2018FA_MSDS_422-DL_SEC58

Practical Machine Learning Assignment 6

Prabhat Thakur

1. Problem Definition

This assignment involves benchmark experiment to calculate performance accuracy and processing time of artificial neural networks classification using the MNIST data by testing alternative network structures, activation functions, optimization methods, and/or hyperparameter settings.

2. Research Design and methods

There MNIST data contains 70,000 images, and each image has 784 features. This is because each image is 28×28 pixels, and each feature simply represents one pixel's intensity, from 0 (white) to 255 (black).

We will assess classification performance accuracy and processing time using Python TensorFlow and/or Python Scikit Learn. Tested neural network structures will be explored within a benchmark experiment, a factorial design with three levels on each of two experimental factors (3x3 completely crossed design: 3 hidden layers and 3 different number of nodes per layer).

3. Implementation and Programming:

Due to technical difficulties in installing Python Tensorflow package, Python Scikit Learn package is used in this assignment. Class MLPClassifier is used for artificial neural networks classifier model.

The first step is to download MNIST data from google archive and read it into python ndarray objects so that it can be used in sklearn package. Features values are rescaled from [0, 255] down to [-0.5, 0.5]. The data is then divided into training set (60000 images) and test set (10000 images).

Multiple MLPClassifier models are designed for different number of hidden layer [2,5,10] and different number of nodes [10,20,40] in each layer. This created total 3x3 = 9 models. Processing time and training & test data set accuracy is captured for model comparison.

4. Findings and recommendations

Total 9 models were designed and trained using 2,5 and 10 hidden layers having 10,20 and 40 nodes in each layer. Out of these 9 models, neural network with 5 layers and 40 nodes in each layer produced best predictive accuracy (training set: 0.992083 , test set : 0.9715) in processing time of 96 seconds. The 2nd best model has 2 layers and 40 nodes in each layer which gives predictive accuracy (training set: 0.989300, test set : 0.9699) and processing time of 71 seconds.

Recommendation: For financial institution extreme high accuracy is very important even if a predictive model takes long time to train. From the experimented models, **Neural Network with 5 layers and 40 nodes in each layer should be used** for optical character recognition model.