[CSC 5825 Fall 2024]

Due. Nov 8, 2024 Homework 3

Full credit: 100 points

October 13, 2024

Question. (100 points) Programming question: Unsupervised Dimension Reduction

In this question, you are asked to run Singular Value Decomposition (SVD) on Fashion-MNIST data set, interpret the output and train generative classifiers for multinomial classification of 10 classes. For the Fashion-MNIST data set, you can find more details in the original GitHub website or Kaggle website. In this assignment, you are allowed to use a library implementation of SVD. For Python users, we recommend scikit-learn's implementation TruncatedSVD.

Tasks:

- Load the training and test data sets from fashion-mnist_train.csv and fashion-mnist_test.csv. Each row uses a vector of dimension 784 with values between 0 (black) and 255 (white) on the gray color scale. (10 points)
- Use SVD function to reduce the number of dimensions of the training data set so that it explains just above 90% of the total variance. Remember to scale the data before performing SVD. Report how many components you select and their variance ratios. (30 points)
- Train generative classifiers (Naive Bayes and KNN) and discriminative classifier (multinomial logistic regression) on both the training data set after SVD and the original data set (without dimension reduction). Fine-tune the hyper-parameters, e.g. learning rate in MLR and k value in KNN, to achieve best performance on a validation set split from the training set. Write a brief description to compare the performances of these classifiers in terms of accuracy, precision, recall and F_1 score on the test set. (60 points)

Guidelines:

In this homework, you are allowed to use scikit-learn's implementations Multinomial Logistic Regression, Naive Bayes, and KNN, and you are allowed to use libraries like `NumPy` for general operations such as matrix manipulation or array handling. However, **you are not allowed to use scikit-learn or any other external libraries to directly implement functions or algorithms for tasks such as data preprocessing, dataset splitting, evaluation metrics, or any other part of the homework not explicitly mentioned above**. All such tasks must be implemented manually.

Submission Instructions

Homework must be submitted electronically through Canvas website on/before the due date/time. Homework must be typed with LaTeX or Word. The code can be submitted as .py file or .ipynb file.