**Application**

**Return to the Vehicle data used in the previous lecture. Use the same split as**

**before.**

set . seed (46685326 , kind =" Mersenne - Twister ")

perm <- sample (x= nrow ( vehdata ))

set1 <- vehdata [ which ( perm <= 3\* nrow ( vehdata )/4) , ]

set2 <- vehdata [ which ( perm > 3\* nrow ( vehdata )/4) , ]

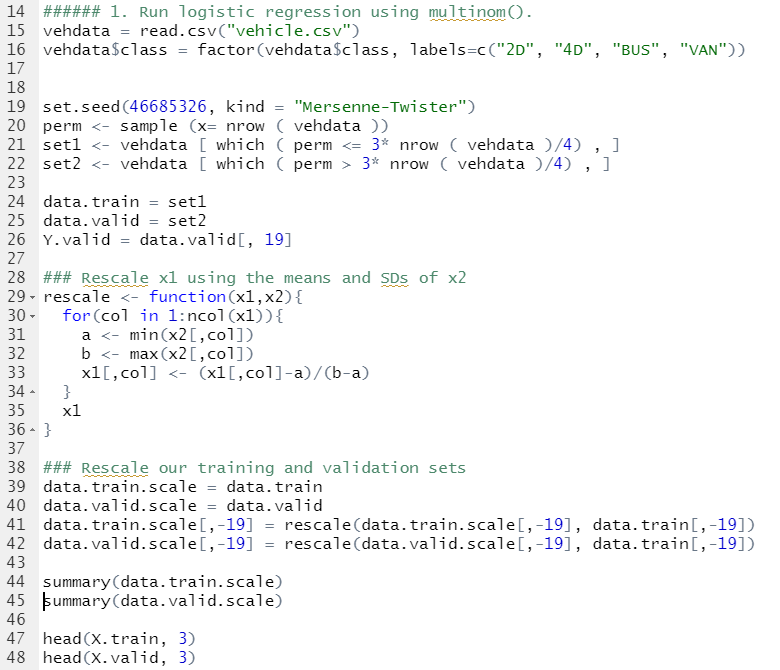
1. Run logistic regression using multinom().

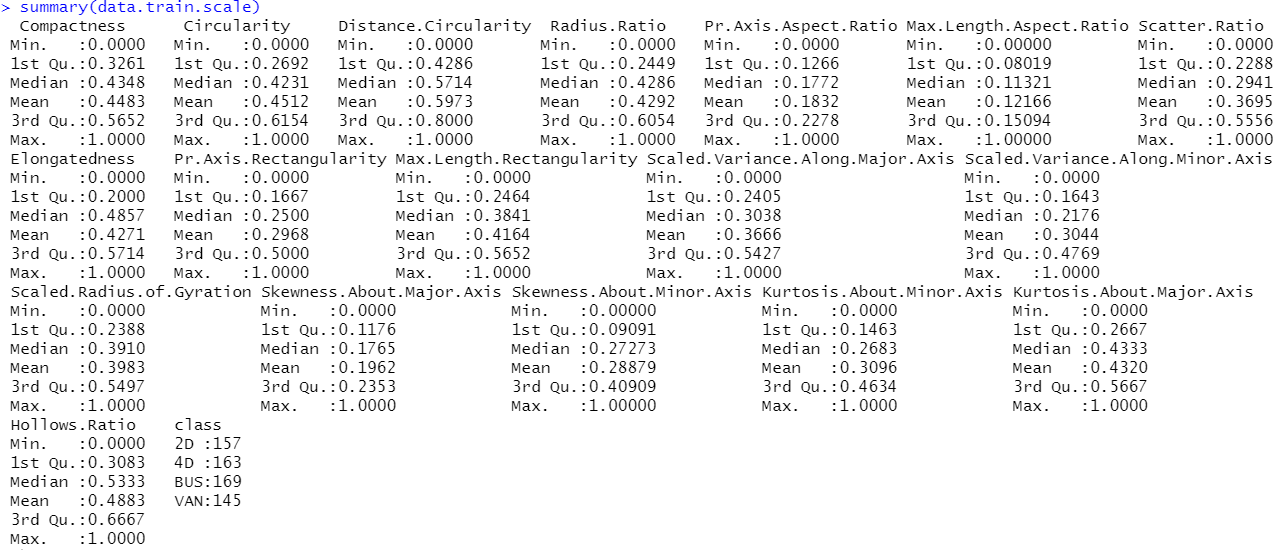
(a) Scale the training data to lie between 0 and 1, and use the same min and max values

