

Task 3: Interpretation

1. Number of Singular Vectors Retained during PCA

I chose to retain 8 singular vectors during PCA. The rationale behind this choice is based on the scree plot generated from the singular values. In the scree plot, I observed an "elbow" or a point where the explained variance starts to level off. Retaining 8 principal components captures a significant portion of the total variance, ensuring a good balance between information retention and dimensionality reduction. Additionally, considering the cumulative explained variance can also be a guide, and in this case, 10 components accounted for a satisfactory percentage.

2. Comparison with and without PCA

If I were to repeat training using the same neural network architecture without performing PCA, the new fit might show some differences compared to the results obtained with PCA.

- **Training without PCA:** The model would directly use the original feature space, which might include some correlated or less informative features. This could lead to a higher-dimensional input space, potentially causing overfitting and increased computational complexity.

- **Training with PCA:** By reducing the dimensionality through PCA, the model focuses on the most significant features that contribute to the variance in the dataset. This can lead to improved generalization, faster training times, and potentially better performance on the test set. The PCA-transformed features serve as a compressed representation that retains essential information.

In summary, using PCA before training the neural network is likely to result in a more efficient and possibly more interpretable model. It helps mitigate the curse of dimensionality and enhances the model's ability to capture meaningful patterns in the data. The choice of the number of principal components, in this case, 8, strikes a balance between retaining sufficient information and avoiding overfitting.