Assignment 5

Brian Morales

October 2, 2022

```r
library(readr)
library(MASS)
Weekly <- read_csv("~/Desktop/Fall-2022/Stats-Learning/ALL-CSV-FILES/Weekly.csv", show_col_types = FALSE
Weekly$Direction = as.factor(Weekly$Direction)
```

Part A

```r
set.seed(1)
weekly.fit = glm(Direction ~ Lag1 + Lag2, Weekly, family= 'binomial')
summary(weekly.fit)
```

```
##
## 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
```

We can see when all the coefficients are zero we have a 22.1% increase in weekly trades when there was no reports on the previous week and 2 weeks. However, we see the weekly market go down when the previous week is included and the market go up when the 2 previous weeks are included.

Part B

```r
 weekly.fit2 = glm(Direction ~ Lag1 + Lag2, Weekly[-1, ], family = 'binomial')
summary(weekly.fit2)
```

```
##
## 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
```

Excluding only the first observation, our results are pretty similar with a slight increase in the intercept coefficient, larger decrease in `Lag1`, and a smaller increase in `Lag2`.

Part C

```r
weekly.predict = predict(weekly.fit2, newdata = Weekly[1,], type = 'response') > 0.5
weekly.predict
```

```
##    1
## TRUE
```

```r
Weekly[1,]$Direction
```

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

Our prediction was classified incorrectly. Our model predicted `true` meaning the market was `UP` when de facto the market was `Down`.

Part D

```r
options(max.print = 1100)
n = nrow(Weekly)
err = rep(0, n)
for(i in 1:n) {
  # part i
  weekly.fit = glm(Direction ~ Lag1 + Lag2, Weekly[-i, ], family = 'binomial')
  # part ii
  posterior = predict(weekly.fit, newdata = Weekly[i, ], type = 'response') > 0.5
  # part iii
  actual_direction = Weekly[i,]$Direction == 'Up'
  if (posterior != actual_direction)
    # part iv
    err[i] = 1
}
err
```

```
##    [1] 1 1 0 1 0 1 0 0 0 1 1 0 1 0 1 0 1 0 1 0 0 0 1 1 1 1 1 1 0 1 1 1 1 0 1 0 0
##   [38] 0 1 0 1 0 0 1 0 1 1 1 0 1 0 0 0 1 0 0 1 1 0 0 0 0 1 0 1 1 0 0 1 0 1 1 0 0
##   [75] 0 1 0 1 1 0 0 1 1 0 1 1 0 0 1 0 0 1 1 1 0 0 0 0 0 1 0 1 1 0 0 1 0 1 0 0 1
##  [112] 1 0 0 1 0 0 1 0 0 1 1 1 1 0 0 0 1 0 1 0 1 1 0 0 0 1 1 1 0 0 0 1 0 0 0 0 0
##  [149] 0 1 1 1 0 1 0 0 1 1 0 1 0 0 1 1 0 0 1 0 0 1 0 0 1 1 1 0 1 0 1 0 0 0 0 0 0
##  [186] 0 0 1 1 0 1 0 1 0 1 0 1 0 0 1 0 0 1 0 0 1 0 1 0 1 1 1 0 0 1 1 0 1 0 0 1 1
##  [223] 0 0 0 1 1 1 0 1 0 1 0 1 0 0 0 1 1 0 1 0 1 0 1 0 1 0 1 1 0 1 0 0 1 0 0 1 0
##  [260] 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 1 0 0 1 0 1 1 0 0 0 0 0 1 0 1 0
##  [297] 0 1 0 0 0 1 0 0 1 1 0 0 1 0 0 0 0 1 0 1 1 0 0 1 0 1 0 1 1 0 0 0 1 0 1 0 0
##  [334] 1 1 1 1 0 1 0 0 1 0 0 0 1 0 1 0 1 0 0 0 0 0 1 1 0 0 1 0 0 1 0 0 0 1 1 0 1
##  [371] 1 1 1 1 0 0 0 1 0 0 0 0 0 0 1 0 1 1 0 0 1 1 0 0 0 0 0 1 0 0 1 1 1 0 1 0 1
##  [408] 0 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 0 1 0 0 1 0 1 0 0 0 0 0 1 1 1 1
##  [445] 0 1 1 0 1 0 1 1 0 1 0 0 1 0 0 1 1 0 0 0 0 1 1 0 0 1 0 1 0 0 0 1 0 0 1 0 0
##  [482] 0 1 1 1 0 1 0 0 0 1 0 1 1 1 0 0 0 0 1 1 1 0 1 1 0 1 0 0 0 1 0 1 0 0 0 1 0
##  [519] 1 1 0 0 1 1 0 0 0 1 1 0 1 0 1 1 1 1 1 0 0 0 1 0 0 0 1 1 0 1 0 0 0 1 1 1 1
##  [556] 1 1 0 1 0 1 0 0 1 0 0 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 0 0 1 0 1
##  [593] 0 0 1 0 0 1 0 1 1 0 1 1 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 1 0 1 1 0 1 0 1
##  [630] 0 1 1 1 1 0 1 1 0 0 0 1 1 1 1 0 1 1 1 0 1 0 0 0 1 1 1 1 1 1 0 1 0 0 1 0 0
##  [667] 0 1 1 0 1 0 1 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 1 1
##  [704] 0 0 0 0 1 0 1 0 1 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 1 1 0 1 1 0
##  [741] 1 1 1 1 0 0 0 1 1 1 1 1 1 0 1 0 0 0 0 0 0 1 0 1 1 1 0 0 0 1 0 0 1 0 0 0 1
##  [778] 1 1 0 0 0 1 0 0 1 1 1 0 0 1 0 1 0 1 0 0 1 0 0 1 0 0 0 0 0 1 0 1 1 0 0 1 1
##  [815] 0 1 1 1 0 0 0 0 0 1 1 0 0 1 0 0 1 0 1 0 0 0 1 1 0 1 1 0 1 0 1 0 1 1 0 0 1
##  [852] 1 1 0 1 1 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 1 0 0 1 1 0 0 1 0 1 0 1 1 0 1 0 1
##  [889] 1 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 1 0 0 0 1 1 1 1 1 0 1 1 0 0 0 0 0 1 0 0 1
##  [926] 0 0 0 0 1 0 1 1 1 0 0 1 1 0 1 1 1 1 0 1 0 1 0 1 0 1 0 1 0 0 1 1 1 1 1 0 1
##  [963] 0 0 0 1 1 1 0 1 1 1 1 0 0 0 0 1 1 0 0 0 0 1 0 0 1 1 1 0 0 1 1 1 0 1 0 0 0
## [1000] 0 1 0 0 1 0 1 0 0 1 1 1 1 0 1 0 0 1 0 0 1 0 0 1 1 0 1 1 1 0 1 1 0 0 0 1 0
## [1037] 1 0 1 1 1 1 0 0 1 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 0
## [1074] 0 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0
```

Part E

```r
mean(err)
```

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

Our average estimate for our LOOCV is 44.9%. This result is pretty poor because we want a test error the is less prone to variance.