Chapter 13

DATA PREPROCESSING

資料集

除了一些內建的資料集

可使用 ” datasets” 這個套件取得更多 dataset

```
1 data (package ="datasets")
```

```
Data sets in package ¡¥datasets¡¦:

AirPassengers                    Monthly Airline Passenger Numbers 1949-1960
BJsales                          Sales Data with Leading Indicator
BJsales.lead (BJsales)           Sales Data with Leading Indicator
BOD                              Biochemical Oxygen Demand
CO2                              Carbon Dioxide Uptake in Grass Plants
ChickWeight                      Weight versus age of chicks on different diets
DNase                            Elisa assay of DNase

iris                             Edgar Anderson's Iris Data
iris3                            Edgar Anderson's Iris Data
islands                          Areas of the World's Major Landmasses
ldeaths (UKLungDeaths)           Monthly Deaths from Lung Diseases in the UK
lh                               Luteinizing Hormone in Blood Samples
longley                          Longley's Economic Regression Data
lynx                             Annual Canadian Lynx trappings 1821-1934
mdeaths (UKLungDeaths)           Monthly Deaths from Lung Diseases in the UK
morley                           Michelson Speed of Light Data
```

資料集

也可使用 ” .package(all.available =TRUE) 找到其他套件中的資料集

```
data(package =.packages(all.available=TRUE))
```

```
Data sets in package ¡¥boot¡¦:

acme                              Monthly Excess Returns
aids                              Delay in AIDS Reporting in England and Wales
aircondit                         Failures of Air-conditioning Equipment
aircondit7                        Failures of Air-conditioning Equipment
amis                              Car Speeding and Warning Signs
aml                               Remission Times for Acute Myelogenous Leukaemia
beaver                            Beaver Body Temperature Data

Data sets in package ¡¥cluster¡¦:

agriculture                       European Union Agricultural Workforces
animals                           Attributes of Animals
chorSub                           Subset of C-horizon of Kola Data
flower                            Flower Characteristics
plantTraits                       Plant Species Traits Data
pluton                            Isotopic Composition Plutonium Batches
ruspini                           Ruspini Data
votes.repub                       Votes for Republican Candidate in Presidential Elections
xclara                            Bivariate Data Set with 3 Clusters
```

資料集

使用 ?+ 資料集名稱 可以看到說明文件

```
?iris
```

R: Edgar Anderson's Iris Data ▾ Find in Topic

iris {datasets} R Documentation

Edgar Anderson's Iris Data

Description

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*.

Usage

```
iris
iris3
```

Format

`iris` is a data frame with 150 cases (rows) and 5 variables (columns) named `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`, and `Species`.

`iris3` gives the same data arranged as a 3-dimensional array of size 50 by 4 by 3, as represented by S-PLUS. The first dimension gives the case number within the species subsample, the second the measurements with names `Sepal L.`, `Sepal W.`, `Petal L.`, and `Petal W.`, and the third the species.

資料前處理

接下來我們要以 mice 套件中的 nhanes2 資料集來進行資料前置處理

1. 安裝並讀取以下四個套件
2. 得到 nhanes2 dataset 的值

```r
install.packages("lattice")
install.packages("MASS")
install.packages("nnet")
install.packages("mice")

library(lattice)
library(MASS)
library(nnet)
library(mice)

ds <- nhanes2

View(ds)
```

2.R × ds ×

Filter

	age	bmi	hyp	chl
1	20-39	NA	NA	NA
2	40-59	22.7	no	187
3	20-39	NA	no	187
4	60-99	NA	NA	NA
5	20-39	20.4	no	113
6	60-99	NA	NA	184
7	20-39	22.5	no	118
8	20-39	30.1	no	187
9	40-59	22.0	no	238
10	40-59	NA	NA	NA
11	20-39	NA	NA	NA
12	40-59	NA	NA	NA
13	60-99	21.7	no	206
14	40-59	28.7	yes	204
15	20-39	29.6	no	NA
16	20-39	NA	NA	NA
17	60-99	27.2	yes	284
18	40-59	26.3	yes	199
19	20-39	35.3	no	218
20	60-99	25.5	yes	NA
21	20-39	NA	NA	NA
22	20-39	33.2	no	229
23	20-39	27.5	no	131
24	60-99	24.9	no	NA
25	40-59	27.4	no	186

