**CASE STUDY 1**

1. a. Data frame dimensions, i.e. Number of rows and columns in data set: (506, 14).

b. **Original column titles: Python Output**

Index(['CRIME', 'ZONE', 'INDUST', 'CHAR RIV', 'NIT OXIDE', 'ROOMS', 'AGE','DISTANCE', 'RADIAL', 'TAX', 'ST RATIO', 'LOW STAT', 'MVALUE','C MVALUE'], dtype='object')

**Modified column titles with no space and one word for titles: Python Output**

Index(['CRIME', 'ZONE', 'INDUST', 'CHAR\_RIV', 'NIT\_OXIDE', 'ROOMS', 'AGE', 'DISTANCE', 'RADIAL', 'TAX', 'ST\_RATIO', 'LOW\_STAT', 'MVALUE', 'C\_MVALUE'], dtype='object')

c. **Modified list of column variables with dummy variables: Python Output**

Index(['CRIME', 'ZONE', 'INDUST', 'NIT\_OXIDE', 'ROOMS', 'AGE', 'DISTANCE', 'RADIAL', 'TAX', 'ST\_RATIO', 'LOW\_STAT', 'MVALUE', 'CHAR\_RIV\_Y', 'C\_MVALUE\_Yes'], dtype='object')

The outcome variable is "MVALUE"

The predictor variables are ['CRIME', 'ZONE', 'INDUST', 'NIT\_OXIDE', 'ROOMS', 'AGE', 'DISTANCE', 'RADIAL', 'TAX', 'ST\_RATIO', 'LOW\_STAT', 'CHAR\_RIV\_Y', 'C\_MVALUE\_Yes']

1. **a. The first five records of the predictors for training partition before and after standardizing/scaling are given below.**

Table

Description automatically generated

The predictor variables of both training and validation partitions are scaled or standardized in order to eliminate the problem of domination of one variable with large values (in 10000’s) over the other variables with low values (in 10’s). The outcome variable need not be scaled in neural network model.

**b.** After fitting the neural network model for the scaled training partition using MLPRegressor() using the following parameters (hidden\_layer\_sizes=9, solver=’lbfgs’, max\_iter=10000, and random\_state=1) in Python, the final intercepts and the weights of this model output are displayed below.

Table

Description automatically generated

There are 13 input layers (predictors), 9 hidden layers and one output layer (outcome) in the neural network.

The first array in the final intercepts represents the thetas (θj-output from node j), i.e. the final node bias values associated with each hidden layer and the 2nd array represent the theta i.e. final node bias value associated with the output layer.

The first array in the network weights represents the final coefficients or weights associated with the connecting arrows from node i of the input layer to node j of the hidden layer and 2nd array represents the final coefficients or weights associated with the connecting arrows from node i of the hidden layer to node j of the output layer.

* **Weights wij** - the final values on the connecting arrows from node i to node j; like coefficients, and subject to iterative adjustments
* **Node bias values θj** – output from node j; a constant that controls the level of contribution of node j like intercept; subject to iterative adjustments

**c.**

Text, table

Description automatically generated

The positive residual values indicate that values of the outcome variable in the validation partition are under-predicted (Actual - Prediction = +ve) and the negative values indicate that values of the outcome variable in the validation partition are over-predicted (Actual – Prediction = -ve ). The first five records of the predictions for validation partition are shown in the table above. In the first five records, the records (343, 67) are under-predicted and the records (307, 47, 362) are over-predicted.

**d.**

Text

Description automatically generated with medium confidence

There is no much difference in the RMSE’s of training and validation sets, we can say there is no possibility of overfitting of the data. It indicates that the neural network model can be used for making predictions. The MAPE’s of training and validation sets are less than 15% i.e. the margin of errors are less than 15% and are not much different, hence based on the MAPE’s, we can say this model has no possibility of overfitting of the data and can be considered for making predictions.

1. **a.** The GridSearchCV() function, with hidden\_layer\_sizes parameter in a range from 2 to 20, is used to identify the best number of nodes for the hidden layer in the Boston Housing neural network model and results of the the best score and best parameter values are presented below.

Best score:0.8759

Best parameter: {'hidden\_layer\_sizes': 2}

**b. The results of the improved neural network model with the best parameter obtained from the Grid Search:**

**Text

Description automatically generated**

**c. The accuracy measures of the improved neural network model with the best parameter obtained from the Grid Search:**

Text

Description automatically generated

There is no much difference in the RMSE’s of training and validation sets, we can say there is no possibility of overfitting of the data. It indicates that the neural network model based on Grid Search can be used for making predictions. The MAPE’s of training and validation sets are less than 15% i.e. the margin of errors are less than 15% and are not much different, hence based on the MAPE’s, we can say this model has no possibility of overfitting of the data and can be considered for making predictions.

**d. Mathematical equation for Linear regression Model based on Exhaustive Search from Case study 1:**

43.89 + 2.13 \* CHAR\_RIV\_Y -0.14 \* CRIME + 11.11 \* C\_MVALUE\_Yes -0.63 \* DISTANCE + 0.11 \* INDUST -0.46 \* LOW\_STAT -16.89 \* NIT\_OXIDE + 0.19 \* RADIAL + 0.86 \* ROOMS -0.61 \* ST\_RATIO -0.01 \* TAX

Accuracy Measures for Validation Set Using Exhaustive Search

Regression statistics

Mean Error (ME) : 0.4505

Root Mean Squared Error (RMSE) : 3.8674

Mean Absolute Error (MAE) : 2.7724

Mean Percentage Error (MPE) : -2.1963

Mean Absolute Percentage Error (MAPE) : 13.3441

Best Variables (11) from Exhaustive Search Algorithm

['C\_MVALUE\_Yes', 'LOW\_STAT', 'CRIME', 'CHAR\_RIV\_Y', 'ST\_RATIO', 'ROOMS', 'DISTANCE', 'NIT\_OXIDE', 'RADIAL', ‘INDUST’, ‘TAX’]

**Model** **No. of predictors RMSE of Valid partition**

1) Improved Neural network model 13 3.0570

with all predictors

2) Model based on Exhaustive search 11 3.8674

The RMSE is lower for the improved neural network model with all the 13 predictors than the RMSE’s of the multiple linear regression model based on the Exhaustive search with 11 number of predictors. The lower RMSE, the better the predictions. Based on RMSE, the improved neural network model with all 13 predictors is recommended for making predictions.

However, the RMSE’s are not much different for two models. The parsimonious model which is the multiple linear regression model based on the Exhaustive search with 11 number of predictors can be recommended for making predictions. The simpler and computationally cheaper model that performs better results.