HW5 - Lucas Fellmeth, Sven Bergmann

2023-11-16

The goal of this problem is to estimate the regression function of acceleration vs time for the `mcycle` data in the package `MASS`.

- Show that the Nadaraya-Watson estimator can be expressed as $\hat{Y} = HY$. Find the "hat matrix" H explicitly.
- For a reasonable range of bandwidths h, compute and plot the generalized cross validation measure $GCV(h)$ and find the optimal bandwidth.

```
require(MASS)
```

```
## Loading required package: MASS
```

```
library(MASS)
require(splines)
```

```
## Loading required package: splines
```

```
library(splines)
```

First, we implement the generalized cross validation measure $GCV(h)$ which is defined by

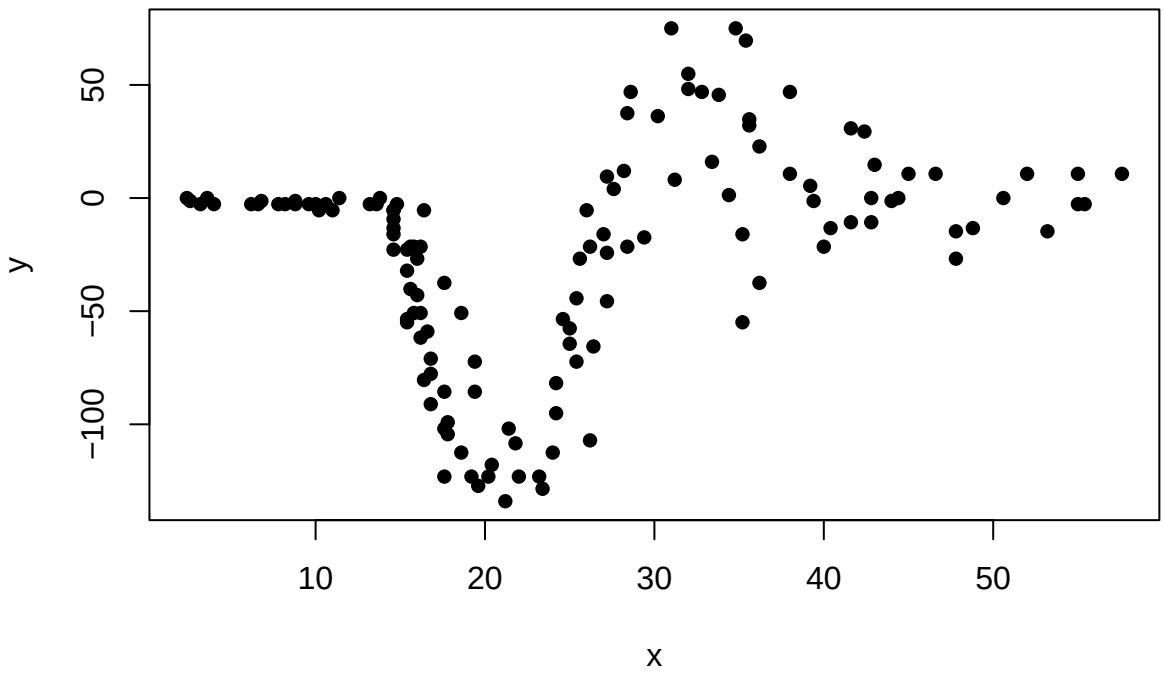
$$GCV(h) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Y_i - \hat{m}_h(x)}{1 - \frac{trS(h)}{n}}\right]^2$$

where

```
GCV <- function(h) {

}
```

```
x <- mcycle$times
y <- mcycle$accel
plot(x, y, pch = 16)
```


y
50
0
−50
−100
10
20
30
40
50
x