

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.**

```
14 ##### 1. Run logistic regression using multinom().
15 vehdata = read.csv("vehicle.csv")
16 vehdata$class = factor(vehdata$class, labels=c("2D", "4D", "BUS", "VAN"))
17
18
19 set.seed(46685326, kind = "Mersenne-Twister")
20 perm <- sample(x = nrow(vehdata))
21 set1 <- vehdata[which(perm <= 3 * nrow(vehdata) / 4), ]
22 set2 <- vehdata[which(perm > 3 * nrow(vehdata) / 4), ]
23
24 data.train = set1
25 data.valid = set2
26 Y.valid = data.valid[, 19]
27
28 ### Rescale x1 using the means and SDs of x2
29 rescale <- function(x1, x2){
30   for(col in 1:ncol(x1)){
31     a <- min(x2[, col])
32     b <- max(x2[, col])
33     x1[, col] <- (x1[, col] - a) / (b - a)
34   }
35   x1
36 }
37
38 ### Rescale our training and validation sets
39 data.train.scale = data.train
40 data.valid.scale = data.valid
41 data.train.scale[, -19] = rescale(data.train.scale[, -19], data.train[, -19])
42 data.valid.scale[, -19] = rescale(data.valid.scale[, -19], data.train[, -19])
43
44 summary(data.train.scale)
45 summary(data.valid.scale)
46
47 head(X.train, 3)
48 head(X.valid, 3)
```

```
> summary(data.train.scale)
  Compactness      Circularity Distance.Circularity Radius.Ratio Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
1st Qu.:0.3261   1st Qu.:0.2692   1st Qu.:0.4286   1st Qu.:0.2449   1st Qu.:0.1266   1st Qu.:0.08019   1st Qu.:0.2288
Median :0.4348   Median :0.4231   Median :0.5714   Median :0.4286   Median :0.1772   Median :0.11321   Median :0.2941
Mean   :0.4483   Mean   :0.4512   Mean   :0.5973   Mean   :0.4292   Mean   :0.1832   Mean   :0.12166   Mean   :0.3695
3rd Qu.:0.5652   3rd Qu.:0.6154   3rd Qu.:0.8000   3rd Qu.:0.6054   3rd Qu.:0.2278   3rd Qu.:0.15094   3rd Qu.:0.5556
Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000

Elongatedness Pr.Axis.Rectangularity Max.Length.Rectangularity Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.2000   1st Qu.:0.1667   1st Qu.:0.2464   1st Qu.:0.2405   1st Qu.:0.1643
Median :0.4857   Median :0.2500   Median :0.3841   Median :0.3038   Median :0.2176
Mean   :0.4271   Mean   :0.2968   Mean   :0.4164   Mean   :0.3666   Mean   :0.3044
3rd Qu.:0.5714   3rd Qu.:0.5000   3rd Qu.:0.5652   3rd Qu.:0.5427   3rd Qu.:0.4769
Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000

Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis Kurtosis.About.Minor.Axis Kurtosis.About.Major.Axis
Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.2388   1st Qu.:0.1176   1st Qu.:0.09091   1st Qu.:0.1463   1st Qu.:0.2667
Median :0.3910   Median :0.1765   Median :0.27273   Median :0.2683   Median :0.4333
Mean   :0.3983   Mean   :0.1962   Mean   :0.28879   Mean   :0.3096   Mean   :0.4320
3rd Qu.:0.5497   3rd Qu.:0.2353   3rd Qu.:0.40909   3rd Qu.:0.4634   3rd Qu.:0.5667
Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000

Hollows.Ratio class
Min.   :0.0000   2D :157
1st Qu.:0.3083   4D :163
Median :0.5333   BUS:169
Mean   :0.4883   VAN:145
3rd Qu.:0.6667
Max.   :1.0000
```

```
> summary(data.valid.scale)
  Compactness      Circularity Distance.Circularity Radius.Ratio Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio
Min.   :0.06522   Min.   :0.0000   Min.   :0.1429   Min.   :-0.006803   Min.   :0.0000   Min.   :0.03774
1st Qu.:0.30435   1st Qu.:0.2692   1st Qu.:0.4286   1st Qu.:0.238095   1st Qu.:0.1266   1st Qu.:0.09434
Median :0.39130   Median :0.4615   Median :0.5643   Median :0.411565   Median :0.1772   Median :0.11321
Mean   :0.45324   Mean   :0.4713   Mean   :0.6133   Mean   :0.452124   Mean   :0.1942   Mean   :0.13065
3rd Qu.:0.60870   3rd Qu.:0.6923   3rd Qu.:0.8286   3rd Qu.:0.659864   3rd Qu.:0.2278   3rd Qu.:0.15094
Max.   :0.93478   Max.   :0.9615   Max.   :1.0286   Max.   :1.551020   Max.   :1.1519   Max.   :0.88679

Scatter.Ratio Elongatedness Pr.Axis.Rectangularity Max.Length.Rectangularity Scaled.Variance.Along.Major.Axis
