CS 280
Programming Assignment 2

Implementing the Backpropagation Algorithm

Instructor: Pros Naval
Due: October 20, 2018

In this Programming Assignment, you will implement an Artificial Neural Network, train it using the Backpropagation Algorithm on a given dataset and compare its performance against a Support Vector Machine classifier.

Instructions

1. Unzip the file CS280_PA2.zip that was provided to you.
2. Study the code named cs280_neural_net.m. Note that this is only a demo code meant to illustrate a simple neural network. You should be guided by the notes you took during our lecture on ANN.
3. Rewrite from scratch the code in Python and revise so that it can accept numeric inputs. You are not allowed to use any code available anywhere else (internet, book, friend, etc).
4. Use a spreadsheet to view the contents of the following dataset files:

   data.csv : 3486 instances each having 354 attributes or features
   data_labels.csv : class labels (1,2,…8) for each of the 3486 instances
   test_set.csv : test instances without labels

5. Partition the dataset *data.csv* into a training set named *training_set.csv* and *training_labels.csv* and a validation set named *validation_set.csv* and *validation_labels.csv*. Please note that we have here a highly imbalanced dataset (i.e., there are huge differences in the number of instances for each class). Ideas on how to deal with this are found in http://stats.stackexchange.com/questions/109177/which-classifiers-work-well-with-unbalanced-data?rq=1
6. Train your ANN on the training set and use the validation set to monitor your prediction errors over training epochs.
7. Change the number of nodes in each hidden layer and the learning rate parameter so that the prediction error over the validation set is minimized. Once the network has learned, predict the labels for each of the test set instances. Save these labels in a csv file named *predicted_ann.csv* with the same format as *data_labels.csv*.
8. Download SVM code (in Python or C) for your SVM classifier.
9. Train the SVM classifier and predict the labels for the test set. Save the predicted labels in a csv file named *predicted_other*.csv.

Deliverables:

a) ANN code (in Python) with self-documenting comments. Please indicate compiler and OS you used.
b) The following csv files:
   i) training_set.csv and *training_labels.csv*
   ii) *validation_set.csv* and *validation_labels.csv*
   iii) *predicted_ann.csv* and *predicted_svm.csv*
c) Documentation (in the same style as the one for PA1) describing the following:
   i) how the dataset was partitioned
   ii) (ANN): how the number of nodes in the hidden layers was selected
   iii) (ANN): plot of training and validation errors vs epoch number
   iv) how you dealt with the highly imbalanced dataset
   v) justification for your kernel choices for the SVM classifier and details on how the classifier was trained
   vi) tables comparing the two classifiers in terms of their accuracies and running times
   vii) Conclusion

Note: Make sure that your ANN code in (a) works correctly, otherwise the rest of your submission will not be checked. Submit the deliverables to submit2pcnaval@gmail.com with "[CS280 PA2 Submission] <Your_Name>" on the subject line by 11:59 pm of October 20, 2017.