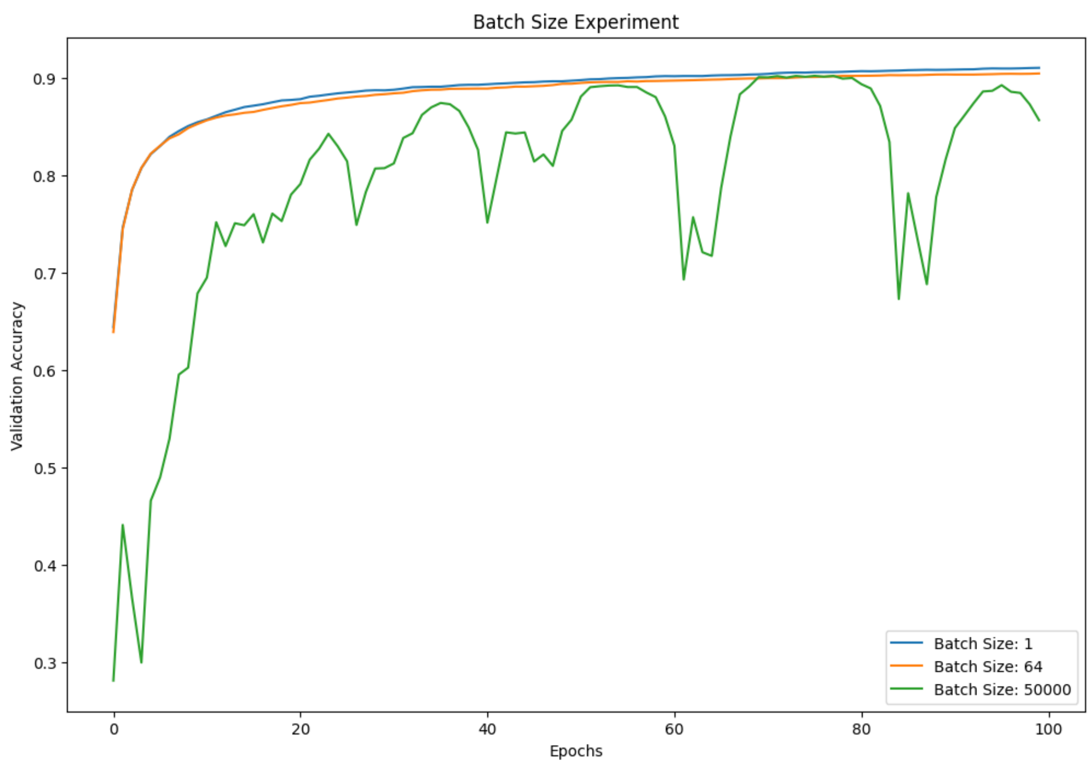


Figure 8 Validation Accuracy with Batch Sizes

İpek Öztaş
22003250
CS464-Sec2

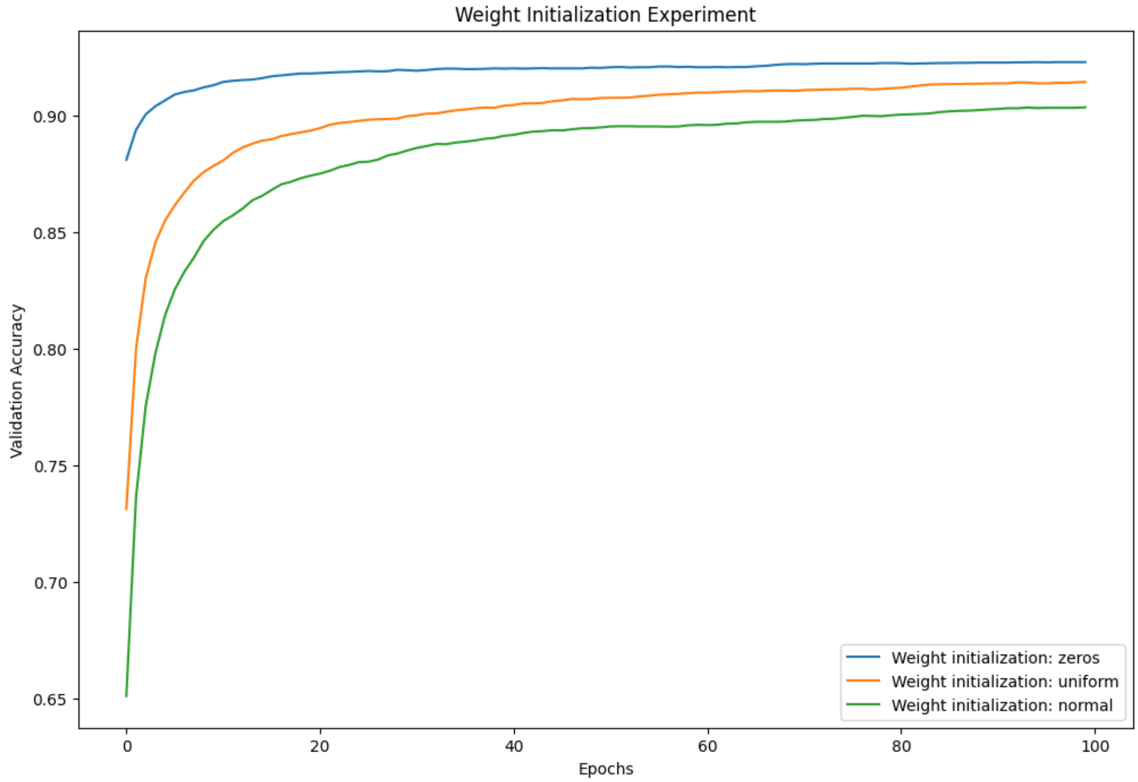