Min.   :0.01961   Min.   :0.0000   Min.   :0.0000   Min.   :-0.01449   Min.   :0.03797
1st Qu.:0.20098   1st Qu.:0.1929   1st Qu.:0.1667   1st Qu.:0.27536   1st Qu.:0.22785
Median :0.29085   Median :0.4857   Median :0.2083   Median :0.40580   Median :0.31013
Mean   :0.37751   Mean   :0.4253   Mean   :0.3039   Mean   :0.43184   Mean   :0.38431
3rd Qu.:0.58170   3rd Qu.:0.5714   3rd Qu.:0.5000   3rd Qu.:0.60870   3rd Qu.:0.56487
Max.   :0.94771   Max.   :0.9429   Max.   :0.9167   Max.   :0.88406   Max.   :1.20253

Scaled.Variance.Along.Minor.Axis Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis
Min.   :0.01079   Min.   :-0.01923   Min.   :0.01471   Min.   :0.00000
1st Qu.:0.14688   1st Qu.:0.23718   1st Qu.:0.13235   1st Qu.:0.09091
Median :0.20923   Median :0.39744   Median :0.19118   Median :0.22727
Mean   :0.31414   Mean   :0.41274   Mean   :0.20325   Mean   :0.29310
3rd Qu.:0.50540   3rd Qu.:0.57212   3rd Qu.:0.23897   3rd Qu.:0.42045
Max.   :0.92686   Max.   :0.97436   Max.   :1.11765   Max.   :0.95455

Kurtosis.About.Minor.Axis Kurtosis.About.Major.Axis Hollows.Ratio class
Min.   :0.0000   Min.   :0.03333   Min.   :0.0000   2D :55
1st Qu.:0.1220   1st Qu.:0.26667   1st Qu.:0.3250   4D :54
Median :0.2683   Median :0.40000   Median :0.5000   BUS:49
Mean   :0.3005   Mean   :0.42830   Mean   :0.4862   VAN:54
3rd Qu.:0.4634   3rd Qu.:0.56667   3rd Qu.:0.6667
Max.   :0.9512   Max.   :0.93333   Max.   :1.0000
```

```
> head(data.train.scale, 3)
  Compactness Circularity Distance.Circularity Radius.Ratio Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio Elongatedness
1  0.4782609  0.5769231      0.6142857    0.4965986      0.3164557      0.1509434    0.3267974    0.4571429
2  0.3913043  0.3076923      0.6285714    0.2448980      0.1265823      0.1320755    0.2418301    0.5428571
3  0.6739130  0.6538462      0.9428571    0.7074830      0.2405063      0.1509434    0.6209150    0.1714286

  Pr.Axis.Rectangularity Max.Length.Rectangularity Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis
1      0.2500000      0.5797101      0.2911392      0.2338129
2      0.1666667      0.3478261      0.2531646      0.1750600
3      0.5000000      0.5652174      0.5886076      0.5407674

  Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis Kurtosis.About.Minor.Axis Kurtosis.About.Major.Axis
1      0.4615385      0.1617647      0.2727273      0.3902439      0.3666667
2      0.2948718      0.1911765      0.4090909      0.3414634      0.4333333
3      0.6923077      0.2058824      0.6363636      0.2195122      0.4000000

  Hollows.Ratio class
1  0.5333333 VAN
2  0.6000000 VAN
3  0.5000000 4D
```

```
> head(data.valid.scale, 3)
  Compactness Circularity Distance.Circularity Radius.Ratio Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio Elongatedness
6  0.7391304  0.9230769      0.9428571    0.4557823      0.03797468      0.0754717    0.93464052    0.0000000
11 0.2826087  0.1153846      0.4285714    0.2585034      0.17721519      0.1320755    0.13725490    0.6857143
12 0.3695652  0.03846154      0.3714286    0.2108844      0.10126582      0.0754717    0.07189542    0.8000000

  Pr.Axis.Rectangularity Max.Length.Rectangularity Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis
6      0.9166667      0.72463768      0.9493671      0.92685851
11     0.08333333      0.15942029      0.1455696      0.09832134
12     0.00000000      -0.01449275      0.1139241      0.04796163

  Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis Kurtosis.About.Minor.Axis Kurtosis.About.Major.Axis
6      0.97435897      0.38235294      0.22727273      0.2195122      0.1666667
11     0.09615385      0.10294118      0.09090909      0.2439024      0.6000000
12     0.03846154      0.08823529      0.22727273      0.6341463      0.6666667

  Hollows.Ratio class
6  0.06666667 BUS
11 0.70000000 VAN
12 0.70000000 4D
```

