

HW2 TJ Jiang

Load necessary libraries

```
library(ISLR)
```

Problem 1

Chapter 4, Exercise 4 (p. 168).

a)

For the cases where $0.5 < X < 0.95$, the average will be 10%. Otherwise, we form an integral as such:

$$\int_0^{0.05} 100x + 5 dx$$

Which equals 0.375, multiplied by 2 for two intervals: when $x < 0.05$ and when $x > 0.95$. Thus on average our prediction is $(0.1 * 0.9 + 0.00375 * 2) * 100 = 9.75\%$

b)

$$0.0975^2 * 100 = 0.95\%$$

c)

$$0.0975^{100} * 100 = (7.95 \text{ e-}100)\%$$

d)

Our results show that as dimensionality increases, the number of datapoints that are close in all dimensions to the response variable decreases exponentially.

e)

The length will be $0.1^{(1/p)}$ e.g. for $p=1$, $l = 0.1^{(1)} = 0.1$ for $p=100$, $l = 0.1^{(1/100)}$

Problem 2

Chapter 4, Exercise 6 (p. 170).

a)

$$\frac{e^{B_0+B_1X_1+\dots+B_pX_p}}{1+e^{B_0+B_1X_1+\dots+B_pX_p}} \frac{e^{-6+0.05*40+3.5}}{1+e^{-6+0.05*40+3.5}} = 0.3775$$

b)

$$0.5 = \frac{e^{-6+0.05X_1+3.5}}{1 + e^{-6+0.05X_1+3.5}} 0.5 * (1 + e^{0.05X_1-2.5}) = e^{0.05X_1-2.5} 0.5 = 0.5e^{0.05X_1-2.5} \log(1) = \log(e^{0.05X_1-2.5}) 0 = 0.05X_1 - 2.5$$

The student in part a) needs to study 50 hours to have a 50% chance of earning an A

Problem 3

Chapter 4, Exercise 8 (p. 170).

With K=1, the decision boundary is highly flexible. In general, as flexibility increases, the training error will decline but the testing error will increase. A KNN classifier with K=1 has a training error of 0, as every observation will simply cluster with itself. This means that the test error was 36%.

Thus, I would choose the 30% test error logistic regression classifier.

Problem 4

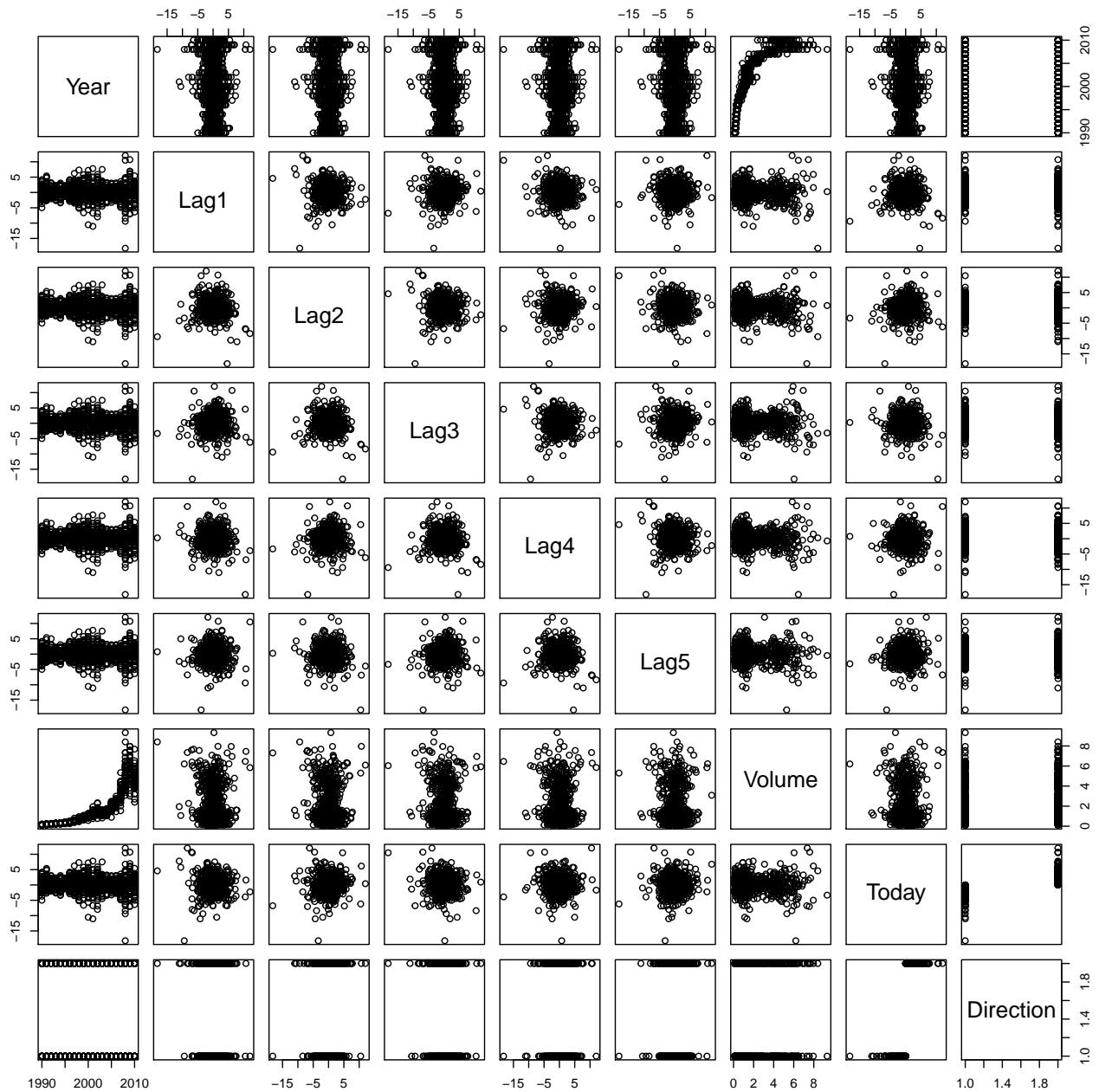
Updated: only a, b, c, d are required Chapter 4, Exercise 10 (p. 171). In part (i), please be concise; only describe and provide the output of your best prediction.

