Image Classification Project

Introduction : The project was associated with ‘Artificial Intelligence Course’ offered by MyCaptain. It contains machine learning algorithm written for classification of the images of handwritten digits. It uses Random Forest Classifier and classifies the images with great accuracy.

In a supervised learning algorithm, a set of input data and its response data (output) is fed to the machine. The machine is thus trained to generate reasonable predictions for the response to new(future) data. Therefore, the image classification is a supervised learning problem.

Random Forests : Random forests or random decision forests are used for both classification and regression. In a random forest algorithm, a multitude of decision trees is constructed at training time. The output is the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual trees

Dataset: The dataset used here is obtained from the Kaggle datasets. It contains images of handwritten digits in the csv file. The data is already processed, doesn’t have any missing values, so that saves lot of time.

Source : https://www.kaggle.com/datasets/mannu1170/minst-dataset

Shape : The dataset has 42000 rows and 785 columns

Calculation of Accuracy :

1. Train Data Size (Total Number of Predictions) = 8400
2. Correct Number of Predictions = 8090
3. Accuracy = 0.9630952380952381