Figure 9 Validation Accuracy with Different Weight Initialization

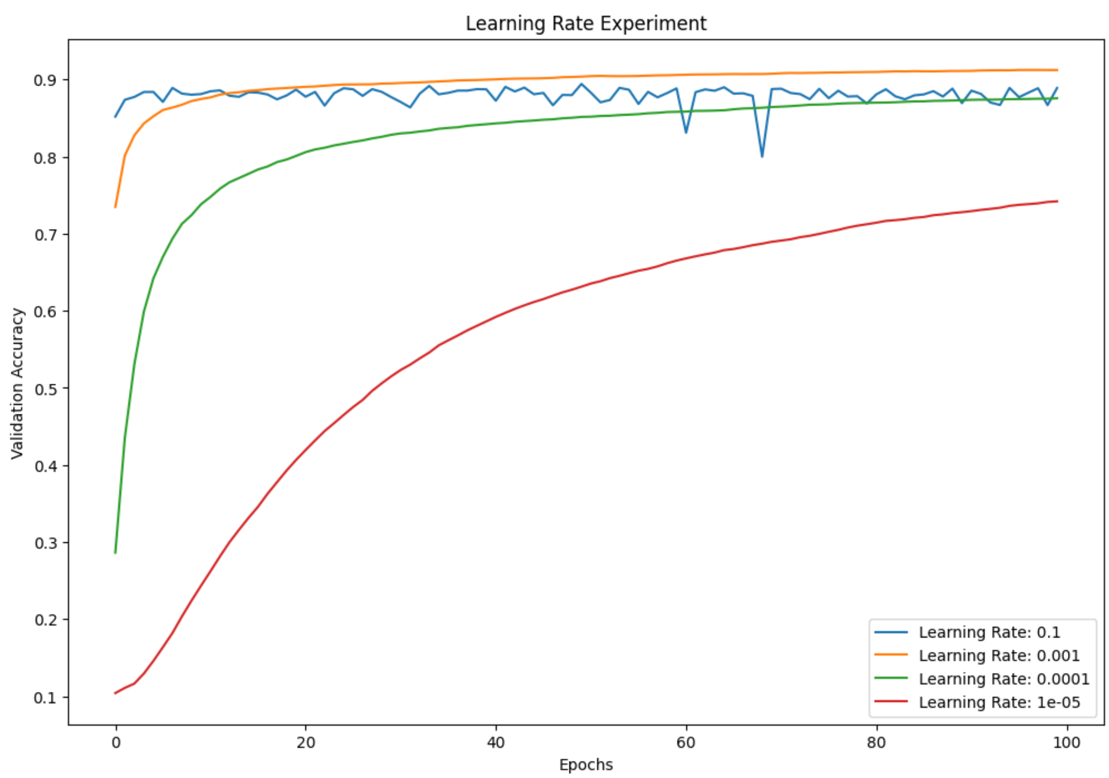


Figure 10 Validation Accuracy with Different Learning Rates

