Kernel Regression: introduction

Kernel regression is a flexible and powerful tool for estimating relationships in data, especially when the form of the relationship is unknown. Suppose that we want to estimate this the function $f(.)$ in this regression problem:

$$y_i = f(x_i) + \epsilon_i$$

where $f(.)$ is some 'smooth' function that we can estimate by the Nedaraya-Watson estimator which is as follows:

$$\hat{f} = \frac{\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x-x_i}{h}\right)}$$

Where h is a bandwith parameter. There exists many choices of kernel function K. In the case of a Gaussing kernel, we would have $K = \frac{1}{2\pi} e^{(1/2)(x-x_i)}$.

'mcycle ' dataset

mcycle: A data frame of 133 observations x 2 variables, giving a series of measurements of head acceleration in a simulated motorcycle accident, used to test crash helmets

times: in milliseconds after impact
accel: in g.

```
> head(mcycle)
  times accel
1   2.4   0.0
2   2.6  -1.3
3   3.2  -2.7
4   3.6   0.0
5   4.0  -2.7
6   6.2  -2.7
```

Source: Silverman, B. W. (1985) Some aspects of the spline smoothing approach to non-parametric curve fitting. Journal of the Royal Statistical Society series B 47, 1–52.

Reference: Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S-PLUS. Fourth Edition. Springer

R code

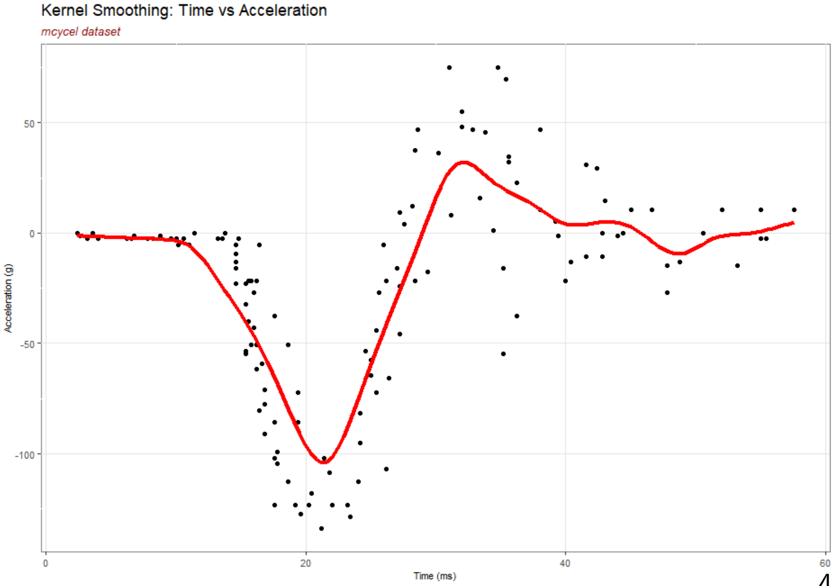
Here is the minimal code to run Kernel regression in R.

```
library(MASS)
library(tidyverse)
data(mcycle)

# fit smoothing splines model (ss) with default number of knots
mod1ks = ksmooth(mcycle$times, mcycle$accel,"normal", 4)
fit = data.frame(times = ks_fit$x, accel = ks_fit$y)

# plot the model
ggplot(mcycle, aes(x = times, y = accel)) +
  geom_point(color = "black") +
  geom_line(data = fit, aes(x = times, y = accel), color = "red", size = 1) +
 ...
```

Plot of the fitted model



Main observations

- The Kernel smoothing reveals a sharp increase in acceleration after 20 milliseconds, indicating a rapid change in speed.
- Following the initial spike, the acceleration decreases and shows a fluctuating pattern over time, suggesting variable speed changes during the motorcycle ride.
- The fitted kernel captures several local maxima and minima, highlighting the periods of acceleration and deceleration throughout the observed time span.
- Towards the latter part of the time series, the kernel regression indicates a gradual stabilization of acceleration, suggesting that the motorcycle's speed changes become less extreme.

References

Faraway, J. J., Extending the Linear Model (2006), ISBN 0-203-62105-0 (e-book)

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S-PLUS. Fourth Edition. Springer

The R Project for Statistical Computing:
https://www.r-project.org/