資料前處理

解析 nhanes2 資料集

```r
summary(ds)
```

```
    age          bmi          hyp          chl       
 20-39:12   Min.   :20.40   no  :13   Min.   :113.0  
 40-59: 7   1st Qu.:22.65   yes : 4   1st Qu.:185.0  
 60-99: 6   Median :26.75   NA's: 8   Median :187.0  
            Mean   :26.56             Mean   :191.4  
            3rd Qu.:28.93             3rd Qu.:212.0  
            Max.   :35.30             Max.   :284.0  
            NA's   :9                 NA's   :10     
> 
```

Factor	meaning
age	年齡
bmi	BMI 指數
hyp	高血壓 0/1
chl	膽固醇 mg/dL

```r
head(ds)
```

```
> head(ds)
    age  bmi  hyp chl
1 20-39   NA <NA>  NA
2 40-59 22.7   no 187
3 20-39   NA   no 187
4 60-99   NA <NA>  NA
5 20-39 20.4   no 113
6 60-99   NA <NA> 184
> 
```

資料前處理

遺漏值處理

1. 檢測資料缺失的情況

```r
library(lattice)
library(MASS)
library(nnet)
library(mice)

ds <- nhanes2

lose <- md.pattern(nhanes2)
print(lose)
```

```
   age hyp bmi chl
13   1   1   1   1  0
1    1   1   0   1  1
3    1   1   1   0  1
1    1   0   0   1  2
7    1   0   0   0  3
     0   8   9  10 27
```

2.R × | ds ×

Filter

	age	bmi	hyp	chl
1	20-39	NA	NA	NA
2	40-59	22.7	no	187
3	20-39	NA	no	187
4	60-99	NA	NA	NA
5	20-39	20.4	no	113
6	60-99	NA	NA	184
7	20-39	22.5	no	118
8	20-39	30.1	no	187
9	40-59	22.0	no	238
10	40-59	NA	NA	NA
11	20-39	NA	NA	NA
12	40-59	NA	NA	NA
13	60-99	21.7	no	206
14	40-59	28.7	yes	204
15	20-39	29.6	no	NA
16	20-39	NA	NA	NA
17	60-99	27.2	yes	284
18	40-59	26.3	yes	199
19	20-39	35.3	no	218
20	60-99	25.5	yes	NA
21	20-39	NA	NA	NA
22	20-39	33.2	no	229
23	20-39	27.5	no	131
24	60-99	24.9	no	NA
25	40-59	27.4	no	186

資料前處理

遺漏值處理

1. 檢測資料缺失的情況

```
library(lattice)
library(MASS)
library(nnet)
library(mice)

ds <- nhanes2

lose <- md.pattern(nhanes2)
print(lose)
```

```
   age hyp bmi chl
13   1   1   1   1  0
1    1   1   0   1  1
3    1   1   1   0  1
1    1   0   0   1  2
7    1   0   0   0  3
     0   8   9  10 27
```

2.R × | ds ×

Filter

	age	bmi	hyp	chl
1	20-39	NA	NA	NA
2	40-59	22.7	no	187
3	20-39	NA	no	187
4	60-99	NA	NA	NA
5	20-39	20.4	no	113
6	60-99	NA	NA	184
7	20-39	22.5	no	118
8	20-39	30.1	no	187
9	40-59	22.0	no	238
10	40-59	NA	NA	NA
11	20-39	NA	NA	NA
12	40-59	NA	NA	NA
13	60-99	21.7	no	206
14	40-59	28.7	yes	204
15	20-39	29.6	no	NA
16	20-39	NA	NA	NA
17	60-99	27.2	yes	284
18	40-59	26.3	yes	199
19	20-39	35.3	no	218
20	60-99	25.5	yes	NA
21	20-39	NA	NA	NA
22	20-39	33.2	no	229
23	20-39	27.5	no	131
24	60-99	24.9	no	NA
25	40-59	27.4	no	186

