SOLUTIONS FOR ASSIGNMENT 3

REGRESSION TREES

UTD NETID: AMK210012

a Neural network with single hidden layer and 2 nodes.

Chart

Description automatically generated

RMSE values for Neural network with single hidden layer and 2 nodes:

Validation Data: 1405.026

Training Data: 1339.066

b. Neural network with a single hidden layer and 5 nodes.

Diagram

Description automatically generated

RMSE values for Neural network with a single hidden layer and 5 nodes.

Training Data: 1209.056

Validation Data: 1358.148

Diagram

Description automatically generated

RMSE values for Neural network with double hidden layer and 5 nodes.

Training Data: 1441.192

Validation Data: 1534.825

Chart, diagram, radar chart

Description automatically generated

RMSE values for Neural network with single hidden layer and 6 nodes.

Training Data: 1465.242

Validation Data: 1735.219

1. The RMS Error initially decreases as we increase the number of nodes from 2 to 5 on training data. But when we increase the number of layers to 2, the RMS Error increases for training data.
2. The RMS Error initially decreases as we increase the number of nodes from 2 to 5 on validation data. But when we increase the number of layers to 2, the RMS Error increases for validation data.
3. The appropriate number of layers would be one and also the appropriate number of nodes would be 5 since these parameters give the least errors amongst all the other parameters.

RMSE Values Comparison

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Training Data RMSE** | **Validation Data RMSE** |
| Neural Network Single Layer 2 nodes | 1339.066 | 1405.026 |
| Neural network with a single hidden layer and 5 nodes | 1209.056 | 1358.148 |
| Neural network with double hidden layer and 5 nodes | 1441.192 | 1534.825 |
|  |  |  |

b. By comparing the RMSE values, we can see that Regression trees have lesser RMSE values for both training and validation datasets as compared to neural networks. Hence, for this particular problem, Regression Trees perform better than Neural Networks.

Regression Analysis summary

The Toyota Corolla dataset has variables such as distance travelled by the car, age of the car, fuel type of the car etc. The task is to predict the price of a car using the variables. Hence, this is a regression problem.

Neural Networks were used to predict the price of a car. Because Neural networks are ‘black box’ algorithms, it hard to ascertain the relationship between the variables.

The following nodes and layers were used to build the model with the threshold parameter being 0.5 .

Single layer 2 nodes

Single layer 5 nodes

Double layer 5 nodes

The RMSE was the lowest on Single layer – 5 nodes. Regression Trees, however, had lower RMSE values compared to neural networks.