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

```
52 # (b) Run the logistic regression model using all explanatory variables.
53 ### Fit a logistic regression model using the multinom() function from the
54 ### nnet package.
55 fit.log.nnet = multinom(class ~ ., data = data.train.scale, maxit = 200)
56
57 summary(fit.log.nnet)
58
59 # i. Run Anova() on the object. Report the table of test results and comment
60 # on which variables seem to be important or unimportant.
61 Anova(fit.log.nnet)
```

i. Run Anova() on the object. **Report the table of test results and comment on which variables seem to be important or unimportant.**

```
> Anova(fit.log.nnet)
```

Analysis of Deviance Table (Type II tests)

Response: class

	LR	Chisq	Df	Pr(>Chisq)	
Compactness	59.697	3	6.825e-13	***	
Circularity	22.947	3	4.142e-05	***	
Distance.Circularity	23.098	3	3.853e-05	***	
Radius.Ratio	135.635	3	< 2.2e-16	***	
Pr.Axis.Aspect.Ratio	145.179	3	< 2.2e-16	***	
Max.Length.Aspect.Ratio	19.384	3	0.0002277	***	
Scatter.Ratio	6.714	3	0.0815880	.	
Elongatedness	8.285	3	0.0404733	*	
Pr.Axis.Rectangularity	12.187	3	0.0067682	**	
Max.Length.Rectangularity	32.623	3	3.868e-07	***	
Scaled.Variance.Along.Major.Axis	35.610	3	9.052e-08	***	
Scaled.Variance.Along.Minor.Axis	5.014	3	0.1708073		
Scaled.Radius.of.Gyration	30.170	3	1.271e-06	***	
Skewness.About.Major.Axis	32.745	3	3.645e-07	***	
Skewness.About.Minor.Axis	17.526	3	0.0005508	***	
Kurtosis.About.Minor.Axis	12.294	3	0.0064404	**	
Kurtosis.About.Major.Axis	103.149	3	< 2.2e-16	***	
Hollows.Ratio	65.587	3	3.757e-14	***	

Variables with 2~3 stars look important and variables without star look 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.)

```
pred.class.1 <- predict(fit.log.nnet, newdata=data.train.scale,
                        type="class")
pred.class.2 <- predict(fit.log.nnet, newdata=data.valid.scale,
                        type="class")
(mul.misclass.train <- mean(ifelse(pred.class.1 == set1$class,
                                   yes=0, no=1)))
(mul.misclass.test <- mean(ifelse(pred.class.2 == set2$class,
                                   yes=0, no=1)))
```



```
> (mul.misclass.train <- mean(ifelse(pred.class.1 == set1$class,
+                                   yes=0, no=1)))
[1] 0.1529968
> (mul.misclass.test <- mean(ifelse(pred.class.2 == set2$class,
+                                   yes=0, no=1)))
[1] 0.2122642
```

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.**

```
### Next, let's investigate the LR's performance on the test set
pred.log.nnet = predict(fit.log.nnet, data.valid.scale)

table(Y.valid, pred.log.nnet,      ### Confusion matrix
      dnn = c("Observed", "Predicted"))

(misclass.log.nnet = mean(pred.log.nnet != Y.valid)) ### Misclass rate

> ### Next, let's investigate the LR's performance on the test set
> pred.log.nnet = predict(fit.log.nnet, data.valid.scale)
> table(Y.valid, pred.log.nnet,      ### Confusion matrix
+       dnn = c("Observed", "Predicted"))
      Predicted
Observed 2D 4D BUS VAN
      2D  34 19  1  1
      4D  18 33  1  2
      BUS  1  1 47  0
      VAN  1  0  0 53
> (misclass.log.nnet = mean(pred.log.nnet != Y.valid)) ### Misclass rate
[1] 0.2122642
```