CS 464 Homework 2 - Görkem Kadir Solun – 22003214

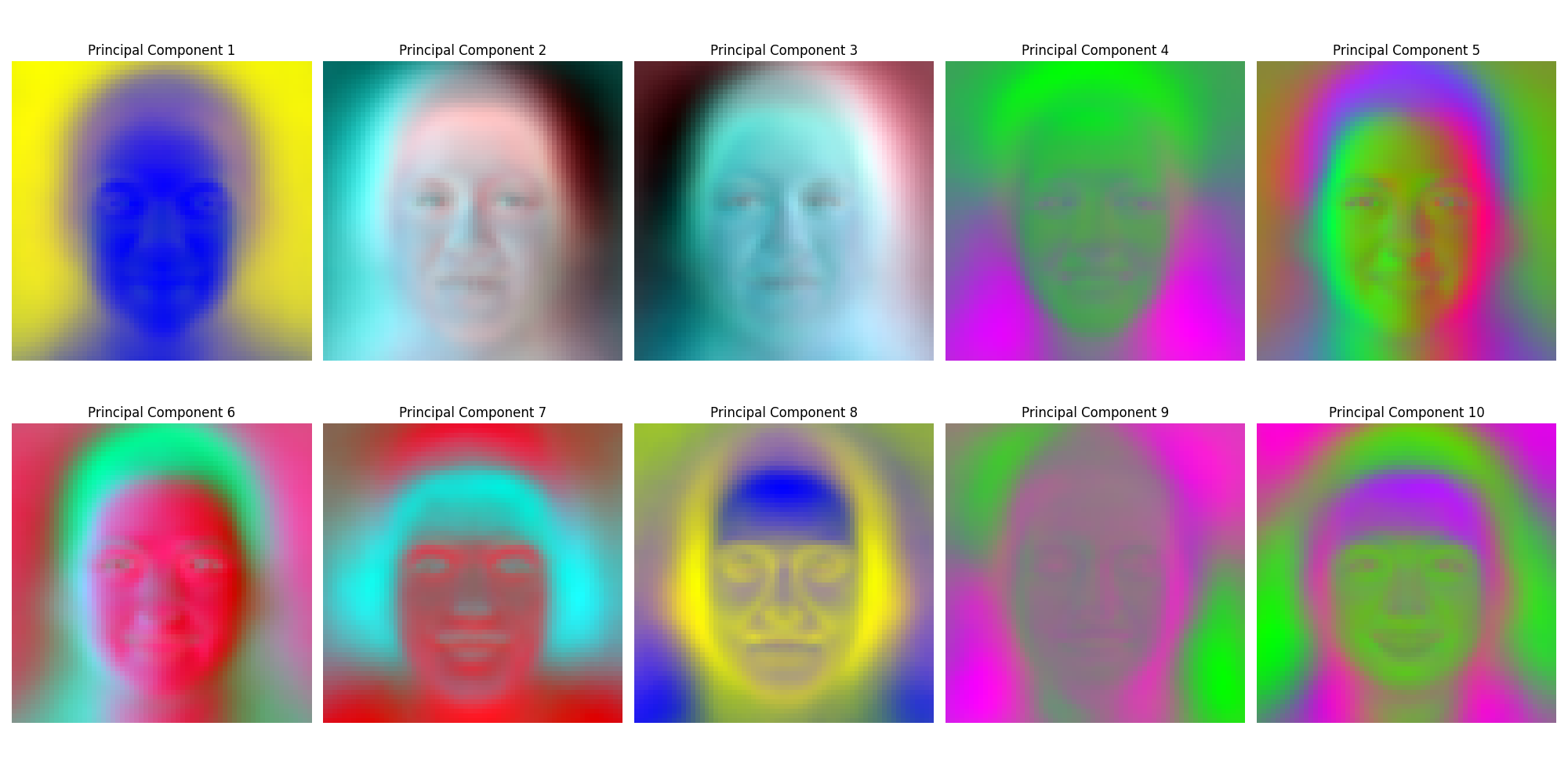
# PCA Question 1.1

A screenshot of a computer program

Description automatically generated

The blue channel achieves a cumulative PVE of 71.9537% with the first 10 components, surpassing the 70% threshold. Consequently, only 9 components are required to meet the requirement, making it the most efficient channel due to its higher variance. Similarly, the green channel explains 70.3356% of the variance with 10 components, just exceeding the threshold, indicating that 10 components are sufficient for this channel. In contrast, the red channel achieves a cumulative PVE of 69.0041% with 10 components, falling short of the 70% target. Analysis shows that 11 components are necessary for the red channel to exceed the threshold, making it the least efficient. Thus, the red channel's requirement of 11 components establishes the minimum number needed to ensure that at least 70% variance is explained across all channels.