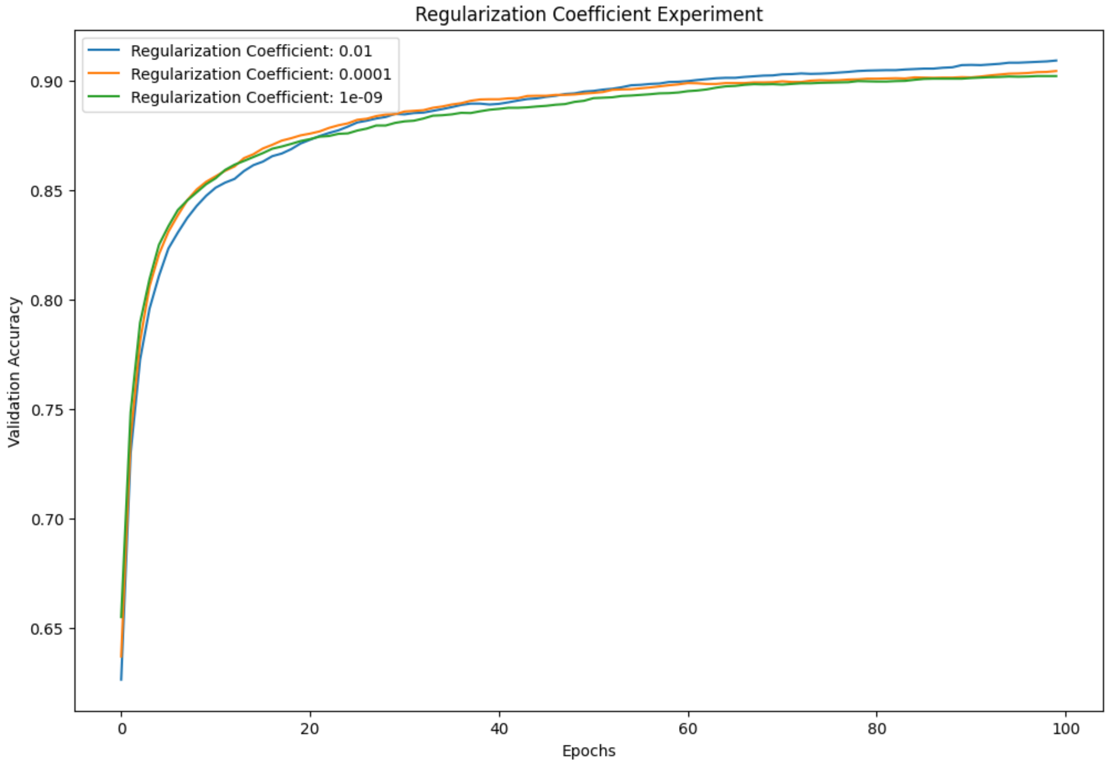


Figure 11 Validation Accuracy with Different Regularization Coefficients

Q2.3

I selected the best values for the hyper parameters as the following:

