STAT4610 - HW5

Michael Ghattas

10/2/2022

Question 7

```
library(ISLR2)
```

```
## Warning: package 'ISLR2' was built under R version 4.1.2
```

```
library(boot)
```

```
summary(Weekly)
```

```
##       Year           Lag1                Lag2                Lag3          
##  Min.   :1990   Min.   :-18.1950   Min.   :-18.1950   Min.   :-18.1950  
##  1st Qu.:1995   1st Qu.: -1.1540   1st Qu.: -1.1540   1st Qu.: -1.1580  
##  Median :2000   Median :  0.2410   Median :  0.2410   Median :  0.2410  
##  Mean   :2000   Mean   :  0.1506   Mean   :  0.1511   Mean   :  0.1472  
##  3rd Qu.:2005   3rd Qu.:  1.4050   3rd Qu.:  1.4090   3rd Qu.:  1.4090  
##  Max.   :2010   Max.   : 12.0260   Max.   : 12.0260   Max.   : 12.0260  
##       Lag4                Lag5              Volume            Today          
##  Min.   :-18.1950   Min.   :-18.1950   Min.   :0.08747   Min.   :-18.1950  
##  1st Qu.: -1.1580   1st Qu.: -1.1660   1st Qu.:0.33202   1st Qu.: -1.1540  
##  Median :  0.2380   Median :  0.2340   Median :1.00268   Median :  0.2410  
##  Mean   :  0.1458   Mean   :  0.1399   Mean   :1.57462   Mean   :  0.1499  
##  3rd Qu.:  1.4090   3rd Qu.:  1.4050   3rd Qu.:2.05373   3rd Qu.:  1.4050  
##  Max.   : 12.0260   Max.   : 12.0260   Max.   :9.32821   Max.   : 12.0260  
##  Direction 
##  Down:484  
##  Up  :605  
##            
##            
##            
##
```

part-(a)

```
logReg1 = glm(Direction ~ Lag1 + Lag2, data = Weekly, family = binomial)
summary(logReg1)
```

```
##
## Call:
## glm(formula = Direction ~ Lag1 + Lag2, family = binomial, data = Weekly)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.623  -1.261   1.001   1.083   1.506
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.22122    0.06147   3.599 0.000319 ***
## Lag1        -0.03872    0.02622  -1.477 0.139672
## Lag2         0.06025    0.02655   2.270 0.023232 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1496.2  on 1088  degrees of freedom
## Residual deviance: 1488.2  on 1086  degrees of freedom
## AIC: 1494.2
##
## Number of Fisher Scoring iterations: 4
```

part-(b)

```
set.seed(123)

logReg2 = glm(Direction ~ Lag1 + Lag2, data = Weekly[-1, ], family =
binomial)
summary(logReg2)
```

```
##
## Call:
```

```
## glm(formula = Direction ~ Lag1 + Lag2, family = binomial, data =
Weekly[-1,
##     ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6258  -1.2617   0.9999   1.0819   1.5071
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.22324    0.06150   3.630 0.000283 ***
## Lag1        -0.03843    0.02622  -1.466 0.142683
## Lag2         0.06085    0.02656   2.291 0.021971 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1494.6  on 1087  degrees of freedom
## Residual deviance: 1486.5  on 1085  degrees of freedom
## AIC: 1492.5
##
## Number of Fisher Scoring iterations: 4
```

part-(c)

```
predict.glm(logReg2, Weekly[1, ], type = "response")

##         1
## 0.5713923

predict.glm(logReg2, Weekly[1, ], type = "response") > 0.5

##    1
## TRUE

Weekly$Direction[1]

## [1] Down
## Levels: Down Up
```

-> The prediction was ~0.57, which is greater than 0.5, meaning the predicted direction is UP.

-> The actual direction was DOWN.

-> The prediction was inaccurate.

part-(d)

```
errorCount = rep(0, dim(Weekly)[1])
for (i in 1:(dim(Weekly)[1])) {
    logReg3 = glm(Direction ~ Lag1 + Lag2, data = Weekly[-i, ], family =
binomial)
    predUp = predict.glm(logReg3, Weekly[i, ], type = "response") > 0.5
    isUp = Weekly[i, ]$Direction == "Up"
    if (predUp != isUp)
        errorCount[i] = 1
}
sum(errorCount)
```

```
## [1] 490
```

-> The algorithm computed a sum of 490 errors.

part-(e)

```
mean(errorCount)
```

```
## [1] 0.4499541
```

-> The Leave-One-Out Cross-Validation (LOOCV) estimate for the test error is 45%.

End.