資料前處理

遺漏值處理

1. 檢測資料缺失的情況

```r
library(lattice)
library(MASS)
library(nnet)
library(mice)

ds <- nhanes2

lose <- md.pattern(nhanes2)
print(lose)
```

```
   age hyp bmi chl
13   1   1   1   1  0
 1   1   1   0   1  1
 3   1   1   1   0  1
 1   1   0   0   1  2
 7   1   0   0   0  3
     0   8   9  10 27
```

2.R × | ds ×

Filter

	age	bmi	hyp	chl
1	20-39	NA	NA	NA
2	40-59	22.7	no	187
3	20-39	NA	no	187
4	60-99	NA	NA	NA
5	20-39	20.4	no	113
6	60-99	NA	NA	184
7	20-39	22.5	no	118
8	20-39	30.1	no	187
9	40-59	22.0	no	238
10	40-59	NA	NA	NA
11	20-39	NA	NA	NA
12	40-59	NA	NA	NA
13	60-99	21.7	no	206
14	40-59	28.7	yes	204
15	20-39	29.6	no	NA
16	20-39	NA	NA	NA
17	60-99	27.2	yes	284
18	40-59	26.3	yes	199
19	20-39	35.3	no	218
20	60-99	25.5	yes	NA
21	20-39	NA	NA	NA
22	20-39	33.2	no	229
23	20-39	27.5	no	131
24	60-99	24.9	no	NA
25	40-59	27.4	no	186

資料前處理

遺漏值處理

2-1. 刪除法 : remove 去有缺值的部份

```r
library(lattice)
library(MASS)
library(nnet)
library(mice)

ds <- nhanes2

test <- complete.cases(ds)
print(test)

rm <- ds[complete.cases(ds), ]
View(rm)
```

	age	bmi	hyp	chl
2	40-59	22.7	no	187
5	20-39	20.4	no	113
7	20-39	22.5	no	118
8	20-39	30.1	no	187
9	40-59	22.0	no	238
13	60-99	21.7	no	206
14	40-59	28.7	yes	204
17	60-99	27.2	yes	284
18	40-59	26.3	yes	199
19	20-39	35.3	no	218
22	20-39	33.2	no	229
23	20-39	27.5	no	131
25	40-59	27.4	no	186

```
> test <- complete.cases(ds)
> print(test)
 [1] FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE
[19]  TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE
```

資料前處理

遺漏值處理

2-2. 插補法：利用統計方法填補缺失值 (ex. 平均值)

Step 1. 算出所有非 NA 的 bmi 平均值

Step 2. 寫入 BMI 有缺少的欄位)

```r
library(lattice)
library(MASS)
library(nnet)
library(mice)

ds <- nhanes2
head(ds)
mean    <- mean(ds[, 2],
                na.rm = T)
mean

ds[is.na(ds[, 2]), 2] <- mean
head(ds)
```

```
> ds <- nhanes2
> head(ds)
    age  bmi  hyp chl
1 20-39   NA <NA>  NA
2 40-59 22.7   no 187
3 20-39   NA   no 187
4 60-99   NA <NA>  NA
5 20-39 20.4   no 113
6 60-99   NA <NA> 184
```

```
> mean    <- mean(ds[, 2],
+                 na.rm = T)
> mean
[1] 26.5625
```