Batch Size: 1

Weight initialization technique: zero initialization

Learning Rate: 0.001

Regularization coefficient (Lambda): 0.01


Test Accuracy: 0.9227

Figure 12 Test Accuracy with Best Hyper Parameters

İpek Öztaş
22003250
CS464-Sec2

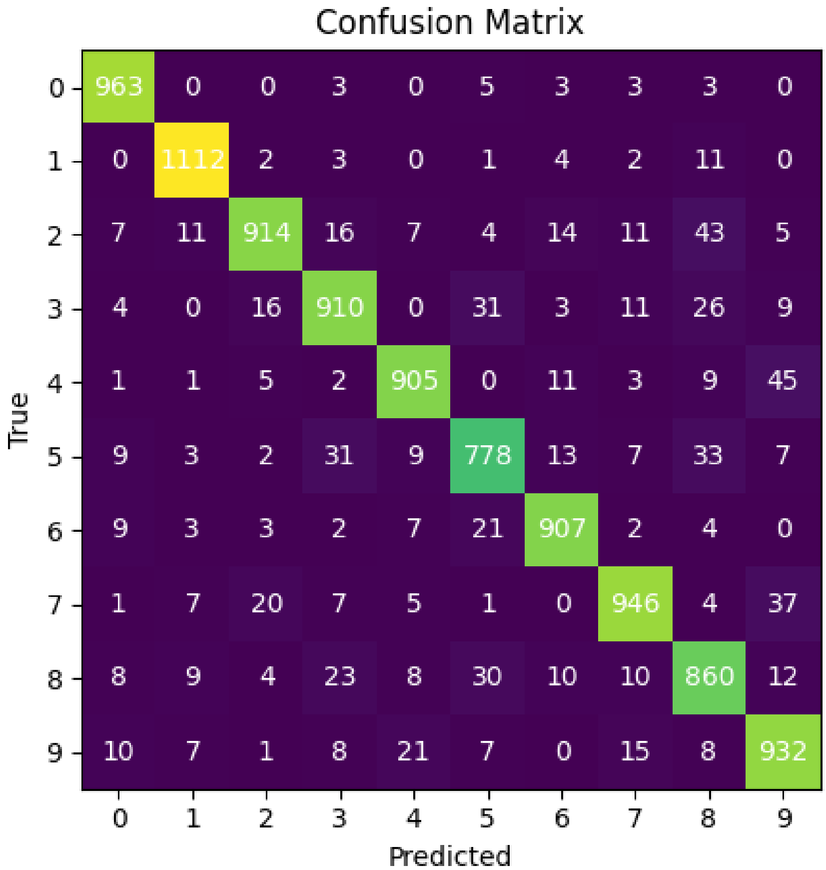
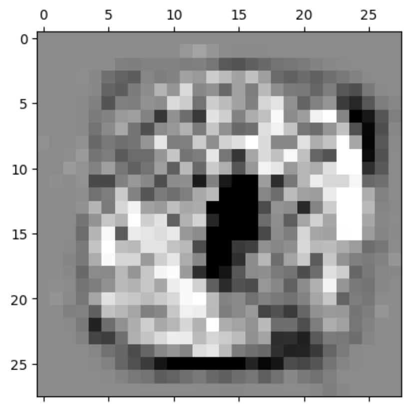
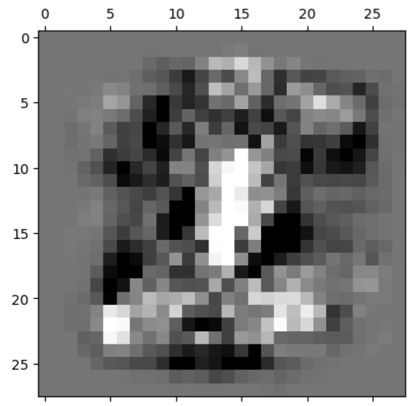


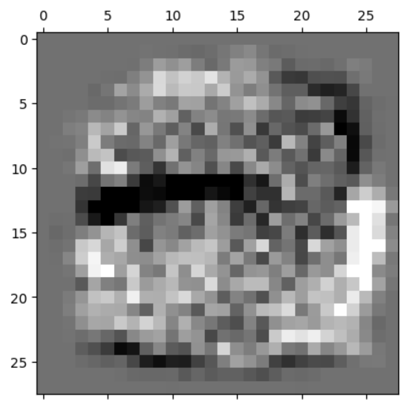
Figure 13 Confusion Matrix of Test Set with Best Hyper Parameters

