

CS 7267- HW 1

In HW1, we will implement KNN using Python. You are given two data sets: [MNIST_training.csv](#) and [MNIST_test.csv](#), where “MNIST_training.csv” contains a training data that you will find the K-nearest neighbors and “MNIST_test.csv” consists of a test data that you need to predict labels. The training data contains 10 classes (i.e., 0, 1, 2, ..., 9), each of which has 90 samples, while there are 10 samples on each class in the test data set.

You can find the description of the MNIST data at <https://www.kaggle.com/c/digit-recognizer/data>, but use the given simplified data sets.

For this homework assignment, please follow the procedure:

1. For each data in “MNIST_test.csv”, compute distances with the training data.
2. Find the K-nearest neighbors, and decide the majority class of them.
3. Compare the prediction with the ground truth
 - a. Correctly classified if the predicted label and ground truth is identical.
 - b. Incorrectly classified if the predicted label and ground truth is NOT identical.
4. Repeat Step 1-4 for all data in the test data
5. Then, you can count how many test data are correctly classified and incorrectly classified.
6. Show the accuracy of your KNN. Compute accuracy by:

$$accuracy = \frac{\text{\# of your predictions correctly classified}}{\text{\# of total test data}}$$

You CANNOT use any libraries or built-in functions of KNN. You have to implement it.

You need to think of what is the optimal “K” in KNN. Describe how you decided the optimal K in the assignment.

You have to submit the followings to D2L:

1. MS word file
 - Describe what you have done for the homework assignment.
 - Accuracy by KNN
2. Python source code file(s)
 - Must be well organized (comments, indentation, ...)
 - You need to upload the “original python file (*.py)” and also its “PDF” version.
 - o For the PDF file, you can just convert the source file to PDF. One way is to print the source file and save to “PDF”.

You have to submit the files SEPERATELY. DO NOT compress into a ZIP file.

Deadline:

The deadline for HW1 is **11:59pm Wednesday, February 7, 2018**. Late assignments will be accepted up to 24 hours after the due date for 50% credit. Assignments submitted more than 24 hours late will not be accepted for credit.