a)

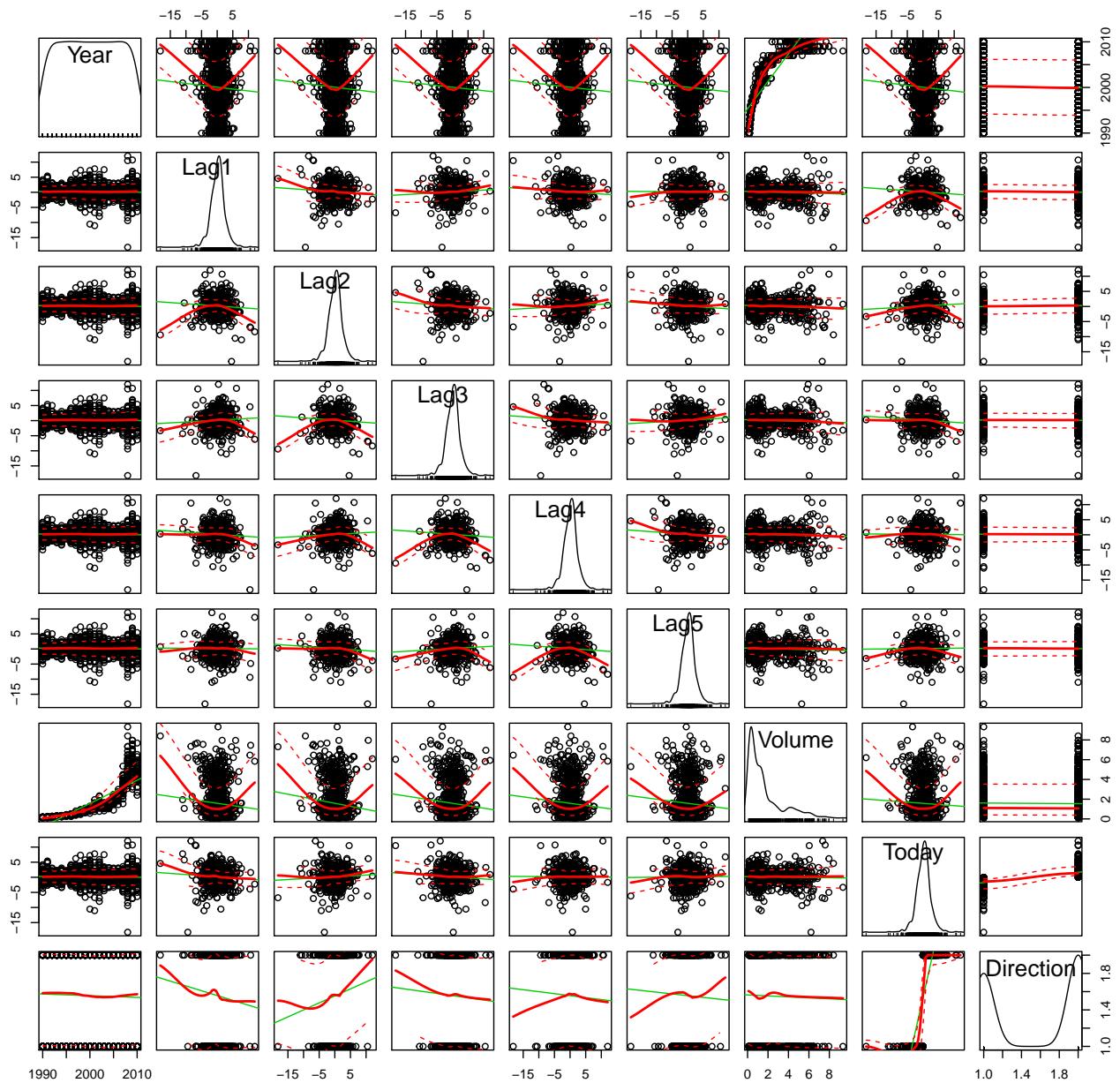
Quick look at correlation matrix

```
class(Weekly)
## [1] "data.frame"
plot(Weekly)
library(car)

## Warning: package 'car' was built under R version 3.2.5
```



Weekly Dataset Scatterplot Matrix



```
# chart.Correlation(Weekly, histogram=TRUE, pch=19)
```

Problem 5

Chapter 5, Exercise 5 (p. 198).

Problem 6

Chapter 5, Exercise 6 (p. 199)

Problem 7

Chapter 5, Exercise 8 (p. 200)

Problem 8

Chapter 5, Exercise 9 (p. 201)