HW1: Regression Modeling

Your Name

due on 10/11 (Tue)

Before working on HW1, you better complete the practice of ISLR ch2-3 R Labs. In particular, you may follow similar fitting steps demonstrated in ISLR ch3 R Lab for the following two data analysis.

Problem1

The first data consist of 4 variables (daily measurements) measured at the same place:

- ozone: response variable
- radiation
- temperature
- wind (speed)

Goal: Predict ozone using regression models.

Your analysis should include the following:

(a) Exploratory data analysis (EDA) among 4 variables
(b) Regression model fitting and model summaries.
(c) Model selection and diagonostics
(d) Comments on your prediction results and scientific findings. (state at least 3 viewpoints with data evidence)

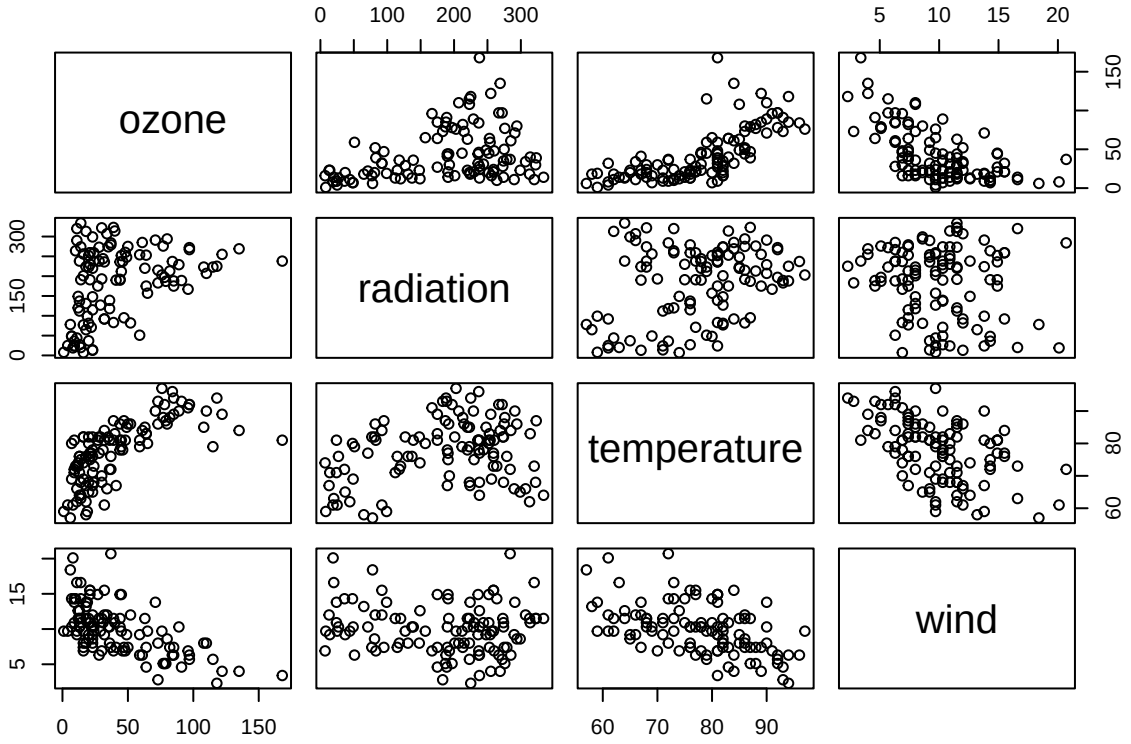
```
dat1 <- read.csv("ozone.csv")
head(dat1)
```

```
##   ozone radiation temperature wind
## 1    41       190          67  7.4
## 2    36       118          72  8.0
## 3    12       149          74 12.6
## 4    18       313          62 11.5
## 5    23       299          65  8.6
## 6    19        99          59 13.8
```

```
dim(dat1)
```

```
## [1] 111   4
```

```
pairs(dat1)
```



View data and start your analysis here...

Problem2

The second problem concerns about the Prostate data (more data descriptions can be found in ESL book chapter 3.2.1). There are 8 input variables and 1 response variable:

- Input variables (columns 1–8)
  - lcavol
  - lweight
  - age
  - lbph
  - svi
  - lcp
  - gleason
  - pgg45
- Response variable: lpsa (column 9)
- Indicator for training set (column 10): there are 97 observations in total, in which 70 obs'n are used as training data set and the rest of 27 observations are used as validation data. **Note: our training index is different from those used in the ESL book chapter 3.2.1.**

Goal: Predict lpsa using regression models.

Your analysis is based on the training data set only and should include the following:

(a) EDA
(b) Determine a good regression model for predicting lpsa
(c) Describe the important main effects and interaction effects.
(d) Predict lpsa for the validation data set based on the fitted model, with their prediction intervals. And compared the prediction results to the true observations. Comment on your model performance.

```
dat2 <- read.csv("Prostate.csv")
head(round(dat2,3))
```

```
##   lcavol lweight age   lbph svi    lcp gleason pgg45   lpsa train.idx
## 1 -0.580   2.769  50 -1.386   0 -1.386       6     0 -0.431         1
## 2 -0.994   3.320  58 -1.386   0 -1.386       6     0 -0.163         0
## 3 -0.511   2.691  74 -1.386   0 -1.386       7    20 -0.163         1
## 4 -1.204   3.283  58 -1.386   0 -1.386       6     0 -0.163         1
## 5  0.751   3.432  62 -1.386   0 -1.386       6     0  0.372         0
## 6 -1.050   3.229  50 -1.386   0 -1.386       6     0  0.765         0
```

```
dim(dat2)
```

```
## [1] 97 10
```