

5.4 Question 7

Matthew Bradley, Ben Howell, Ayanna Fisher, Hayley Zorkic

3/13/2022

a:

```
library(ISLR2)
```

```
## Warning: package 'ISLR2' was built under R version 4.0.5
```

```
data("Weekly")
```

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

```
##
```

```
## 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
```

b:

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

```
##
```

```
## 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:	on	df	degrees of freedom
##	1494.6	1087		

	Residual deviance:	on	df	degrees of freedom
##	1486.5	1085		

```
## AIC: 1492.5
```

```
##
```

```
## Number of Fisher Scoring iterations: 4
```

c:

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

```
##      1
```

```
## 0.5713923
```

This observation was not correctly classified. The real value of “Direction” for this observation is “Down”, but our prediction was greater than 0.5, which corresponds to “Up”.

d:

```
totalError = 0
```

```
for(i in 1:nrow(Weekly)){
```

```
  currentLogistic <- glm(Direction ~ Lag1 + Lag2, data = Weekly[-i,], family = binomial)
```

```
  prediction <- predict.glm(currentLogistic, Weekly[i,], type = "response")
```

```
  if(prediction>0.5){
```

```
    market = "Up"
```

```
  }
```

```
  if(prediction < 0.5){
```

```
    market = "Down"
```

```
  }
```

```
  if(market != Weekly[i,9]){
```

```
    totalError = totalError + 1
```

```
  }
```

```
}
```

```
print(totalError)
```

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

e:

```
averageError = totalError/nrow(Weekly)
```

```
print(averageError)
```

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

The LOOCV is 44.995%, so our model is correct less than half of the time, which seems fairly unhelpful for predicting a binary variable.