DATA621 Extended LMR Ex11.3

Chun Yip

2022/5/2

R Markdown

The dataset prostate is from a study of 97 men with prostate cancer who were due to receive a radical prostatectomy. Predict the 1 weight using the age. How do the methods deal with the outlier?

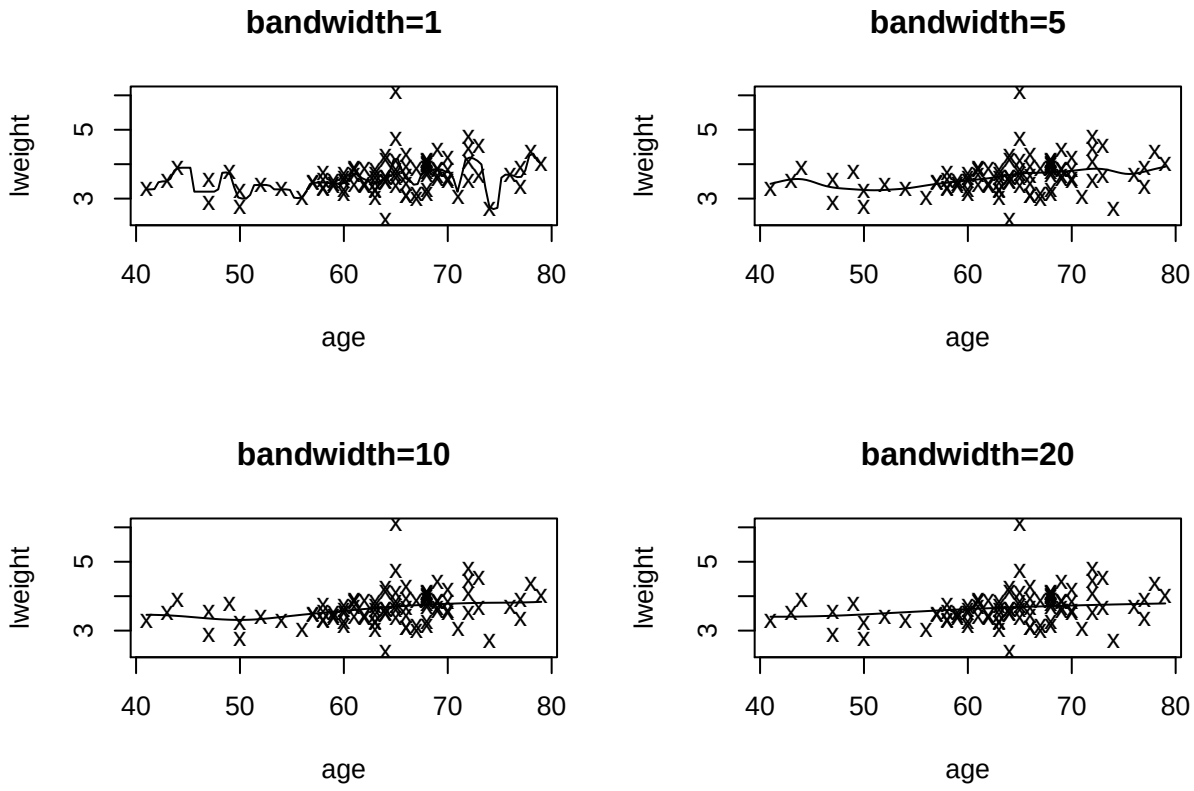
```
data(prostate, package="faraway")
head(prostate)
```

```
##       lcavol lweight age      lbph svi      lcp gleason pgg45     lpsa
## 1 -0.5798185  2.7695  50 -1.386294   0 -1.38629       6     0 -0.43078
## 2 -0.9942523  3.3196  58 -1.386294   0 -1.38629       6     0 -0.16252
## 3 -0.5108256  2.6912  74 -1.386294   0 -1.38629       7    20 -0.16252
## 4 -1.2039728  3.2828  58 -1.386294   0 -1.38629       6     0 -0.16252
## 5  0.7514161  3.4324  62 -1.386294   0 -1.38629       6     0  0.37156
## 6 -1.0498221  3.2288  50 -1.386294   0 -1.38629       6     0  0.76547
```

```
summary(prostate)
```

```
##      lcavol           lweight           age             lbph
##  Min.   :-1.3471   Min.   :2.375   Min.   :41.00   Min.   :-1.3863
##  1st Qu.: 0.5128   1st Qu.:3.376   1st Qu.:60.00   1st Qu.:-1.3863
##  Median : 1.4469   Median :3.623   Median :65.00   Median : 0.3001
##  Mean   : 1.3500   Mean   :3.653   Mean   :63.87   Mean   : 0.1004
##  3rd Qu.: 2.1270   3rd Qu.:3.878   3rd Qu.:68.00   3rd Qu.: 1.5581
##  Max.   : 3.8210   Max.   :6.108   Max.   :79.00   Max.   : 2.3263
##       svi              lcp             gleason          pgg45
##  Min.   :0.0000   Min.   :-1.3863   Min.   :6.000   Min.   :  0.00
##  1st Qu.:0.0000   1st Qu.:-1.3863   1st Qu.:6.000   1st Qu.:  0.00
##  Median :0.0000   Median :-0.7985   Median :7.000   Median : 15.00
##  Mean   :0.2165   Mean   :-0.1794   Mean   :6.753   Mean   : 24.38
##  3rd Qu.:0.0000   3rd Qu.: 1.1786   3rd Qu.:7.000   3rd Qu.: 40.00
##  Max.   :1.0000   Max.   : 2.9042   Max.   :9.000   Max.   :100.00
##       lpsa
##  Min.   :-0.4308
##  1st Qu.: 1.7317
##  Median : 2.5915
##  Mean   : 2.4784
##  3rd Qu.: 3.0564
##  Max.   : 5.5829
```

```
par(mfrow=c(2,2))
plot(lweight~age, prostate, main="bandwidth=1", pch="x")
lines(ksmooth(prostate$age, prostate$lweight, kernel = "normal", bandwidth = 1))
plot(lweight~age, prostate, main="bandwidth=5", pch="x")
lines(ksmooth(prostate$age, prostate$lweight, kernel = "normal", bandwidth = 5))
plot(lweight~age, prostate, main="bandwidth=10", pch="x")
lines(ksmooth(prostate$age, prostate$lweight, kernel = "normal", bandwidth = 10))
plot(lweight~age, prostate, main="bandwidth=20", pch="x")
lines(ksmooth(prostate$age, prostate$lweight, kernel = "normal", bandwidth = 20))
```



Bandwidth equals to 10 is a relatively good fit. It basically ignore outliner.