HW6 Solutions

Dante Wu

Problem 1

(a) $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 1$, and $x_1 = 40$, $x_2 = 3.5$. Therefore, $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = -0.5$, and

$$\hat{p}(x) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2}} = \frac{e^{-0.5}}{1 + e^{-0.5}} = 0.378$$

.

(b) $\hat{p}(x) = 0.5$, $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 1$, and $x_2 = 3.5$. We have

$$0 = \log(p(x)/(1 - p(x))) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = 0.05 x_1 - 2.5.$$

Therefore $x_1 = 50$.

Problem 2

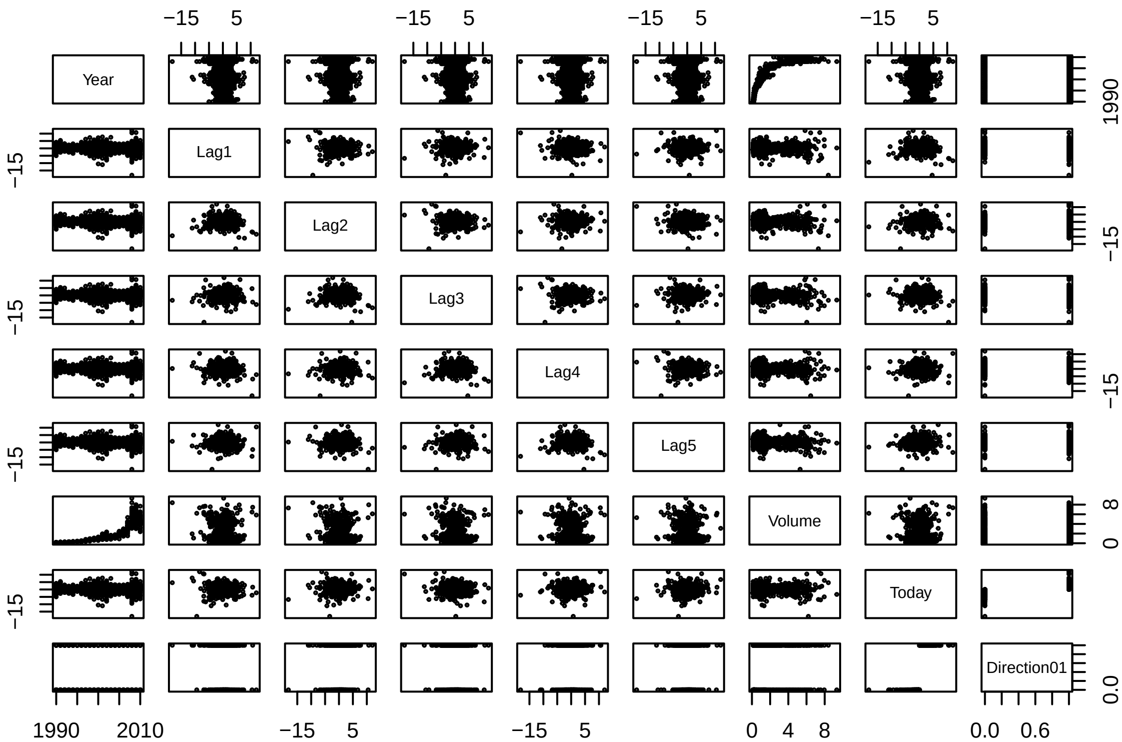
```
library(ISLR)
library(knitr)
head(Weekly)
```

```
##   Year   Lag1   Lag2   Lag3   Lag4   Lag5    Volume  Today Direction
## 1 1990  0.816  1.572 -3.936 -0.229 -3.484 0.1549760 -0.270      Down
## 2 1990 -0.270  0.816  1.572 -3.936 -0.229 0.1485740 -2.576      Down
## 3 1990 -2.576 -0.270  0.816  1.572 -3.936 0.1598375  3.514        Up
## 4 1990  3.514 -2.576 -0.270  0.816  1.572 0.1616300  0.712        Up
## 5 1990  0.712  3.514 -2.576 -0.270  0.816 0.1537280  1.178        Up
## 6 1990  1.178  0.712  3.514 -2.576 -0.270 0.1544440 -1.372      Down
```

```
knitr::kable(summary(Weekly))
```

