AMATH 482/582: HOME WORK 3

XINYUE LYU

Amazing Department, University of Washington, Seattle, WA
xlyu6@uw.edu

ABSTRACT. This assignment explores handwritten digit classification using the MNIST dataset. Principal Component Analysis (PCA) is used for dimensionality reduction, followed by classification using Ridge regression, k-Nearest Neighbors (KNN), and Support Vector Machines (SVM). Cross-validation is used to ensure optimal performance. The results show that KNN outperforms Ridge regression, achieving 97.60% accuracy. The findings highlight the effectiveness of different machine learning models for image classification and demonstrate the importance of PCA in reducing computational complexity while preserving classification accuracy.

1. Introduction and Overview

Handwritten digit classification is a fundamental problem in machine learning and serves as a benchmark for evaluating classification algorithms. The MNIST dataset, introduced by LeCun et al., consists of 28×28 grayscale images of handwritten digits (0-9) and is widely used in machine learning research. Traditional classification approaches struggle with high-dimensional image data, making dimensionality reduction techniques like Principal Component Analysis (PCA) crucial. This study applies PCA to reduce dimensionality and evaluates Ridge regression, k-Nearest Neighbors (KNN), and Support Vector Machines (SVM) for classification for specific selecting digits. The goal is to determine which algorithm provides the highest accuracy while balancing computational efficiency.

2. Theoretical Background

Singular value decomposition(SVD) is a fundamental technique in linear algebra and data analysis, widely used for dimensionality reduction and data compression. Given a matrix A of shape $m \times n$, SVD decomposes it into three matrices:

$$A = U\Sigma V^T$$

where:

- U $(m \times m)$ contains the left singular vectors (eigenvectors of AA^T), a rotation,
- Σ $(m \times n)$ is a diagonal matrix of singular values, a stretching, representing the importance of each mode and order the value from high to low,
- V^T $(n \times n)$ contains the right singular vectors (eigenvectors of A^TA), a rotation.

PCA finds a set of orthogonal basis vectors, called principal components, which capture the directions of maximum variance in the data. Given a dataset represented as a matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ (where rows represent observations and columns represent features), PCA is performed by computing the covariance matrix:

$$\mathbf{C} = \frac{1}{n}\mathbf{X}^T\mathbf{X}.$$

Date: March 3, 2025.

Generally, people assume that data measurements with a high variance capture more important features. Calculating PAC by SVD transforms orthogonal data points into ellipse.

Each principal component (PC) in PCA is associated with an "explained variance," which measures how much of the total variance in the data is accounted for by that component.Cumulative Energy is the running total of the explained variances of the first few PCs. In other words, cumulative energy gives a measure of how much of the original data's variability is captured by the principal components. PCA allows people to reduce the number of features in a dataset without losing significant information, making the dataset smaller and more manageable.

In this assignment, I project the original data of images on to a new coordinate system defined by the principal components. This projection is a a linear transformation of the original data.

Classification is a supervised learning task where the goal is to assign a label(0 to 9 in this assignment) to a given input(images) based on its features.

The Ridge classifier uses linear regression with L2 regularization to prevent overfitting by penalizing large coefficients. K-Nearest Neighbors (KNN) classifies data based on the majority label of the K closest neighbors, making it flexible but computationally expensive. Support Vector Machines (SVM) find the hyperplane that maximizes the margin between classes, using kernel functions for non-linear boundaries. Each method differs in assumptions, computational cost, and suitability for different types of datasets. SVM is a linear classifier that finds a hyperplane that maximizes the margin between classes. It can handle non-linear decision boundaries using kernel functions, mapping the data into higher-dimensional spaces.

3. Algorithm Implementation and Development

For this project, Principal Component Analysis from `sklearn.decomposition` was applied to reduce the dimensionality of the MNIST dataset. Ridge regression from `sklearn.linear_model`, K-Nearest Neighbors from `sklearn.neighbors`, and Support Vector Machines from `sklearn.svm` were used for classification. Additionally, `matplotlib` and `numpy` were used for visualization and numerical computations. All implementations were conducted using Python and Scikit-learn, a widely used library for machine learning applications.

4. Computational Results

The dataset contains 60,000 images with 28×28 features for each. I performed PCA to reducing the features of each image sample. The Figure 1 shows the first 16 PC modes. To have 85% energy or variance, I need to use 59 PC modes. The Figure 2 shows a comparison between the original and the reconstructed images after applying PCA with a reduced number of components. We can clearly recognize each digit from the reconstructed images under the original images, so using truncated modes is reasonable.

digits	cross-validation	test accuracy
(1, 8)	0.9643 ± 0.0027	0.9801
(3, 8)	0.9588 ± 0.0061	0.9642
(2, 7)	0.9797 ± 0.0014	0.9748

Table 1. This table presents the cross-validation and test set accuracies calculating by Ridge classifier for three different digit pairs: (1, 8), (3, 8), and (2, 7).

I applied Ridge Classifier for binary classification of specific digit pairs. For each digit pair (1, 8), (3, 8), and (2, 7), I used cross-validation to evaluate the classifier's performance on the training set, and trained the Ridge Classifier on the full training set and evaluated it on the test set, printing

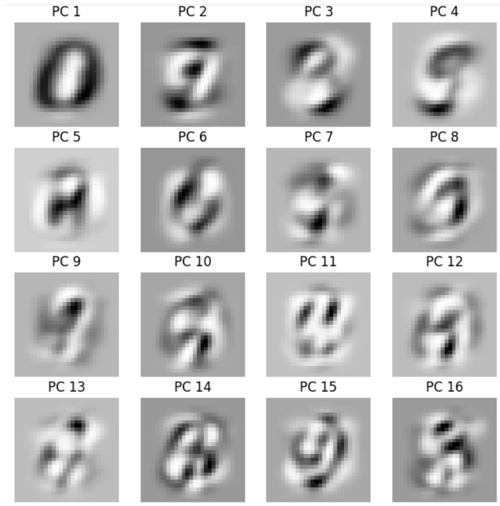

FIGURE 1. The figure visualizes the first 16 principal components (PCs) learned by the PCA algorithm. Each subplot shows a 28x28 image, which represents one principal component.

FIGURE 2. The figure shows a comparison between the original and the reconstructed images after applying PCA with a reduced number of components with the 85% of the variance.

Classifier	**Accuracy**
Multi-class Ridge	0.8561
Multi-class KNN	0.9760
Multi-class SVM	0.9858

TABLE 2. Multi-class classification accuracy for different classifiers

the cross-validation accuracy and test set accuracy for each pair. The Table 1 accuracy for Ridge classifier. It performed very well with test accuracy higher than 95%.

To compare how different classifiers behave, we can infer from the Table 2 that the accuracy of SVM > KNN > Ridge.

5. Summary and Conclusions

Overall, the SVM and KNN classifiers show better generalization and accuracy on this problem, while Ridge may need more tuning or a different feature set to improve performance.

Acknowledgements

I include the explanation of PCA from the last HW.

The author is thankful to Dr. Frank for instructing the PCA, three classifier concepts, and coding with the textbook[1].

References

[1] J. N. Kutz. *Data-driven modeling & scientific computation: methods for complex systems & big data*. OUP Oxford, 2013.