# PCA Question 1.2



Every principal component (PC) captures a maximum variance direction, emphasizing the dataset's critical features. The predominant color patterns and most notable global variations are represented by PC 1, which mainly shows subtle color intensity changes, such as gradients from yellow to blue. We can conclude that PC 8 is the opposite of PC 1. Sharper contrasts and complex spatial patterns, like variations in particular regions like facial characteristics, are highlighted in PCs 2 and 3. PCs 4, 5, 9, and 10 emphasize intricate color transitions with less emphasis on general structure, concentrating on more delicate and localized patterns, such as nose or hair texture. In addition to representing hair and the bottom portion of the face, PCs 6 and 7 feature intriguing colors. By preventing rendering distortions from raw values exceeding acceptable ranges, min-max scaling normalizes eigenvector values to a range between 0 and 1, guaranteeing proper RGB image display. This standardization improves the clarity and comparability of color patterns among PCs. The normalized red, green, and blue components are arranged to create meaningful RGB graphics that show how variances in the dataset are dispersed across the spectrum and visualize relationships among color channels. Red-green and blue-yellow contrasts provide essential information on the color dynamics causing the variability in the dataset.

# PCA Question 1.3

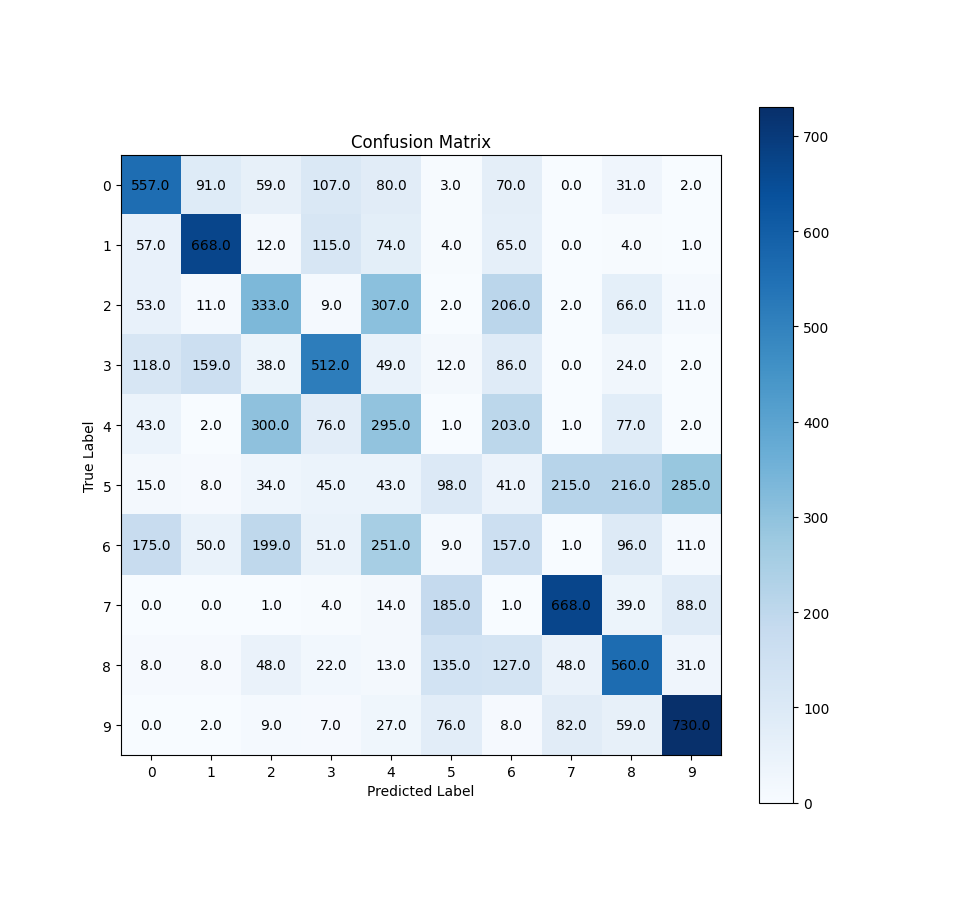


A collage of a person's face

Description automatically generated

Reconstruction using PCA involves projecting an image onto its principal components (eigenvectors) and reprojecting it back into the original space using the first k eigenvectors. These eigenvectors correspond to the directions of maximum variance in the dataset, with earlier components carrying more significant information as they are associated with higher eigenvalues. The choice of k, the number of principal components, controls the level of detail in the reconstructed image. A smaller k captures broader patterns, while a larger k adds finer information, achieving a complete reconstruction when k equals the total number of original components. This process demonstrates how PCA effectively reduces dimensionality while retaining critical information.

