

Tutorial 10

Machine Learning and Big Data for Economics and Finance

This week's tutorial will only consist of practical activities using R and an exercise.

List of activities

I. Complete the activity from the textbook **Section 7.8 Lab: Non-linear Modeling**.

II. Exercise:

We observe an n -sample of the continuous random variables X and Y . We are interested in predicting Y given X using the model $Y = f(X) + \varepsilon$ for some unknown function f . Consider the kernel regression estimator

$$\hat{f}(x) = \frac{\sum_{i=1}^n y_i K_\lambda(x_i, x)}{\sum_{i'=1}^n K_\lambda(x_{i'}, x)},$$

where x is the point at which we are interested in making the prediction and $K_\lambda(x_i, x)$ is a kernel function with bandwidth $\lambda > 0$.

1. Show that $\hat{f}(x)$ can be written as a weighted sample average of Y .
2. Consider the special kernel function:

$$K_\lambda(x_i, x) = \frac{1}{2\lambda} I(-\lambda < x - x_i < \lambda),$$

- a. Draw the kernel function.
- b. Deduce that when λ is big enough, \hat{f} is just the usual sample average

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n y_i$$

Interpret.