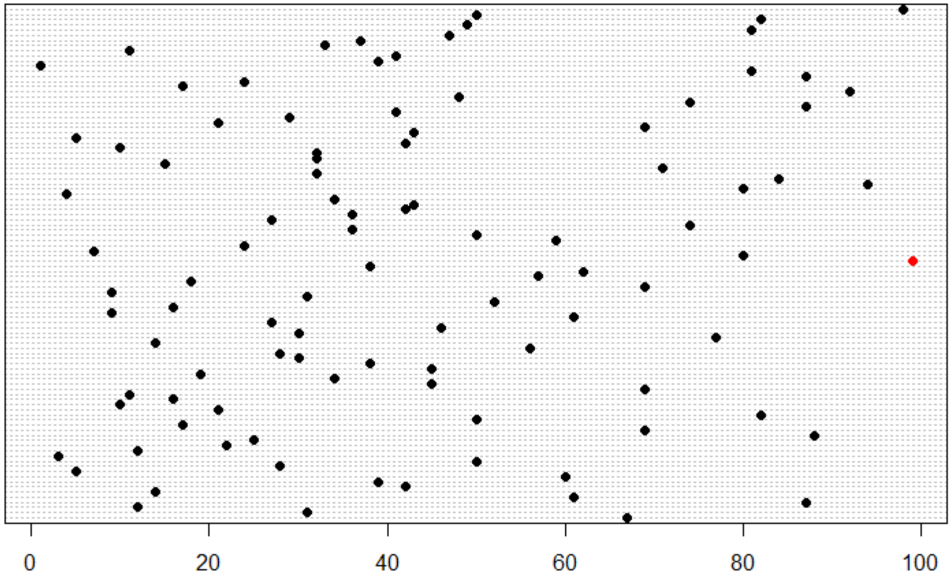
```
> ds[is.na(ds[, 2]), 2] <- mean
> head(ds)
    age     bmi  hyp chl
1 20-39 26.5625 <NA>  NA
2 40-59 22.7000   no 187
3 20-39 26.5625   no 187
4 60-99 26.5625 <NA>  NA
5 20-39 20.4000   no 113
6 60-99 26.5625 <NA> 184
```

資料前處理

雜訊資料處理

1. 找出資料中的例外點

```
rm(list=ls())
library(lattice)
library(MASS)
library(nnet)
library(mice)
library(DMwR)
library(outliers)

y=round(runif(100)*100)
print(y)
dotchart(y,col=ifelse(y==outlier(y),
                     "red","black"),pch=19)
```

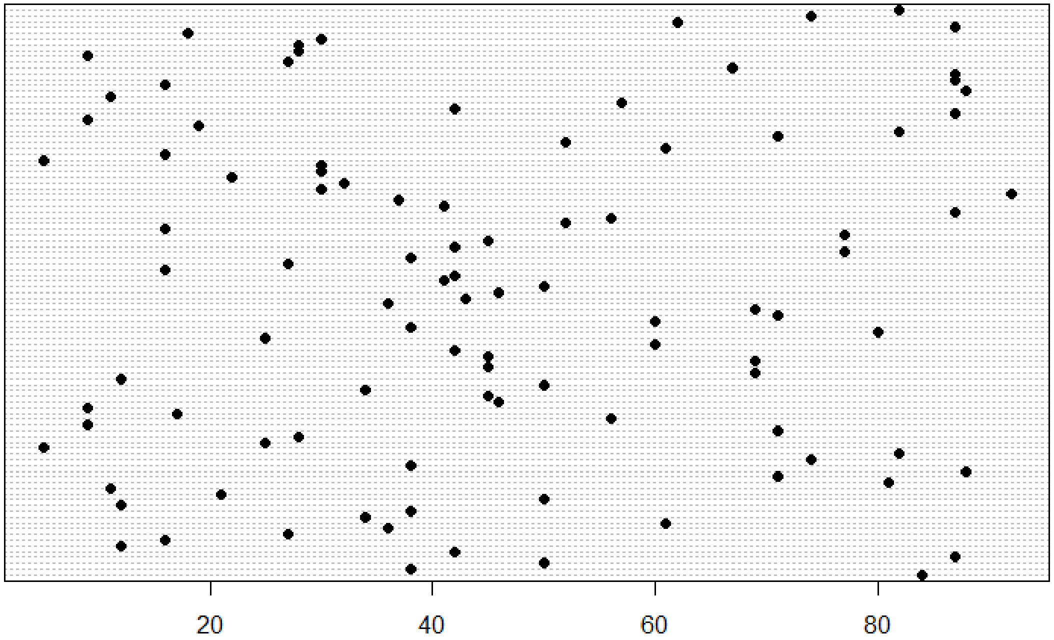


資料前處理

雜訊資料處理

2-1. 刪除法 : remove 例外點

```r
rm(list=ls())
library(lattice)
library(MASS)
library(nnet)
library(mice)
library(DMwR)
library(outliers)

y=round(runif(100)*100)
print(y)
dotchart(y,col=ifelse(y==outlier(y),
                      "red","black"),pch=19)
rmd <- outlier(y)
yy  <- y
yy  <- yy[yy[-rmd]]
print(yy)
dotchart(yy,col=ifelse(yy==outlier(y),
                       "red","black"),pch=19)
```



