STAT40810 — Stochastic Models

Brendan Murphy

Week 3

Locally Weighted Regression (LOWESS)

Locally Weighted Regression Smoothing (LOWESS)

- Suppose we have a dataset with values $(x_1, y_1), \ldots, (x_n, y_n)$.
- We want to study the relationship between the x and the y values.
- In other words, we want to estimate the function $f(x_i)$, where $y_i = f(x_i) + \epsilon_i$.
- A crucial input into LOWESS is the the *span* (s).
  The span determines what proportion of the data is used when estimating $f(x)$ near a point $x = x_i$.
- The span is a number between 0 and 1; it controls the smoothness of the estimate $\hat{f}(x)$.

LOWESS (I)

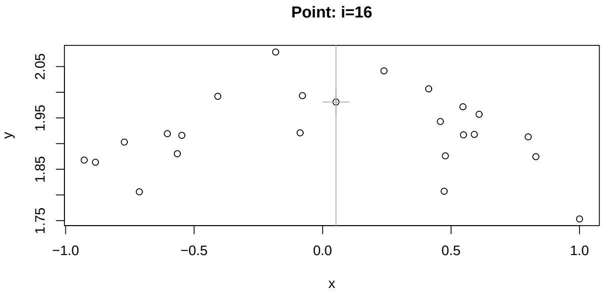
- To estimate the value of f at the point x_i, that is $\hat{f}(x_i)$ we take the following steps:
  1. Compute the value of $k = \lceil sn \rceil$.
  2. Find the k nearest neighbors of x_i. These points constitute a neighborhood $N(x_i)$ of x_i.
  3. Calculate the largest distance between x_i and another point in the neighborhood:
  $$\Delta(x_i) = \max_{x \in N(x_i)} |x_i - x|.$$

LOWESS (II)

4. Assign weights to each point in $N(x_i)$ using the tri-cube weight function:
$$w\left(\frac{|x - x_i|}{\Delta(x_i)}\right)$$
where
$$w(u) = \begin{cases} (1 - |u|^3)^3 & \text{for } |u| \le 1 \\ 0 & \text{otherwise} \end{cases}.$$
5. Calculate the weighted least squares fitting line on the data in the neighborhood $N(x_i)$.
6. Let $\hat{f}(x_i) = \hat{y}_i = \hat{a} + \hat{b}x_i$.

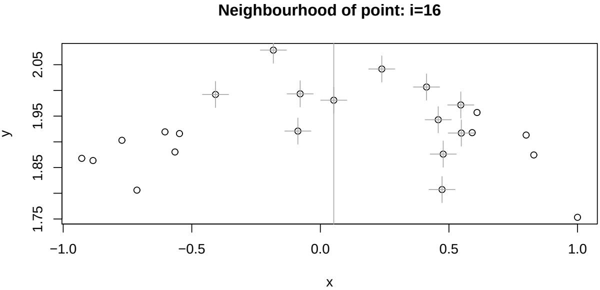
LOWESS Steps (1)

- Suppose we want to compute $\hat{f}(x_i)$. Suppose $s = 0.5$ and $n = 24$. Hence, $k = 12$.



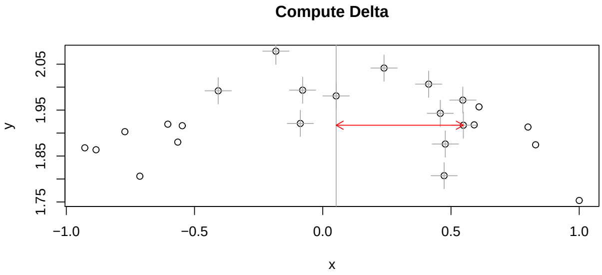
LOWESS Steps (2)

- Compute the neighbourhood $N(x_i)$.



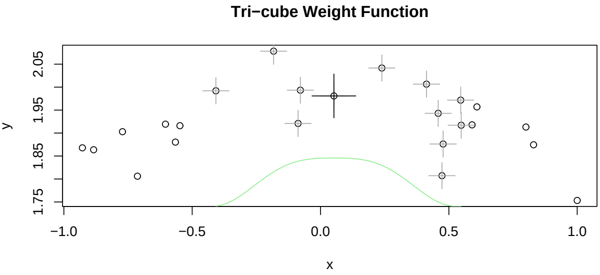
LOWESS Steps (3)

- Compute the value of Δ



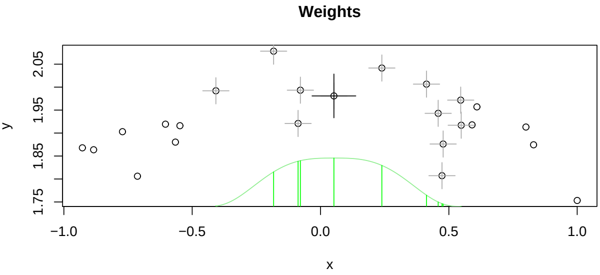
LOWESS Steps (4a)

- The tri-cube weight function is computed.



