SWE 582: Sp. Tp. Machine Learning for Data Analytic
Homework #2
Due May 26, 2024 by 23:59

In this assignment, you will implement the following models. You can use libraries such as numpy, scipy, matplotlib in your experiments. Please submit a PDF report containing the link to our code, your answers, and references. Please cite all the resources used in the assignment. If you ever use an AI tool such as ChatGPT, please acknowledge.

Support Vector Machines (50 pts)

You will implement the SVM classifier for the MNIST [1] dataset in this task. MNIST has 50,000 training and 10,000 test images of 10 classes. Please consider the digits 2, 3, 8, and 9 in this section. Thus, the total number of samples will be 20,000 (5000 for each class) in the training set and 4,000 in the test set. Please flatten the gray-scale images and feed these vectors directly to your soft-margin SVM model.

1. Please train a 4-class SVM using the scikit-learn's soft margin primal SVM function with linear kernel. Please tune the hyperparameters and report your training and test accuracy.

2. Please train a 4-class SVM using the scikit-learn's soft margin dual SVM function with a non-linear kernel. You may choose any kernel you like. Please tune the hyperparameters and report your training and test accuracy. Compare the test accuracies with part 1.

3. Please find the support vectors using the dual SVM model you trained in part 2 and inspect the images. Please discuss whether there is any visual difference between the support vectors and other images in the same class.

K-Means Clustering (50 pts)

Please download kmeans_data.zip on Moodle. In this problem, groundtruth class assignments are given in labels.npy file. Please do the following.

1. Plot the data using a scatter plot. Assign different colors to different labels.

2. Implement k-means algorithm from scratch and run it. You may use the Euclidean distance as the distance metric.

3. Plot the final clustering assignments as a scatter plot.

---

[1] https://en.wikipedia.org/wiki/MNIST_database