

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.

We know that the Nadaraya-Watson estimator of $\hat{m}(x)$ is defined by

$$\hat{m}(x) = \frac{\sum_{i=1}^n K_h(X_i - x)Y_i}{\sum_{i=1}^n K_h(X_i - x)},$$

where

$$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$$

with h as associated bandwidth.

- For a reasonable range of bandwidths h , compute and plot the generalized cross validation measure $GCV(h)$ and find the optimal bandwidth.

```
library(MASS)
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)
```