資料前處理

雜訊資料處理

2-2. 等寬箱平均值光滑方法

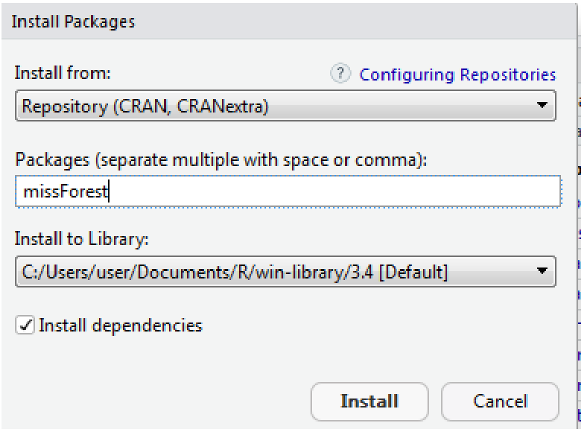
```r
rm(list=ls())
library(lattice)
library(MASS)
library(nnet)
library(mice)
library(DMwR)
library(outliers)

y=round(runif(100)*100)
print(y)
dotchart(y,col=ifelse(y==outlier(y),
                      "red","black"),pch=19)
y <- sort(y)
dim(y)=c(10,10)
print(y)
y[1,] = apply(y,1,mean)[1]
y[2,] = apply(y,1,mean)[2]
y[3,] = apply(y,1,mean)[3]
y[4,] = apply(y,1,mean)[4]
y[5,] = apply(y,1,mean)[5]
y[6,] = apply(y,1,mean)[6]
y[7,] = apply(y,1,mean)[7]
y[8,] = apply(y,1,mean)[8]
y[9,] = apply(y,1,mean)[9]
y[10,] = apply(y,1,mean)[10]
print(y)
```

```
> y <- sort(y)
> dim(y)=c(10,10)
> print(y)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]    3    8   16   31   41   52   64   77   81    90
 [2,]    3    9   18   31   42   52   65   77   83    92
 [3,]    3   10   21   32   42   54   66   79   84    95
 [4,]    4   10   23   36   46   54   68   79   86    95
 [5,]    4   12   25   38   48   55   69   79   86    97
 [6,]    5   13   26   38   48   56   69   79   86    98
 [7,]    6   15   26   39   49   57   69   79   87    98
 [8,]    6   15   27   39   50   58   70   80   88    98
 [9,]    7   15   28   39   50   61   72   80   90    98
[10,]    8   16   29   40   51   62   74   80   90    99
```

```
> y[1,] = apply(y,1,mean)[1]
> y[2,] = apply(y,1,mean)[2]
> y[3,] = apply(y,1,mean)[3]
> y[4,] = apply(y,1,mean)[4]
> y[5,] = apply(y,1,mean)[5]
> y[6,] = apply(y,1,mean)[6]
> y[7,] = apply(y,1,mean)[7]
> y[8,] = apply(y,1,mean)[8]
> y[9,] = apply(y,1,mean)[9]
> y[10,] = apply(y,1,mean)[10]
> print(y)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,] 46.3 46.3 46.3 46.3 46.3 46.3 46.3 46.3 46.3  46.3
 [2,] 47.2 47.2 47.2 47.2 47.2 47.2 47.2 47.2 47.2  47.2
 [3,] 48.6 48.6 48.6 48.6 48.6 48.6 48.6 48.6 48.6  48.6
 [4,] 50.1 50.1 50.1 50.1 50.1 50.1 50.1 50.1 50.1  50.1
 [5,] 51.3 51.3 51.3 51.3 51.3 51.3 51.3 51.3 51.3  51.3
 [6,] 51.8 51.8 51.8 51.8 51.8 51.8 51.8 51.8 51.8  51.8
 [7,] 52.5 52.5 52.5 52.5 52.5 52.5 52.5 52.5 52.5  52.5
 [8,] 53.1 53.1 53.1 53.1 53.1 53.1 53.1 53.1 53.1  53.1
 [9,] 54.0 54.0 54.0 54.0 54.0 54.0 54.0 54.0 54.0  54.0
[10,] 54.9 54.9 54.9 54.9 54.9 54.9 54.9 54.9 54.9  54.9
```

