

CS4092 Machine Learning Laboratory

Assignment II

Date of posting : 10 March 2022

MNIST database of handwritten digits

Description of the dataset

The MNIST database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. The database is widely used for training and testing in the field of machine learning.

It was created by "re-mixing" the samples from NIST's original datasets. The creators felt that since NIST's training dataset was taken from American Census Bureau employees, while the testing dataset was taken from American high school students, it was not well-suited for machine learning experiments. Furthermore, the black and white images from NIST were normalized to fit into a 28x28 pixel bounding box and anti-aliased, which introduced grayscale levels.

Dataset download from the URL : <http://yann.lecun.com/exdb/mnist/>

Tasks :

- 1. Build and SVM to classify the MNIST dataset into classes 0 to 9.***
- 2. You are required to implement various kernel methods.***
- 3. Analyze the error rate for various kernels.***
- 4. Note down your observations.***

[Note: You can utilise on the Python implementation of SVM model].