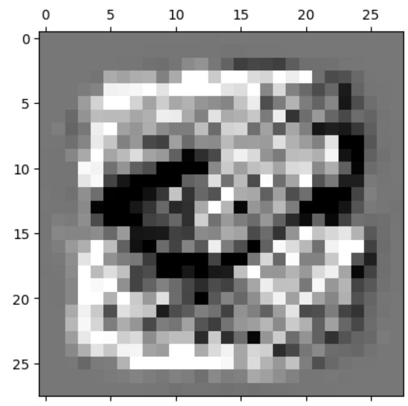
İpek Öztaş
22003250
CS464-Sec2

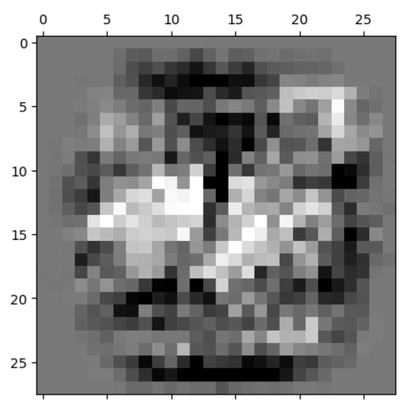
Q2.4

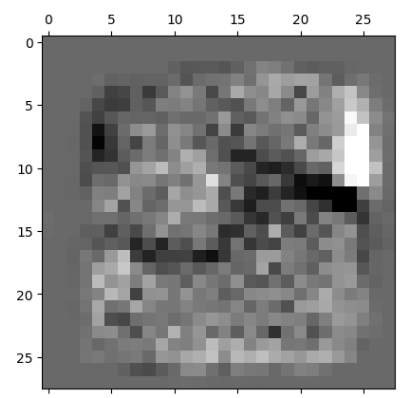












İpek Öztaş
22003250
CS464-Sec2

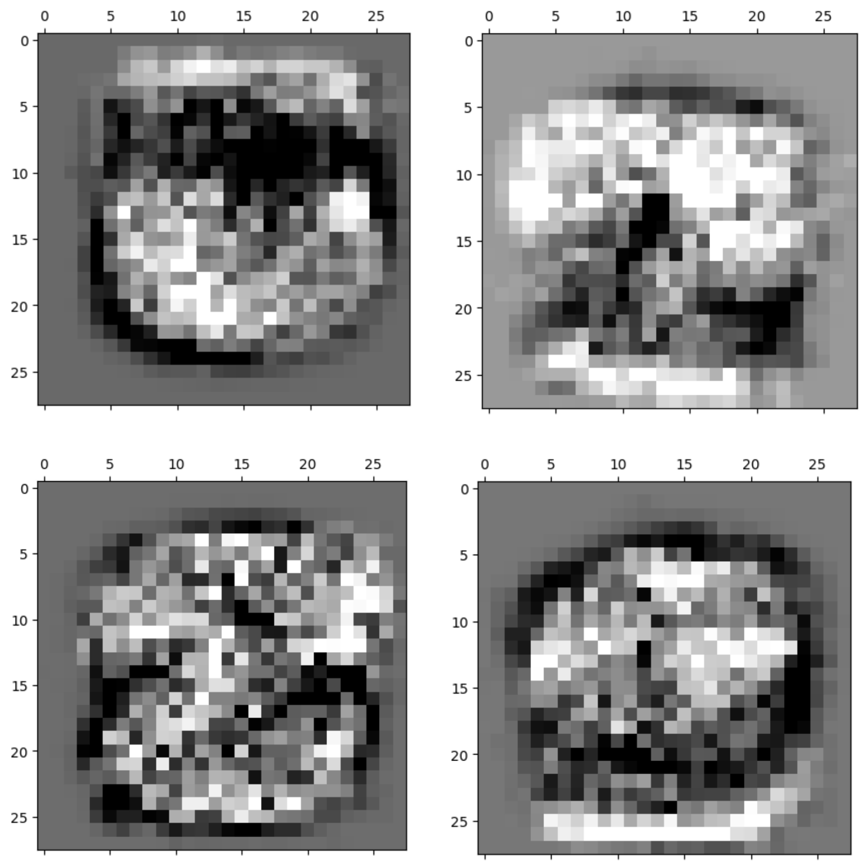


Figure 14 Finalized Weight Vectors

Q2.5

Class 0 Precision: 0.9515810276679841
Class 0 Recall: 0.9826530612244898
Class 0 F1 Score: 0.9668674698795181
Class 0 F2 Score: 0.9762773722627739

Class 1 Precision: 0.9644405897658282
Class 1 Recall: 0.9797356828193833
Class 1 F1 Score: 0.9720279720279721
Class 1 F2 Score: 0.9766379764623222

İpek Öztaş
22003250
CS464-Sec2
Class 2 Precision: 0.9451913133402275
Class 2 Recall: 0.8856589147286822
Class 2 F1 Score: 0.9144572286143071
Class 2 F2 Score: 0.8969578017664377

Class 3 Precision: 0.9054726368159204
Class 3 Recall: 0.900990099009901
Class 3 F1 Score: 0.903225806451613
Class 3 F2 Score: 0.9018830525272548

Class 4 Precision: 0.9407484407484408
Class 4 Recall: 0.9215885947046843
Class 4 F1 Score: 0.9310699588477367
Class 4 F2 Score: 0.9253578732106339

Class 5 Precision: 0.8861047835990888
Class 5 Recall: 0.8721973094170403
Class 5 F1 Score: 0.8790960451977401
Class 5 F2 Score: 0.8749437696806116

Class 6 Precision: 0.9398963730569948
Class 6 Recall: 0.9467640918580376
Class 6 F1 Score: 0.9433177327093084
Class 6 F2 Score: 0.9453825307483845

Class 7 Precision: 0.9366336633663367
Class 7 Recall: 0.9202334630350194
Class 7 F1 Score: 0.9283611383709519
Class 7 F2 Score: 0.9234673955486138

Class 8 Precision: 0.8591408591408591
Class 8 Recall: 0.8829568788501027
Class 8 F1 Score: 0.8708860759493671
Class 8 F2 Score: 0.8780886256891974

Class 9 Precision: 0.8901623686723973
Class 9 Recall: 0.9236868186323092
Class 9 F1 Score: 0.9066147859922179
Class 9 F2 Score: 0.9167814282903797

İpek Öztaş
22003250
CS464-Sec2
Analysis of the Confusion Matrix and Weight Images:

