Neural Network Modeling Using R

1. Explore the Data

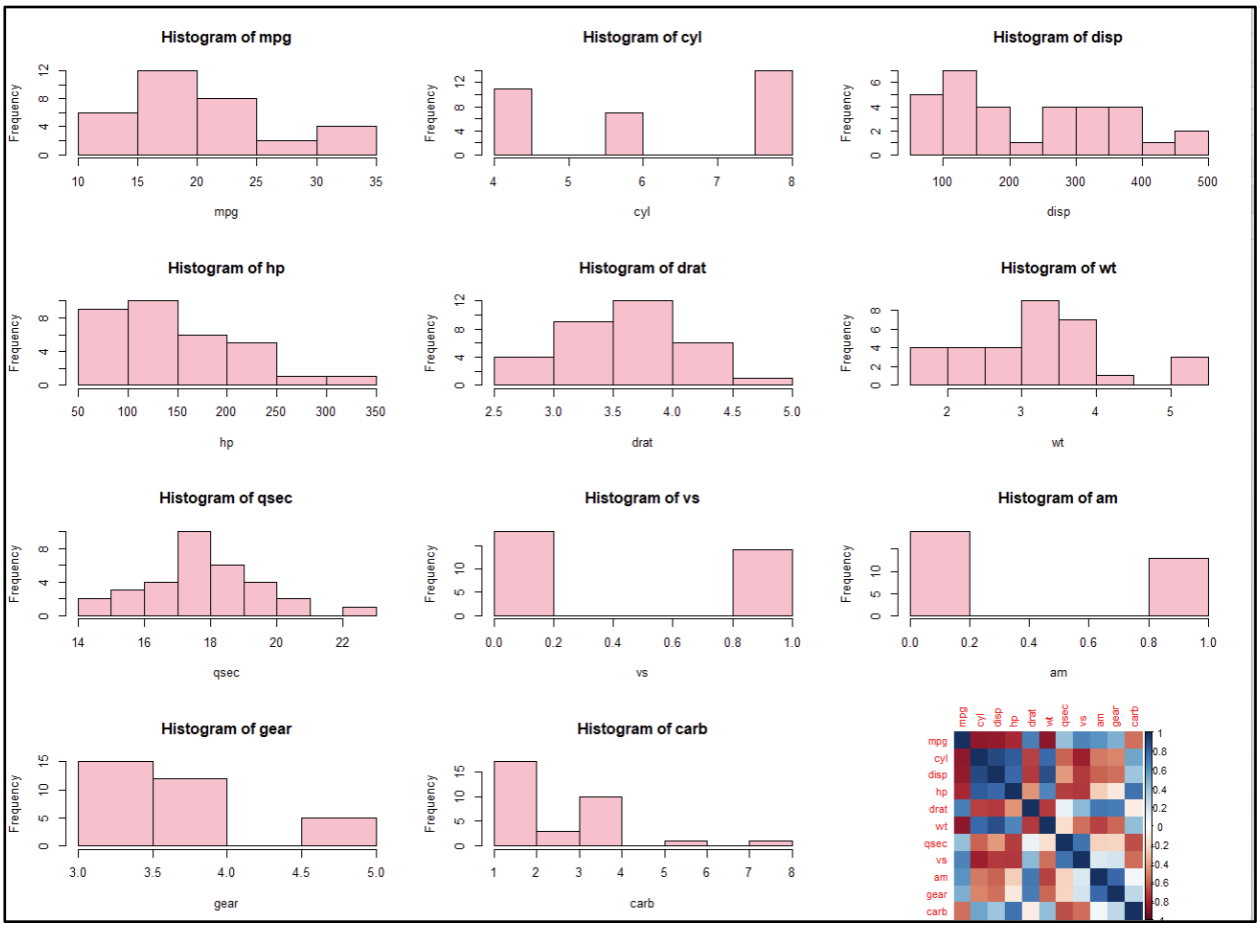

Figure 1: Histograms and correlation plot

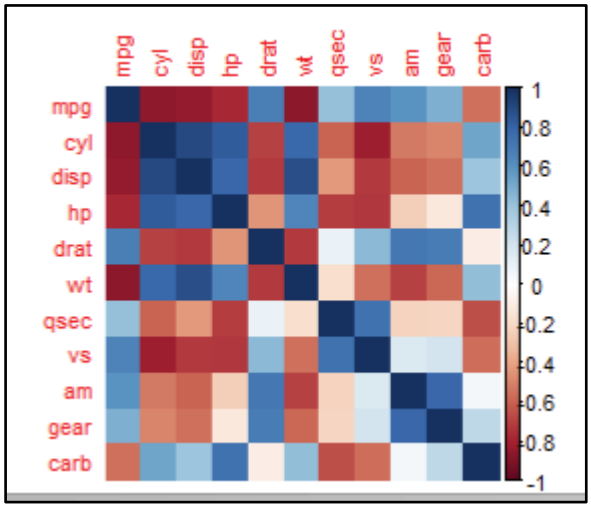

Figure 2: Correlation plot

mpg	cyl	disp	hp	drat
Min. :10.40	Min. :4.000	Min. : 71.1	Min. : 52.0	Min. :2.760
1st Qu.:15.43	1st Qu.:4.000	1st Qu.:120.8	1st Qu.: 96.5	1st Qu.:3.080
Median :19.20	Median :6.000	Median :196.3	Median :123.0	Median :3.695
Mean :20.09	Mean :6.188	Mean :230.7	Mean :146.7	Mean :3.597
3rd Qu.:22.80	3rd Qu.:8.000	3rd Qu.:326.0	3rd Qu.:180.0	3rd Qu.:3.920
Max. :33.90	Max. :8.000	Max. :472.0	Max. :335.0	Max. :4.930

wt	qsec	vs	am	gear	carb
Min. :1.513	Min. :14.50	Min. :0.0000	Min. :0.0000	Min. :3.000	Min. :1.000
1st Qu.:2.581	1st Qu.:16.89	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:3.000	1st Qu.:2.000
Median :3.325	Median :17.71	Median :0.0000	Median :0.0000	Median :4.000	Median :2.000
Mean :3.217	Mean :17.85	Mean :0.4375	Mean :0.4062	Mean :3.688	Mean :2.812
3rd Qu.:3.610	3rd Qu.:18.90	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:4.000	3rd Qu.:4.000
Max. :5.424	Max. :22.90	Max. :1.0000	Max. :1.0000	Max. :5.000	Max. :8.000

Figure 3: Summary statistics

The goal of this assignment is to predict a car's fuel efficiency by comparing how many miles it can drive per gallon of gas with other factors. These factors include, but are not limited to, weight, forward gears, and the number of cylinders in the engine.

When examining the summary statistics (Figure 3), we can see that the average mpg for a car is 20, with a minimum of 10.4 and a maximum of approximately 34 mpg. On average, the car's engine has 6 cylinders and 4 forward gears. The average weight of the vehicles in the dataset is 3,217 lbs, with a maximum of 5,424 lbs and a minimum of 1,513 lbs. The histograms do not show any clear outliers, and there are no missing datapoints to note.

Finally, we can see how each variable correlates with another when examining Figure 2. In this analysis, it is most important that we compare the dependent variable (mpg) with all the independent variables. The MPG of the car has a high negative correlation with the number of cylinders, the engine displacement, the horsepower of the car, and the weight of the car. The MPG of the car has a high positive correlation with the rear axle ratio, the engine shape, the transmission type, and the number of forward gears.

2. Neural Network Modeling:
    a. Model 1:
        - Model 1 is the most stable, with low prediction errors and a moderate gap (0.115) between RMSE training (0.050) and validation data (0.165). This suggests that the model has learned and generalized well. However,

