

CS Homework 3

Team:

Nishanth Reddy Pinnapareddy (011426181)

Abhinaya Koduri (011440299)

Experiment 1:

Hypothesis: Increase in training size increases the accuracy.

Description: The neural network consists of five dense layers, the last of which is softmax layer. The hidden layers use rectified linear units. The models are trained using stochastic gradient descent.

We have trained the model with varying training sizes i.e, 5000, 10000 and 20000 and observed the accuracies when tested against testfolder_01 which contains 5000 records.

Results:

Training size	Accuracy
5000	46.53%
10000	70.35%
20000	74.02%
60000	75.25%

Conclusion: The hypothesis is true.

As the size of the data increases, the models learns better and the accuracy increases.

Experiment 2:

Hypothesis 1: The model performs well when the distribution of the classes in the data is randomly chosen.

Description: The neural network consists of five dense layers, the last of which is softmax layer. The hidden layers use rectified linear units. The models are trained using stochastic gradient descent.

Trained using random train data from trainfolder_04. Below are the observed accuracies for the test folders.

Results:

Class	Data size	Accuracy
Testfolder_01	5000	75.25%
Testfolder_02	10000	74.01%
Testfolder_03	20000	74.65%

Conclusion: The hypothesis is True.

Hypothesis 2: Randomly generated class distribution of test data performs better than test data is chosen that is likely to be equally representative of the classes that we are training for.

Description: The neural network consists of five dense layers, the last of which is softmax layer. The hidden layers use rectified linear units. The models are trained using stochastic gradient descent.

We trained our model using evenly chosen train data and tested against randomly chosen test data of 60000 records. Below are the results.

Results: Testing

Class	Data size	Accuracy
Trainfolder_01	5000	61.84%
Trainfolder_02	10000	69.71%
Trainfolder_03	20000	73.84%

Conclusion: The hypothesis is false. The model performs better when the test data is chosen that is likely to be equally representative of the classes that we are training for.

Experiment 3:

Hypothesis: Accuracy increases when cross-validation is used for testing versus separate test data.

Description: The neural network consists of five dense layers, the last of which is softmax layer. The hidden layers use rectified linear units. The models are trained using stochastic gradient descent.

Results:

Testing	Data Size	Accuracy
Cross-validation	10000	82.3%
Separate test data	10000	72.7%

Conclusion: The hypothesis is true.

The accuracy is better when cross-validation is used when compared to separate test data as we have run the experiments on the same set of data 5 times as cross-validation is 5-fold.