IT350 Data Analytics

Assignment 5

Part 1:

Design a feed-forward neural network with two hidden-layer nodes and one output node, trains it using the back-propagation algorithm. The problem that the neural network will handle is a multi-class classification problem for identifying the rating of reviews (Dataset-1) *recognizing images of handwritten digits* (Dataset-1). All inputs to the neural network will be converted to numeric. Assume there is a sigmoid activation function at the hidden layer nodes and at the output layer node and Learning rate is 0.1.

Dataset:

https://www.kaggle.com/omkarsabnis/yelp-reviews-dataset

http://yann.lecun.com/exdb/mnist/

1. For the given dataset, based on the outputs of the output nodes, classifies the instance as the index of the output node with the maximum value. Evaluate the predicted values and tabulate the results.
2. Apply dimensionality reduction technique, PCA to the dataset before training the feed-forward neural network. Compare and tabulate the results.