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

```
Weekly$Direction01 = ifelse(Weekly$Direction == "Up", 1, 0)
pairs(Weekly[,-9], cex = 0.3)
```




```
cor(Weekly[, -9])
```

```
##                    Year         Lag1        Lag2        Lag3         Lag4
## Year         1.00000000 -0.032289274 -0.03339001 -0.03000649 -0.031127923
## Lag1        -0.03228927  1.000000000 -0.07485305  0.05863568 -0.071273876
## Lag2        -0.03339001 -0.074853051  1.00000000 -0.07572091  0.058381535
## Lag3        -0.03000649  0.058635682 -0.07572091  1.00000000 -0.075395865
## Lag4        -0.03112792 -0.071273876  0.05838153 -0.07539587  1.000000000
## Lag5        -0.03051910 -0.008183096 -0.07249948  0.06065717 -0.075675027
## Volume       0.84194162 -0.064951313 -0.08551314 -0.06928771 -0.061074617
## Today       -0.03245989 -0.075031842  0.05916672 -0.07124364 -0.007825873
## Direction01 -0.02220025 -0.050003804  0.07269634 -0.02291281 -0.020549456
##                    Lag5      Volume        Today Direction01
## Year        -0.030519101  0.84194162 -0.032459894 -0.02220025
## Lag1        -0.008183096 -0.06495131 -0.075031842 -0.05000380
## Lag2        -0.072499482 -0.08551314  0.059166717  0.07269634
## Lag3         0.060657175 -0.06928771 -0.071243639 -0.02291281
## Lag4        -0.075675027 -0.06107462 -0.007825873 -0.02054946
## Lag5         1.000000000 -0.05851741  0.011012698 -0.01816827
## Volume      -0.058517414  1.00000000 -0.033077783 -0.01799521
## Today        0.011012698 -0.03307778  1.000000000  0.72002470
## Direction01 -0.018168272 -0.01799521  0.720024704  1.00000000
```

We can see almost all variables are uncorrelated, except for volume and year, direction and today.

Do logistic regression

```
glm_fits = glm(Direction ~ Lag1 + Lag2 + Lag3 + Lag4 + Lag5 + Volume, data = Weekly, family = "binomial
summary(glm_fits)
```

```
##
## Call:
## glm(formula = Direction ~ Lag1 + Lag2 + Lag3 + Lag4 + Lag5 +
##     Volume, family = "binomial", data = Weekly)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6949  -1.2565   0.9913   1.0849   1.4579
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.26686    0.08593   3.106   0.0019 **
## Lag1        -0.04127    0.02641  -1.563   0.1181
## Lag2         0.05844    0.02686   2.175   0.0296 *
## Lag3        -0.01606    0.02666  -0.602   0.5469
## Lag4        -0.02779    0.02646  -1.050   0.2937
## Lag5        -0.01447    0.02638  -0.549   0.5833
## Volume      -0.02274    0.03690  -0.616   0.5377
## ---
## 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: 1486.4  on 1082  degrees of freedom
## AIC: 1500.4
##
## Number of Fisher Scoring iterations: 4
```

Confusion matrix

```r
Probs = predict(glm_fits, type='response')
contrasts(Weekly$Direction)
```

```
##      Up
## Down  0
## Up    1
```

```r
Pred_trend = ifelse(Probs>0.5, "Up", "Down")
table(Pred_trend, Weekly$Direction) # The row names are predicting labels and the column names are the
```

```
##
## Pred_trend Down  Up
##       Down   54  48
##       Up    430 557
```

```r
mean(Pred_trend == Weekly$Direction) #accuracy
```

```
## [1] 0.5610652
```

Therefore the accuracy is 56.1%. Since we define "Down" as "Positive", we have 48 false positives(type I errors) and 430 false negatives(type II errors).

Training and Prediction

```r
Data_train = Weekly[Weekly$Year <= 2008,]
Data_test = Weekly[Weekly$Year > 2008,]
glm_fits2 = glm(Direction ~ Lag2, data = Data_train, family = binomial)
```

```
Probs_2 = predict(glm_fits2, Data_test, type = "response")
Pred_trend2 = ifelse(Probs_2>0.5, "Up", "Down")
table(Pred_trend2, Data_test$Direction)
```

```
##
## Pred_trend2 Down Up
##        Down    9  5
##        Up     34 56
```

```
mean(Pred_trend2 == Data_test$Direction)
```

```
## [1] 0.625
```

Therefore the accuracy is 62.5%, and we have 5 type I errors and 34 type II errors.