Name: Aditi Helekar UNCC ID: 800966667

Course: Machine Learning(ITCS_6156)

Neural Networks Algorithm Report

Name: Aditi Helekar UNCC ID: 800966667

Problem Statement:

Design a Decision Tress classification algorithm for the given datasets.

Approach:

- I have implemented neural networks in Python using scikit-learn library. Scikit-learn is library designed for implementing Machine learning techniques in Python. I have used multilayer percerptron classifier for basic algorithm with Stochastic gradient descent function. The activation function used is 'relu', the rectified linear unit function, returns f(x) = max(0, x).

  The basic approach is described below:

  1. Read the csv file using pandas.io library and store it as a Dataframe.
  2. Classify the attributes as samples and features to apply it to MLPClassifier. Features are also referred as predictors and samples as responses.
  3. Create an object of MLPClassifier imported from scikit learn library. The number of hidden layer neurons is set to mean of input and output. The learning rate is set to 0.001.
  4. Using fit function, fit the dataframe to tree model. Fit function finds patterns in the given dataset.
  5. For cross-validation, divide the training dataset in train and test data using train_test_split function.
  6. Predict the behavior of model using predict function with test data.
  7. After cross validation use the test dataset and predict again.
  8. Calculate accuracy using score function.
  9. Plot the loss graphs using matplotlib library.

  For amazon dataset, I have used bag of words concept to classify the words using CountVectorizer. It gives the occurrences of the words and stores it. This is basically used to feed to algorithm the string input converted to integer form.

Conclusion:

Digit Recognition Dataset:

For digit recognition dataset, the number of inputs is 64 and output is 10. So, the number of hidden layer is mean of input and output rounded to 35. As we increase the hidden layer neurons accuracy increases but it tends to overfit. Therefore, mean is the appropriate choice. Learning rate is set to 0.001 and as we decrease this rate the accuracy increases. Accuracy is directly proportional to number of maximum iterations.

Amazon dataset:

For amazon dataset, the number of inputs is considered as all three names, reviews and ratings and output will be positive or negative review. So, the number of hidden layer neurons is set to 2. We have tested by increasing this number to 10. Accuracy does not increase considerably but the loss reduces.

Per my analysis, Neural Networks algorithm gives much more accuracy than Decision tree algorithm for both the datasets. The best fitted model is the one that most accurately fits your data. Because these trees are so easy to understand, they are very useful as modeling techniques and provide visual representations of the data. Neural networks tends to fit the data better than decision tree model.

References:

1. http://scikit-learn.org/stable/modules/classes.html
2. http://pandas.pydata.org/pandas-docs/stable/
3. http://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html
4. http://rianadam.web.ugm.ac.id/2016/10/21/neural-net-in-4-lines-using-scikit-learn/