

# [CSC 5825 Fall 2022]

## Homework 3

Due 11:59 pm, November 4, 2022

Full credit: 100 points

**Question.** (100 points) Programming assignment: Dimension Reduction

In this assignment, you are required to run Singular Value Decomposition (SVD) on MNIST data set, interpret the output and train classifiers for multinomial classification of 10 classes. For the MNIST data set, you can find more details in the original data set 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 `mnist_train.csv` and `mnist_test.csv`. Each row uses a vector of dimension 784 with values between 0 (black) and 255 (white) on the gray intensity scale. (10 points)
- Use SVD 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. Generate a plot with x-axis representing number of principle components and y-axis representing the proportion of the variance explained (Refer the Slide #16 of Lecture 6 for an example). 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 dimension reduced data set using 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 on the test set. (60 points)

### Guidelines:

In this assignment, you are allowed to use scikit-learn's implementations Multinomial Logistic Regression, Naive Bayes, and KNN directly.

**Submission Instructions**

Homework must be submitted electronically through Canvas website on/before the due date/time. Homework must be typed with LaTeX, or a Python IDE, or Word. The code can be submitted as .py file or .ipynb file. Late homeworks will not be accepted unless with legitimate excuses with documents.