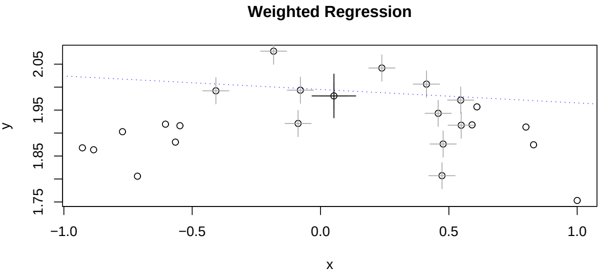
LOWESS Steps (4b)

- The weights are computed.



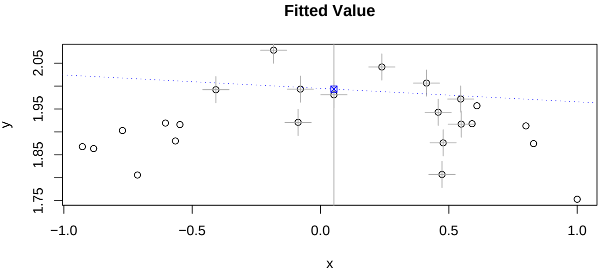
LOWESS Steps (5)

- A weighted least squares fit is found.



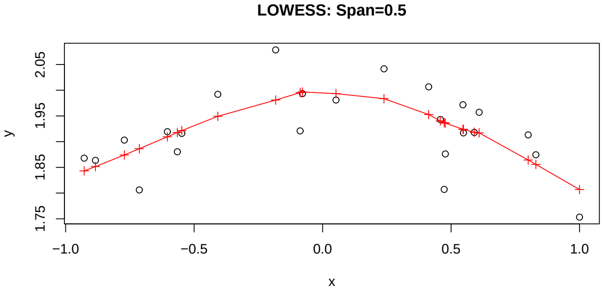
LOWESS Steps (6)

- The fitted value is found.



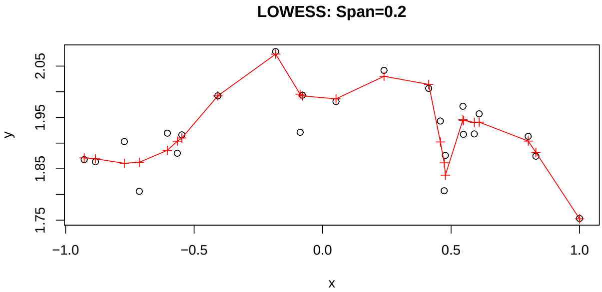
LOWESS Fits

- The steps are repeated for each $i = 1, 2, \ldots, n$. This gives the LOWESS curve.



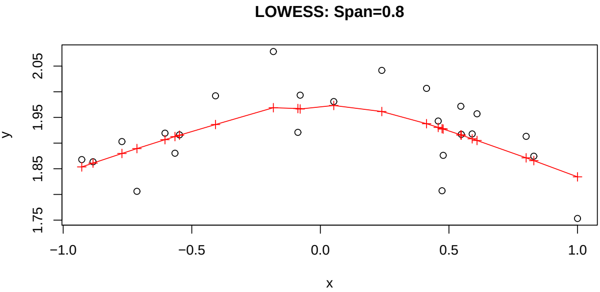
LOWESS (Span=0.2)

- If the span is smaller, then we get a more "wiggly" fit.



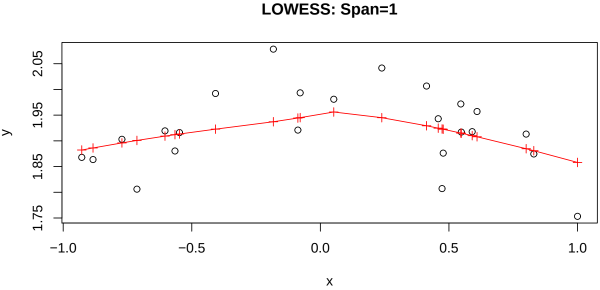
LOWESS Fits (Span=0.8)

- If the span is larger, then we get a more "smooth" fit.



LOWESS Fits (Span=1.0)

- If the span is even larger, then we get an even “smoother” fit.



LOWESS

- LOWESS is very easy to implement in R.
- Here's code to model the motorcycle data.

```
# Load the data
library(MASS)
data(mcycle)

# Plot the data
plot(mcycle)

# Fit a lowess regression
# The value of f controls the span
fit <- lowess(mcycle$times,mcycle$accel,f=2/3)

# Add the fitted values to the plot
points(fit,pch=3,col="red")

# Assess fit using mean squared error (MSE)
MSE <- mean((mcycle$accel-fit$y)^2)
MSE
```