如何產生缺失資料

使用 prodNA 函數產生 15% IRIS data 的 miss value




```r
rm(list=ls())
library(missForest)
library(DMwR)


iris_data <- prodNA(iris, noNA = 0.15)
head(iris_data)
```

```
> iris_data <- prodNA(iris, noNA = 0.15)
> head(iris_data)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0           NA          NA  setosa
3          4.7         3.2          1.3          NA  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0          NA           NA         0.2    <NA>
6          5.4         3.9          1.7         0.4  setosa
>
```

隨堂練習 1

1. 利用平均方法來填補所有數值的 missing value

2. 利用最近鄰居方法 (KNN) 來填補所有的 missing value

   hint: 需要安裝 DMwR 套件

資料前處理

資料整合

資料整合是將多個來源中的資料合併並儲存，整合前要先進行比對的動作 (ex t-test, 卡方檢定 chisq.test)

```r
d1 <- round(runif(10)*100)
d2 <- round(runif(10)*100)

d  <- cbind(d1,d2)
print(d)

test <- chisq.test(d1)
print(test)

test2 <- t.test(d1,d2)
print(test2)
```

```
> print(d)
      d1 d2
 [1,] 24 29
 [2,] 68 20
 [3,]  8 57
 [4,] 68 20
 [5,] 14 51
 [6,] 15 92
 [7,] 61 15
 [8,] 65 93
 [9,] 90 85
[10,] 12 86
>
> test <- chisq.test(d1)
> print(test)

	Chi-squared test for given probabilities

data:  d1
X-squared = 198.51, df = 9, p-value < 2.2e-16

>
> test2 <- t.test(d1,d2)
> print(test2)

	Welch Two Sample t-test

data:  d1 and d2
t = -0.87324, df = 17.946, p-value = 0.3941
```

資料前處理

資料整合

資料整合是將多個來源中的資料合併並儲存，整合前要先進行比對的動作 (ex t-test, 卡方檢定 chisq.test)

```r
d1 <- seq(1:10)
d2 <- seq(1:10)
d  <- cbind(d1,d2)
print(d)
test <- chisq.test(d)
print(test)

test2 <- t.test(d1,d2)
print(test2)

```

```
> print(d)
      d1 d2
 [1,]  1  1
 [2,]  2  2
 [3,]  3  3
 [4,]  4  4
 [5,]  5  5
 [6,]  6  6
 [7,]  7  7
 [8,]  8  8
 [9,]  9  9
[10,] 10 10
> test <- chisq.test(d)
Warning message:
In chisq.test(d) : Chi-squared approximation may be incorrect
> print(test)

        Pearson's Chi-squared test

data:  d
X-squared = 0, df = 9, p-value = 1

>
> test2 <- t.test(d1,d2)
> print(test2)

        Welch Two Sample t-test

data:  d1 and d2
t = 0, df = 18, p-value = 1
```

資料前處理

資料整合

```r
set.seed(50)
a1 <- round(runif(3)*5)
a2 <- round(runif(3)*5)
b1 <- runif(3)*1
b2 <- runif(3)*1
a <- c(a1,a2)
b <- c(b1,b2)
print(a)

dt <- cbind(a,b)
print(dt)

rept <-unique(dt[,1])
lens <-length(rept)
print(rept)

for(i in 1:lens)
{
  if(length(which(dt[,1]==rept[i]))>=2)
  {
    x<-which(dt[,1]==rept[i])
    break
  }
}

dt <- dt[-x[2],]
print(dt)
```

```
> print(a)
[1] 4 2 1 4 3 0
>
> dt <- cbind(a,b)
> print(dt)
     a          b
[1,] 4 0.69991279
[2,] 2 0.64634362
[3,] 1 0.04202962
[4,] 4 0.10754123
[5,] 3 0.39055818
[6,] 0 0.26976581
>
> rept <-unique(dt[,1])
> lens <-length(rept)
> print(rept)
[1] 4 2 1 3 0
```

```
> for(i in 1:lens)
+ {
+   if(length(which(dt[,1]==rept[i]))>=2)
+   {
+     x<-which(dt[,1]==rept[i])
+     break
+   }
+ }
>
> dt <- dt[-x[2],]
> print(dt)
     a          b
[1,] 4 0.69991279
[2,] 2 0.64634362
[3,] 1 0.04202962
[4,] 3 0.39055818
[5,] 0 0.26976581
>
```