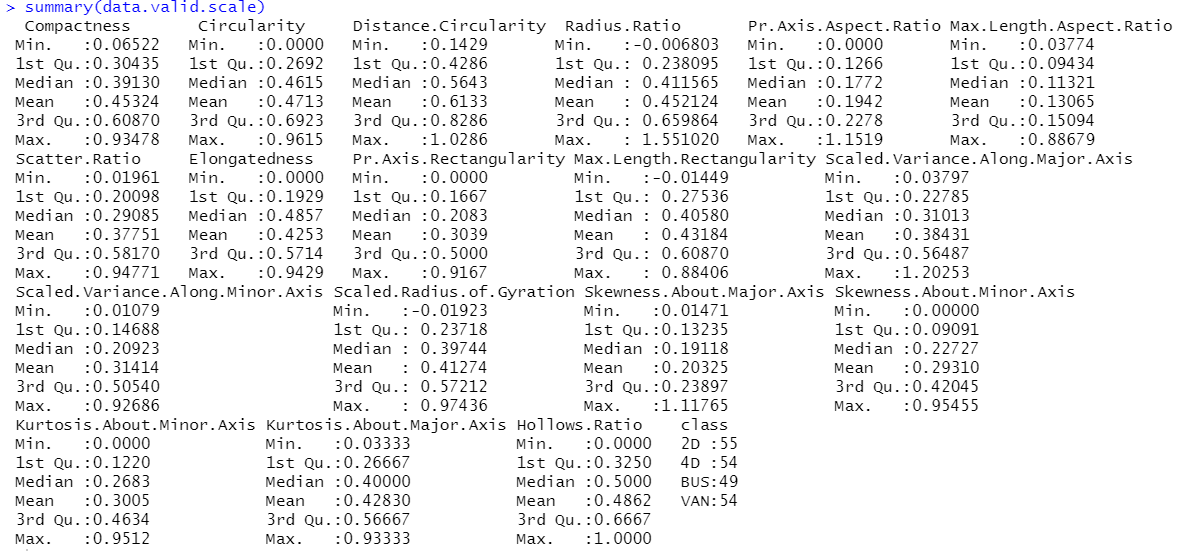
to scale the test data. Run summary() in each scaled data set to confirm that

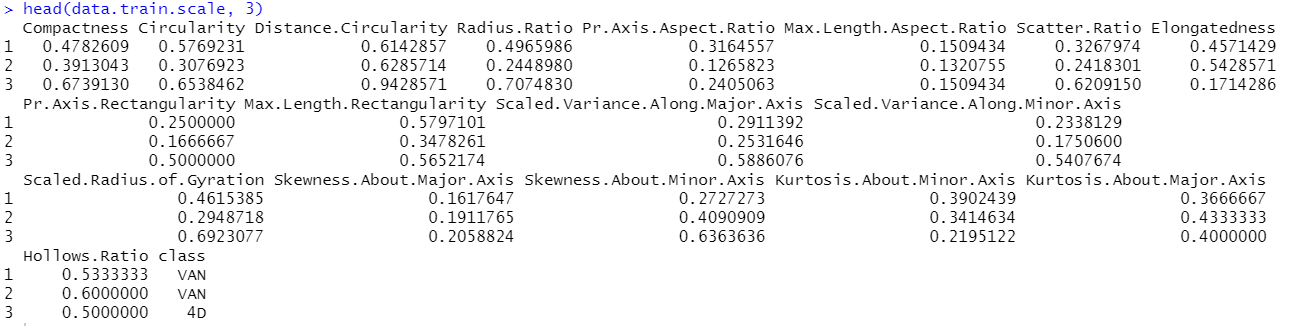
you are doing this correctly. **Report the summary of the first 3 variables**