The confusion matrix, as shown in Fig. 11, provides insights into the model's performance across different digit classes. Notably, digits 5 and 8 exhibit a lower count of correctly classified instances, contributing to an overall decrease in accuracy. This aligns with observations from the weight images obtained in Question 2.4, where the patterns for these digits appear less distinct compared to others.

Precision, measuring positive prediction accuracy, reveals that digits 1 and 6 achieve high precision values, indicating fewer false positives which is consistent with their well-defined weight images. The corresponding weight images for these classes may capture well-defined features, contributing to their accurate positive predictions.

Recall, assessing the model's ability to capture all positive instances, shows high values for digits 0 and 1. These digits exhibit effective positive instance capture, reflecting the clarity of distinguishing features in their respective weight images.

Digits 5 and 8, with lower F1 and F2 scores, align with the blurry and less distinct patterns observed in their weight vectors.

According to precision, recall, F1 and F2 scores of digits 0 and 1 the model performs well predicting these two classes. The weight vectors also supports the high scores. Although digit 2 has a similar high precision, the lower recall indicates that the model sometimes fails to label. In Fig. 11, it is often confused with digit 8 due to similar curves.

In summary, the correspondence between the confusion matrix metrics and the characteristics of weight images indicates a correlation between the model's performance and its ability to capture distinctive features for each digit. The model excels in recognizing digits with clear patterns, while challenges arise for digits with less distinct features, evident in both the confusion matrix and weight images.