# Logistic Regression Question 2.1



A screen shot of a computer

Description automatically generated

The logistic regression model achieved a test accuracy of 46.42% in a 10-class classification problem. This indicates moderate overall performance, but the model faces consistent difficulties making accurate predictions across all classes. Despite this, certain classes have some notable differences in prediction accuracy.

The confusion matrix reveals that specific classes, such as classes 1 and 7, are more accurately predicted. This is evidenced by their diagonal solid values in the matrix. Conversely, other courses, like class 2 and class 6, need to be more accurate. These errors suggest that the model needs help differentiating courses with similar features.

The model uses default configuration parameters, including a learning rate 0.0005, a regularization strength 0.0001, and a batch size 200. During training, the model demonstrates notable improvement in the validation dataset with each epoch, reaching convergence. However, the low regularization strength may not be sufficient to prevent overfitting, potentially hindering the model’s ability to generalize, particularly when handling complex datasets.

The model employs Gaussian normal initialization for its weights, a standard approach. However, experiments described in question 2.2 suggest that alternative weight initialization methods could yield better optimization outcomes. This indicates potential room for improvement in model initialization strategies to enhance overall performance.

# Logistic Regression Question 2.2

A graph of a number of colored lines

Description automatically generated with medium confidence

A graph of different colored lines

Description automatically generated

A screen shot of a computer

Description automatically generated

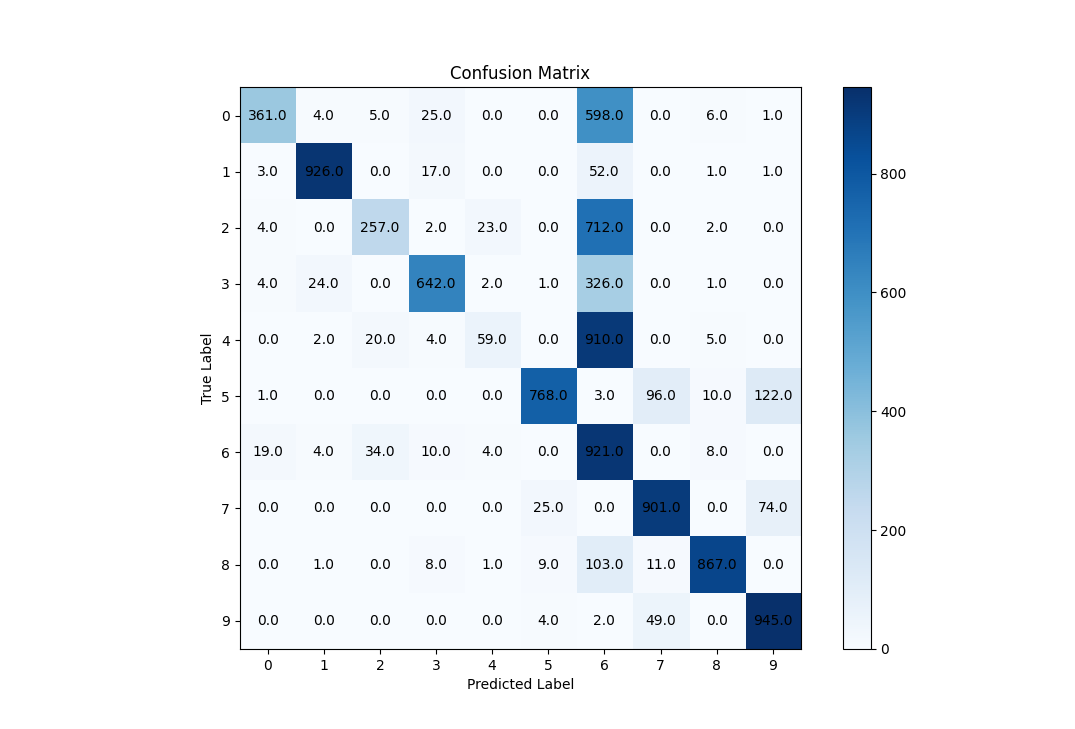
The first is the impact of batch size variation on model accuracy over 100 epochs. Higher accuracy (about 80%) and a faster convergence rate are achieved with smaller batch sizes, mainly when the batch size is 1. Smaller batches allow for more frequent weight updates, which improves the model's ability to learn intricate patterns given the default hyperparameters. It's crucial to remember that the model performed best overall when additional hyperparameters were changed, and batch sizes ranged from 16 to 32. However, a batch size 64 stabilizes at about 60% and exhibits reasonable accuracy and convergence. Batch sizes of 3000 or more are huge, which significantly slows down the learning process and produces accuracy that is only 20%. This implies that the model's capacity to capture a variety of properties in the dataset may need to be improved by large batch sizes, which would lower its capacity for generalization.