predictions are less accurate than in model 2 which achieves a MAE of 3.22 compared to model 1's 3.27. The rescale prediction value shows that there is overprediction and underprediction, but the model demonstrates a reasonable prediction accuracy overall.

b. Model 2:
   - Model 2 has balancing models with the lowest MAE (3.22), indicating more accurate forecasts with smaller average errors than other models. The gap between normalized training (0.041) and validation (0.169) RMSE is moderate, which shows reasonable generalization. The rescaled RMSE values are 0.97 for training and 3.97 for validation. Model 2 has better training performance than model 1 due to tis lower RMSE, making it a better predictor of actual MPG values..

c. Model 3:
   - Model 3 performs the worst. The reason for that is they have smallest RMSE training data (0.038), meaning the model learns the training data too perfectly as it fits almost every fluctuation, noise and pattern while it performs very poorly on RMSE validation data as it is significantly larger than training data (0.213), showing a sign of overfitting and poor generalization. We can also find overfitting if we look at the normalized predictions as some values are very close to 0. Furthermore, rescaled MAE and RMSE are also the highest compared to other two (4.04 and 5.00 approximately).

3. Model Evaluation:
   - Model 1 has simple architecture. It produced moderate training and validation error, resulting in good generalization, however model 2 performs somewhat better in predicting real value. Model 2 got deeper and added small layers. It also produced moderate training and validation errors, though slightly worse than simple ones. However, when it comes to predicting real values, it outperforms the competition with the lowest MAE being 3.22. Model 3 got more complex (3 wider layers) when it got less deep but wider layers than model 2. Although it gave lowest training error (0.89) it also gave biggest validation error (5.00 rescaled RSME), leading to overfitting and reduced generalization. It also has worst performance when predicting actual values as highest rescaled MAE (4.04) and RMSE (5.00).
   - Model 2 is the best choice for optimal performance. Because it not only has similar and steady training and validation data to Model 1, but it also has the lowest MAE when predicting original mpg values.
   - There are some relationships between predictors and MPG. For example, the weight variable influences mpg since big cars have low MPG whereas small cars have high MPG, implying that more mass equals more gasoline required to run. Cars with high horsepower, cylinders, and displacement also consume more gasoline, resulting in low MPG.