



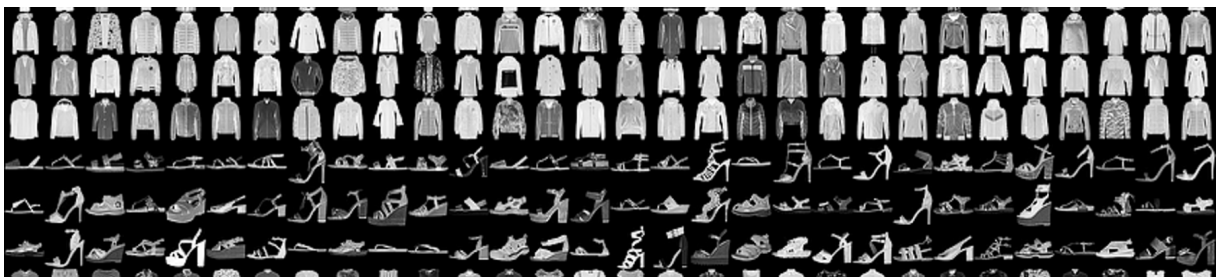
PCA, SVM, Bayes, KNN, and MLP

DeadLine: 30 Ordibehesht 1402

Assignment 2

1 Classification on Fashion-MNIST Images Data Set

The Fashion-MNIST dataset is a collection of images consisting of 70,000 grayscale images of fashion products, each with a size of 28x28 pixels. The dataset is split into a training set of 60,000 images and a test set of 10,000 images. The images are classified into 10 different categories, including T-shirts/tops, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots.



2 Instructions

1. Load the data and separate the labels from the training features. y corresponds to the first column, and X corresponds to the data in the other columns. Name the training data X_{train} and y_{train} , and the test data X_{test} and y_{test} .
2. Perform pre-processing on the data if necessary.
3. Perform Principal Component Analysis (PCA) on the training data using your own implementation. You may use built-in functions to compute the eigenvectors and eigenvalues.
4. Plot the data points in the newly transformed n -dimensional space.
5. Explain your results in your report file. What do the new principal components represent? Which features of the original dataset contribute most to each principal component?
6. Compare your results with those obtained by performing PCA using a built-in PCA function in your report file.
7. Train four models on the selected features: support vector machines (SVM), Bayes, k-nearest neighbors (KNN), and multilayer perceptron (MLP).
8. Use 5-fold cross validation on the training data to find the best hyperparameters for each model. Some of the hyperparameters to tune are:
 - SVM: The kernel type (linear, polynomial, or radial basis function).
 - Bayes: No hyperparameter.
 - KNN: Number of neighbors.
 - MLP: Number of hidden layers, number of neurons in each hidden layer, Learning rate, Activation function for each layer, Batch size, Number of epochs (i.e., training iterations)

9. In your report file, explain the differences between the different activation functions used in the MLP model.
10. Perform classification using all four models with the best parameters obtained using 5-fold cross-validation. Mention the best parameters for these models in your report.
11. Provide these 4 evaluation metrics: accuracy, precision, recall, and f1 score. Along with a confusion matrix for both training and test data.
12. Analyze all the obtained confusion matrices in your report file.
13. How does the performance of the MLP model compare to the other models used in the assignment?

3 Additional Guidance

- Make sure your code is in .ipynb format.
- Along with your code, please include a report file that thoroughly analyzes your results.
- Use appropriate visualizations and statistics to support your analysis and conclusions.