Homework 6

IE 7275 Data Mining in Engineering

Note: Read the following literature before you attempt to solve the problems in this homework.

Neural network related documents

Documentation on neuralnet.pdf
Examples on neuralnet.pdf
Reference manual on neuralnet.pdf
Documentation on rsnns.pdf (focus on mlp)
Reference manual on rsnns.pdf (focus on mlp)
Reference manual on nnet.pdf
Visualizing neural networks in R.pdf
Using neural networks for credit scoring.pdf

Lift charts related documents

Tutorial on lift charts with R.pdf
Documentation on lift.pdf
Lift charts to compare binary predictive models.pdf
Documentation on gain.pdf

Problem 1 (Car Sales, Neural Networks) [50 points]

Consider the data on used cars given in **`ToyotaCorolla.xlsx`**. The data has 1436 records and details on 38 attributes, including Price, Age, KM, HP, and other specifications. The goal is to predict the price of a used Toyota Corolla based on its specifications using a multilayer neural network. Select appropriate predictor variables.

Use 75% of the data for training a multilayer neural network and 25% to validate the network performance. Use a traditional (standard) backpropagation algorithm.

Record the RMS error for the training data and the validation data. Repeat the process, changing the number of epochs or steps (and only the epochs keeping all other network parameters the same) to 200, 2000, 4000, 6000, 8000, and 10,000. Set the `stepmax` to desired number of steps (iterations), and set `threshold` to a very small value, say 0.000001. Set the number of repetitions equal to 1 (`rep` = 1) in each case.

(a) What happens to the RMS error (or Sum of Squares Error) for the training data as the number of epochs (or steps) increases?

(b) What happens to the RMS error Sum of Squares Error for the validation data?

(c) Comment on the appropriate number of epochs for the model.

(d) Conduct a similar experiment to assess the effect of changing the number of layers from 1 to 2 in the network.

(e) Study the effect of gradient descent step size (`learningrate`) on the training process and the network performance.

Problem 2 (Airline Customers, Neural Networks) [50 points]

East-West Airlines has entered into a partnership with the wireless phone company Telcon to sell the latter's service via direct mail. The file **`East West Airlines.xlsx`** contains a subset of a data sample of who has already received a test offer. About 13% accepted.

You are asked to develop a multilayer neural network model to classify East-West customers as to whether they purchased a wireless phone service contract (target variable Phone_Sale). Using this model you are asked to predict classifications for additional customers. Use 75% of the data for training a multilayer neural network and 25% to validate the network performance. Use a traditional (standard) backpropagation algorithm. Select appropriate predictor variables. Set the number of repetitions equal to 1 (`rep` = 1) in each case.

Run a neural net model on these data, after normalize the data, setting the number of epochs or steps at 3000. Set the `stepmax` to desired number of steps, and set `threshold` to a very small value, say 0.000001. Use `lift.chart` function in R to create lift chart (with `cumulative` response) and decile lift chart (with `incremental` response).

(a) Construct a lift charts for both the training and validation data.

(b) Construct a decile lift charts for both the training and validation data.

(c) Interpret the meaning (in business terms) of the leftmost bar of the validation lift chart (the bar chart).

(d) Comment on the difference between the training and validation lift charts.

(e) Run a second neural net model on the data, this time setting the number of epochs equal to 100.

(f) Repeat steps (a) and (b) for the second neural network.

(g) Comment now on the difference between this model and the model you ran earlier, and how over-fitting might have affected results.

(h) What sort of information, if any, is provided about the effect of the variables?

Files Included in the Folder:

1. **Homework 6.pdf**
2. **Documentation on neuralnet.pdf**
3. **Examples on neuralnet.pdf**
4. **Reference manual on neuralnet.pdf**
5. **Documentation on rsnns.pdf (focus on mlp)**
6. **Reference manual on rsnns.pdf (focus on mlp)**
7. **Reference manual on nnet.pdf**
8. **Visualizing neural networks in R.pdf**
9. **Using neural networks for credit scoring.pdf**
10. **Tutorial on lift charts with R.pdf**
11. **Documentation on lift.pdf**
12. **Lift charts to compare binary predictive models.pdf**
13. **Documentation on gain.pdf**
14. **Toyta Corolla.xlsx**
15. **East West Airlines.xlsx**