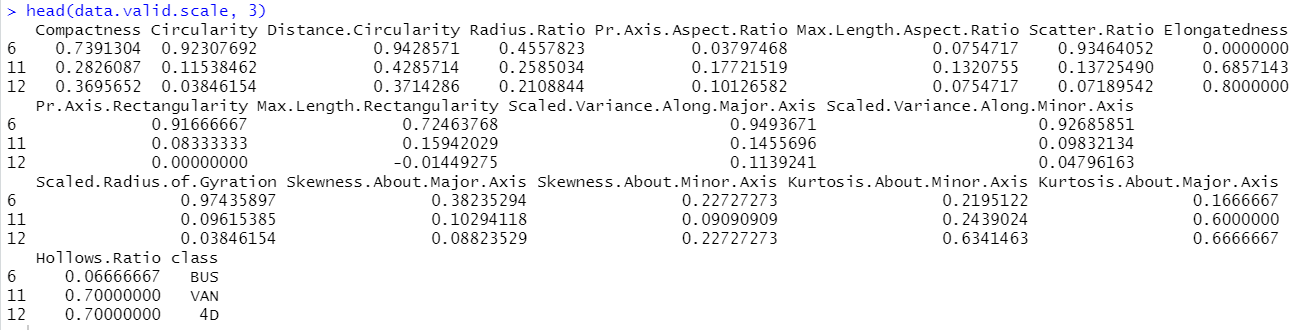
**in each set.**



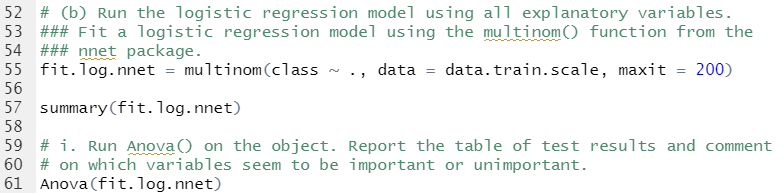






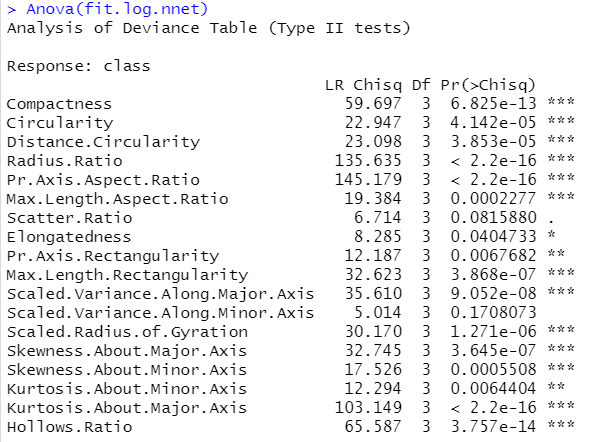


(b) Run the logistic regression model using all explanatory variables.



i. Run Anova() on the object. **Report the table of test results and comment**

**on which variables seem to be important or unimportant.**

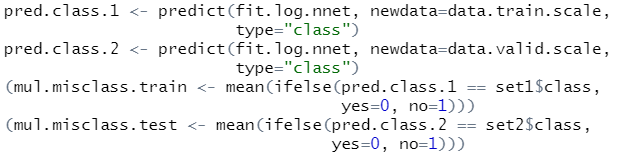


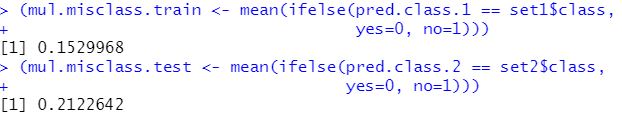
Variables with 2~3 starts look important and variables without star looks unimportant.

ii. Compute and **report training and test error**. **Does test error seem**

**better or worse than optimal KNN?** (Use the standard error computed

before to help you make a sensible comment here.)





These are better than optimal KNN. Values from KNN were above 0.35.

iii. **Report the confusion matrix and comment sensibly on what it tells**

**you.**

