Assignment #2: R code

Tristan De Alwis

4/16/2020

Loading Neccessary Libraries

```
suppressPackageStartupMessages(library(MASS))
suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library("Rlab"))
suppressPackageStartupMessages(library(class))
```

Q3

```
carsb = Cars93[, c(4, 5, 6, 7, 8, 12, 13, 14, 15, 17, 19:22, 25, 26)]
names(carsb)
```

```
##  [1] "Min.Price"          "Price"              "Max.Price"
##  [4] "MPG.city"           "MPG.highway"        "EngineSize"
##  [7] "Horsepower"         "RPM"                "Rev.per.mile"
## [10] "Fuel.tank.capacity" "Length"             "Wheelbase"
## [13] "Width"              "Turn.circle"        "Weight"
## [16] "Origin"
```

```
carsb[, -16] = log(carsb[, -16])
```

(a)

```
lrfp = function(m) {
    (m[1, 1]/(m[1, 1] + m[2, 1]))/(m[1, 2]/(m[2, 2] + m[1, 2]))
}
lrfn = function(m) {
    (m[2, 1]/(m[1, 1] + m[2, 1]))/(m[2, 2]/(m[2, 2] + m[1, 2]))
}
f0 = function(m) {
    c(1 - sum(diag(m))/sum(m), lrfp(m), lrfn(m))
}
```

(b)

```
fit.lda = lda(Origin ~ ., data = carsb)

confusion.matrix.lda = table(predict(fit.lda)$class, carsb$Origin)

m = confusion.matrix.lda

vector = f0(m)
cat("CE:  ", vector[1], "\n")
```

```
## CE:   0.1075269
```

```
cat("LR+: ", vector[2], "\n")
```

```
## LR+:  8.0625
```

```
cat("LR-: ", vector[3], "\n")
```

```
## LR-:  0.1171875
```

(c)

```
fit.qda = qda(Origin ~ ., data = carsb)

confusion.matrix.qda = table(predict(fit.qda)$class, carsb$Origin)

m = confusion.matrix.qda

vector = f0(m)
cat("CE:  ", vector[1], "\n")
```

```
## CE:   0.03225806
```

```
cat("LR+: ", vector[2], "\n")
```

```
## LR+:  43.125
```

```
cat("LR-: ", vector[3], "\n")
```

```
## LR-:  0.04261364
```

Yes, the QDA appears to produce less error

(d)

```
fit.lda = lda(Origin ~ ., data = carsb, CV = TRUE)

confusion.matrix.lda = table(fit.lda$class, carsb$Origin)

m = confusion.matrix.lda

vector = f0(m)
cat("CE:  ", vector[1], "\n")
```

```
## CE:   0.1397849
```

```
cat("LR+: ", vector[2], "\n")
```

```
## LR+:  6.40625
```

```
cat("LR-: ", vector[3], "\n\n")
```

```
## LR-:  0.1682692
```

```
fit.qda = qda(Origin ~ ., data = carsb, CV = TRUE)

confusion.matrix.qda = table(fit.qda$class, carsb$Origin)

m = confusion.matrix.qda
```

```r
vector = f0(m)
cat("CE:  ", vector[1], "\n")
```

```
## CE:   0.2473118
```

```r
cat("LR+: ", vector[2], "\n")
```

```
## LR+:  3.068182
```

```r
cat("LR-: ", vector[3], "\n")
```

```
## LR-:  0.3308824
```

lda() does better than qda() with CV because it is better at generalizing

Q4

(a)

(b)

(c)

(d)

```r
# norm_feat <- apply(pima1[-6], 2, function(x) { return((mean(x))/sd(x)) } )
# knn_mat <- knn_func(k.list, norm_feat, pima1[,6]) plot(seq(1, 125, 2),
# knn_mat[,1], pch=16, ylab='Classification Error', xlab = 'K', + main =
# 'Normalized Features') lines(seq(1, 125, 2), knn_mat[,1], lty=1) abline(h
# = 0.34, lty = 2, lwd = 3, col = 'dark blue')
```