# STAT 5244 – Unsupervised Learning

**Homework 1**

Name: Chuyang Su UNI: cs4570

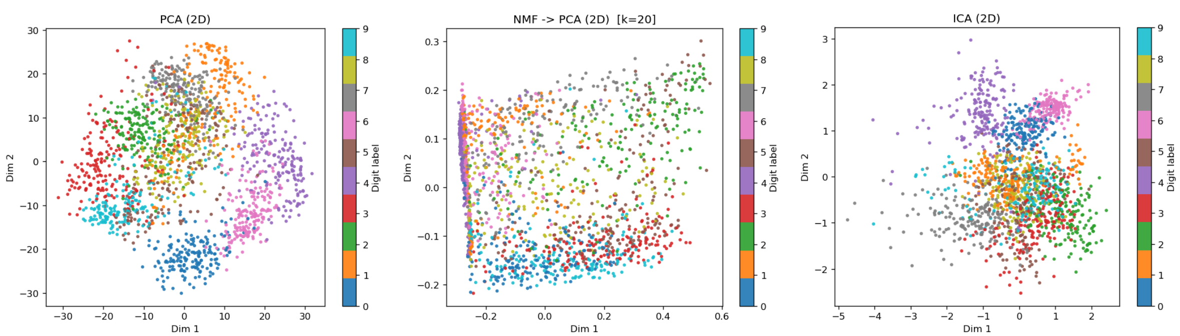
# 1 Dimension Reduction on Digits Data.

## 1.1 Apply linear dimension reduction techniques including PCA, NMF, and ICA.

In this experiment, I applied three linear dimension reduction methodsand compared their performance on the `scikit-learn` *Digits* dataset ($n = 1797$, $p = 64$).

Each method projects the data into a two-dimensional latent space, on which I visualized the results and quantitatively evaluated their ability to separate the ten digit classes.

The results of this experiment are summarized below.




(a) PCA (b) NMF→PCA (c) ICA

Figure 1: 2D embeddings of the digits data using PCA, NMF→PCA, and ICA. Colors denote true digit labels.

| Method | ARI | NMI | Silhouette |
|---|---|---|---|
| PCA | **0.3614** | **0.5190** | 0.3993 |
| NMF → PCA | 0.1588 | 0.2961 | **0.4080** |
| ICA | 0.3175 | 0.4567 | 0.3673 |

Table 1: Quantitative comparison of linear dimension reduction methods. The best scores for each metric are bolded.

**PCA.** For PCA, I retained the first two principal components and projected the samples into the 2D subspace they span, which preserves the primary directions of variance. As for hyperparameters, PCA has very few tunable parameters—the main one being the number of principal components. For ease of visualization, I set the number of components to 2 (PC=2) in this experiment. Figure 1a shows that the data are well dispersed, and digits such as 0, 3, 4, 6, and 9 form clearly separated clusters. Quantitatively, PCA achieved an ARI of 0.3614, an NMI of 0.5190, and a Silhouette score of 0.3993. Except for the Silhouette score, PCA obtained the highest values among the three methods. This indicates that PCA effectively separates the digits and maintains a high level of consistency with the true labels. Although its Silhouette value (approximately 0.4) is not the highest, it still suggests reasonably compact and well-separated clusters. This minor difference can be attributed to slight overlaps between neighboring clusters in the 2D embedding, even though the overall structure aligns well with the ground-truth classes.

In addition to the 2D embedding, I plotted the PCA scree plot and a bar chart of the top ten principal components' explained variance, as shown in Figure 2. Both plots reveal that the first three components contribute substantially more variance than the rest, with the third component explaining slightly less variance than the first two but significantly more than the fourth. This supports the observation that the 2D projection loses some discriminative information, which explains why the best ARI achieved by PCA remains moderate (0.3614).

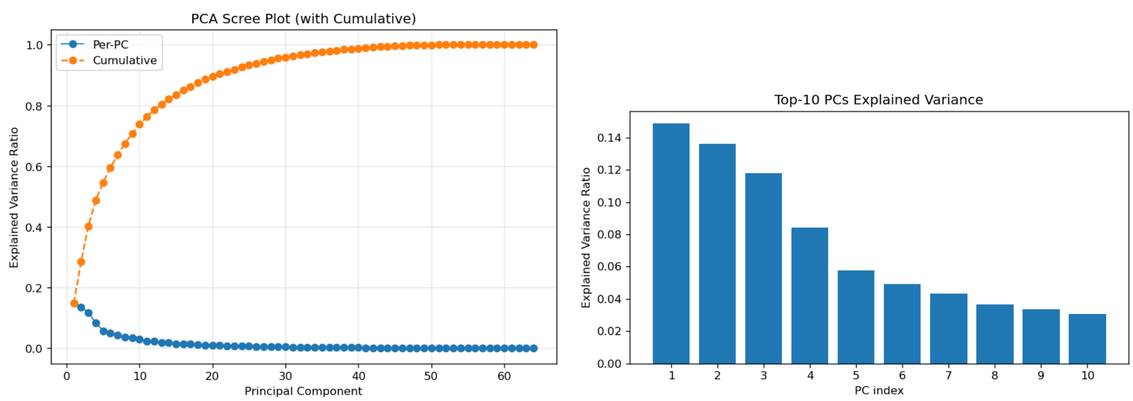



Figure 2: (Left) Scree plot showing the variance explained by each principal component; (Right) bar chart of the top-10 PCs' explained variance ratios.

Furthermore, I visualized the top ten PCA component images (Figure 3), which illustrate the principal modes of variation across digits.

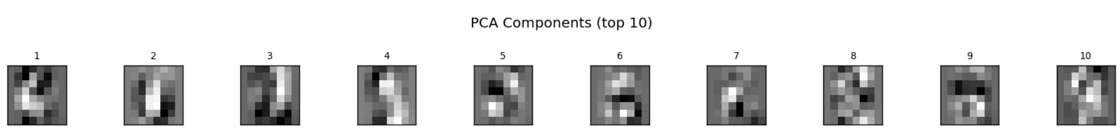



Figure 3: Top ten PCA components visualized as $8 \times 8$ basis images.

**NMF.**

# A Appendix: Code Implementation