資料前處理

資料轉換

資料整合是將資料轉換成接下來要探勘的樣式 (如之前介紹的光滑方法)

現在用將 uci 資料集 (wine) 的正規化來當成另一個例子

https://archive.ics.uci.edu/ml/datasets/wine

UCI

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Wine Data Set

Download: Data Folder, Data Set Description

Abstract: Using chemical analysis determine the origin of wines

Data Set Characteristics:	Multivariate	**Number of Instances:**	178	**Area:**	Physical
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	13	**Date Donated**	1991-07-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	850710

資料前處理

資料轉換

現在用將 uci 資料集 (wine) 的正規化來當成另一個例子

https://archive.ics.uci.edu/ml/datasets/wine

```r
wine_uci <- "https://archive.ics.uci.edu/ml
/machine-learning-databases/wine/wine.data"

wine_data <- read.table(wine_uci, header=TRUE, sep =",")

head(wine_data,10)

```

```
> head(wine_data,10)
   X1 X14.23 X1.71 X2.43 X15.6 X127 X2.8 X3.06 X.28 X2.29 X5.64 X1.04 X3.92 X1065
1   1  13.20  1.78  2.14  11.2  100 2.65  2.76 0.26  1.28  4.38  1.05  3.40  1050
2   1  13.16  2.36  2.67  18.6  101 2.80  3.24 0.30  2.81  5.68  1.03  3.17  1185
3   1  14.37  1.95  2.50  16.8  113 3.85  3.49 0.24  2.18  7.80  0.86  3.45  1480
4   1  13.24  2.59  2.87  21.0  118 2.80  2.69 0.39  1.82  4.32  1.04  2.93   735
5   1  14.20  1.76  2.45  15.2  112 3.27  3.39 0.34  1.97  6.75  1.05  2.85  1450
6   1  14.39  1.87  2.45  14.6   96 2.50  2.52 0.30  1.98  5.25  1.02  3.58  1290
7   1  14.06  2.15  2.61  17.6  121 2.60  2.51 0.31  1.25  5.05  1.06  3.58  1295
8   1  14.83  1.64  2.17  14.0   97 2.80  2.98 0.29  1.98  5.20  1.08  2.85  1045
9   1  13.86  1.35  2.27  16.0   98 2.98  3.15 0.22  1.85  7.22  1.01  3.55  1045
10  1  14.10  2.16  2.30  18.0  105 2.95  3.32 0.22  2.38  5.75  1.25  3.17  1510
```

資料前處理

資料轉換

正規化

```r
rm(list=ls())
nor <- function(x){(x-min(x))/(max(x)-min(x))}

wine_uci <- "https://archive.ics.uci.edu/ml
/machine-learning-databases/wine/wine.data"

wine_data <- read.table(wine_uci, header=TRUE, sep =",")

print(wine_data[,14])
wine_data[,14] <- nor(wine_data[,14])

head(wine_data,10)
```

```
> print(wine_data[,14])
 [1] 1050 1185 1480  735 1450 1290
[22] 1035 1015  845  830 1195 1285
[43]  680  885 1080 1065  985 1060
```

```
> head(wine_data,10)
  X1 X14.23 X1.71 X2.43 X15.6 X127 X2.8 X3.06 X.28 X2.29 X5.64 X1.04 X3.92     X1065
1  1  13.20  1.78  2.14  11.2  100 2.65  2.76 0.26  1.28  4.38  1.05  3.40 0.5506419
2  1  13.16  2.36  2.67  18.6  101 2.80  3.24 0.30  2.81  5.68  1.03  3.17 0.6469330
3  1  14.37  1.95  2.50  16.8  113 3.85  3.49 0.24  2.18  7.80  0.86  3.45 0.8573466
```

隨堂練習 2

1. 讀取 IRIS data, Wine data

2. 將 IRIS 的資料集和 Wine data 合併

   Hint: 將 IRIS 增加維度並補足 NA 的資料後再合併

Any Questions !?