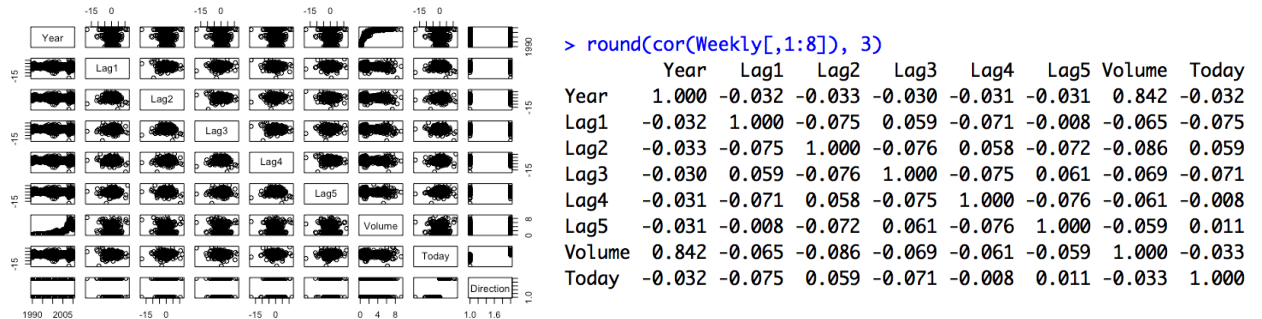


Statistical Learning and Data mining

Homework 6

M052040003 鍾冠毅

4.10. (a) 由下圖與相關係數矩陣得知，Year 與 Volume 有高度正相關。



(b) Lag2 較為顯著。

```
> summary(lmb)

Call:
glm(formula = Direction ~ Volume + Lag1 + Lag2 + Lag3 + Lag4 +
     Lag5, family = binomial)

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 **
Volume      -0.02274    0.03690  -0.616  0.5377
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
---
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
```

(c) test error rate is 0.4389.

```
> table(dir.prdct, Direction)
      Direction
dir.prdct Down Up
Down     54  48
Up      430 557
> (48+430)/1089
[1] 0.4389348
```

(d) test error rate is 0.375.

```
> table(dir.prdct, test$Direction)
      test$Direction
dir.prdct Down Up
Down      9   5
Up       34  56
> (5+34)/104
[1] 0.375
```

- (e) test error rate is 0.375.

```
> table(lda.e.prdct$class, test$Direction)

      Down Up
Down    9  5
Up     34 56
> (5+34)/104
[1] 0.375
```

- (f) test error rate is 0.4135.

```
> table(qda.f.prdct$class, test$Direction)

      Down Up
Down    0  0
Up     43 61
> (43+0)/104
[1] 0.4134615
```

- (g) test error rate is 0.5.

```
> table(knn.prdct, test$Direction)

knn.prdct Down Up
Down     21 30
Up       22 31
> (22+30)/104
[1] 0.5
```

- (h) The logistic regression and the LDA perform the best with the test error rate, 0.375.

- (i) 前三個模型改使用與Volume的交互作用項作為預測變數，並在KNN中討論K=10、100等情形，可得知(d)、(e)之結果依然優於其他預測模型。

```
> table(dir.prdct, test$Direction)

dir.prdct Down Up
Down      9  6
Up       34 55
> mean(dir.prdct != test$Direction)
[1] 0.3846154
> table(lda.i.prdct$class, test$Direction)

      Down Up
Down     8  6
Up      35 55
> mean(lda.i.prdct$class != test$Direction)
[1] 0.3942308
> table(qda.i.prdct$class, test$Direction)

      Down Up
Down    24 28
Up     19 33
> mean(qda.i.prdct$class != test$Direction)
[1] 0.4519231
> table(knn.prdct.i10, test$Direction)

knn.prdct.i10 Down Up
Down     17 18
Up       26 43
> mean(knn.prdct.i10 != test$Direction)
[1] 0.4230769
> table(knn.prdct.i100, test$Direction)

knn.prdct.i100 Down Up
Down      9 13
Up       34 48
> mean(knn.prdct.i100 != test$Direction)
[1] 0.4519231
```