MM811 Assignment 3 Report

Ruyi Wang 1277680

1. Description of dataset

The dataset ilpd.csv is downloaded from UCI Machine Learning Repository.
URL:
http://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29#
This dataset contains 416 liver patient and 167 non-liver patient records. It has 10 variables that are age, gender, total Bilirubin, direct Bilirubin, Alkaline Phosphatase, Alamine Aminotransferase, Aspartate Aminotransferase, total proteins, albumin and A/G ratio. It also has a class selector at the 11th column. The selector is labeled by the experts.
Note: There are around 5 lines in data contains missing value. Those lines has been manually deleted.

2. Description of problem

The problem is to split the data into two classes: liver patient and non-liver patient.

3. Description of condition the inputs

To load the data, function loadtxt in numpy module is used. The delimiter is set to ‘,’. The first 10 columns are assigned to variable inputs.

4. Description of interpret outputs

The last column of data is assigned to variable outputs. There are three classes of outputs: 0, 1 and 2. Since there is no input data of class 0, there is no output for class 0.

5. Description of performance of each architecture

For one hidden layer, 3 neurons: ('tp', 39), ('tn', 0), ('fp', 0), ('fn', 0).
Accuracy: 0.7000
For two hidden layers, each has 3 neurons: ('tp', 40), ('tn', 0), ('fp', 0), ('fn', 0).
Accuracy: 0.7545
For three hidden layers, each has 3 neurons: ('tp', 43), ('tn', 0), ('fp', 0), ('fn', 0).
Accuracy : 0.7953

6. Possible ways of improving

There are two ways to improve result.
One is to increase the number of hidden layers. From the testing result, the accuracy seems increased in the case of higher number of hidden layers. The value of true positive seems increased as well.

Another way is to let the number of neurons equals to the times of class numbers.

7. Summary

From the testing result, the deep learning solution solves the problem with satisfactory accuracy. Higher accuracy and true positive value can be obtained by increasing the number of hidden layers but costing longer running time of program.

8. Reference

Code is written based on the theanets-tutorial from:
https://github.com/abramhindle/theanets-tutorial.git
Dataset is downloaded from:
http://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29#