Project Proposal of ECE 532

Xin Wang

October 2020

1 The data set

The data set that will be used is MINIST database of handwritten digits, the link is here.

This data set contains a training set of 60,000 examples, and a test set of 10,000 examples.

The image has 28×28 pixels and is gray-scaled. The pixel value is from 0 (white) to 255 (black). The digits have been size-normalized and the center of the mass of the pixels has been centered in the image. The sets of writers of the training set and test set have been ensured to be disjoint.

The classification problem is trying to recognize the handwritten digits and classify the image to digits.

2 Algorithms

Three algorithms will be implemented on this data set:

- ridge regression
- LASSO
- neural network

The training set of the data set will be used to train the model. The parameters of the model will be determined using cross validation on the training set. Then the model will be applied to the test set to get the error rate as the metric of accuracy.

2.1 Evaluation

The models will be evaluated based on the following metrics:

- error rate
- training time of the model
- how much human effort is needed
- computation complexity of classification after training

3 Github page

my Github page is here.

4 Timeline

The tentative timeline is as follows:

- Nov. 1st: Finish ridge regression
- Nov. 17th: Finish LASSO
- Dec. 5th: Finish neural network
- Dec. 6th-Dec. 11th: Write the final report