The second point examines how various weight initialization techniques affect the model's accuracy over 100 epochs. Zeros, uniform distribution, and Gaussian distribution are the names of the initialization techniques. The findings demonstrate that within the first 25 epochs, the uniform initialization performs better than the others in terms of accuracy. Although it converges more slowly, the Gaussian initialization does so gradually. Remarkably, after 100 epochs, the zero initialization produces the highest ultimate accuracy, close to 0.8. This implies that this method stabilizes best with zero initialization. These results emphasize the importance of using the proper initialization technique to balance convergence speed and ultimate performance.

The third point shows the impact of various learning rates on the accuracy of the model over 100 epochs: 0.01, 1e-03, 1e-04, and 1e-05. The findings show that a higher learning rate of 0.01 produces the highest final accuracy of roughly 0.8 and accelerates convergence. A learning rate of 0.001 results in poorer ultimate accuracy and slower convergence. The lowest learning rate (1e-05) performs the worst, and smaller learning rates (1e-04 and 1e-05) lead to much slower learning and poorer end performance. Accordingly, more excellent learning rates must be carefully chosen to prevent problems like overshooting minima or creating instability in the training process, even though they might speed up learning and increase ultimate accuracy.

The fourth point outlines how different regularization coefficients influence accuracy over 100 epochs. Interestingly, the differences between the tested regularization values—0.01, 1e-04, and 1e-5—are negligible, with all configurations leading to similar accuracy levels of around 50%. This demonstrates that the model's performance in this scenario is relatively insensitive to the regularization parameter within this range. Despite the minimal impact observed, regularization still plays a crucial role in preventing overfitting by penalizing large weights, which helps the model generalize better to unseen data.

# Logistic Regression Question 2.3



# Logistic Regression Question 2.4

A close-up of a body scan

Description automatically generated

A graph of a weight scale

Description automatically generated with medium confidenceA close-up of a weight

Description automatically generatedA close-up of a weight chart

Description automatically generatedA close-up of a weight chart

Description automatically generatedA close-up of a weight chart

Description automatically generatedA close-up of a weight chart

Description automatically generatedA graph with a number of objects

Description automatically generated with medium confidenceA close-up of a weight chart

Description automatically generatedA graph of weight for class 9

Description automatically generated

It is expected that the weight maps may appear hazy. Regularization methods used during training, including L2 regularization, cause this blurriness. By punishing abnormally high weight values, regularization helps avoid overfitting and improves the model's ability to generalize to new data. Brighter areas on the weight maps show higher positive weights. The classification of that class is greatly aided by these regions, which correlate more with input features. Darker areas, on the other hand, indicate negative weights, emphasizing characteristics that actively lower the input's chance of falling into that class. Grey areas indicate little impact on the class's classification. These regions stand for neutrality when the associated pixels don't strongly support or disagree with the classification choice. Each weight map visualization highlights the traits necessary for differentiating one class from another. These maps help the model make classification judgments by capturing each class's distinctive qualities. The color bars show that various classes have varied weight values. This variety might reflect how different or complex it is to categorize classes. For instance, certain classes might have more distinctive traits than others, while others might share qualities that make them more challenging to differentiate. Every class is associated with a specific category, such as shoes, t-shirts, or pants. The weight maps highlight the significance of regions for categorization and assist in visualizing the model's emphasis on the salient characteristics that set these categories apart.

# Logistic Regression Question 2.5

A screenshot of a computer screen

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 Score | F2 Score |
| 0 | 0.920918 | 0.361 | 0.518678 | 0.410974 |
| 1 | 0.96358 | 0.926 | 0.944416 | 0.93328 |
| 2 | 0.813291 | 0.257 | 0.390578 | 0.297729 |
| 3 | 0.90678 | 0.642 | 0.751756 | 0.681818 |
| 4 | 0.662921 | 0.059 | 0.108356 | 0.072145 |
| 5 | 0.951673 | 0.768 | 0.850028 | 0.798835 |
| 6 | 0.253929 | 0.921 | 0.398098 | 0.603776 |
| 7 | 0.852412 | 0.901 | 0.876033 | 0.890844 |
| 8 | 0.963333 | 0.867 | 0.912632 | 0.884694 |
| 9 | 0.826772 | 0.945 | 0.881941 | 0.918724 |