

Problem Set 5

Non-parametric Regression

Consider the following data generating process:

$$y_i = a + \beta x_i + \varepsilon_i \tag{1}$$

where initially $\beta = 1.5$, $\varepsilon_i \sim \mathcal{N}(0, 1)$ and $x_i \sim \mathcal{N}(0, 1)$ for each $i = 1, \dots, n$ with $n = 200$.

1. Implementation

- Implement the Nadaraya-Watson estimator as introduced in the lecture using a Kernel function of your choice.
- Plot the results for one sample using three different bandwidths in one plot and compare the results.

Interpret the effect of X on y using your plots.

2. Causal Inference

- Consider a new DGP with $\beta = 1$, $Xe \sim \mathcal{N}(\mathbf{0}, \Sigma)$, where $x_i = Xe[i, 1]$ and $\varepsilon = Xe[i, 2]$ and

$$\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

- Estimate the effect of x_i on y_i using the nonparametric regression estimator above using again three different bandwidths.
- Use `set.seed(1)`, generate the data using the data generating process from 1. and 2..
- Compare your results, i.e. using both plots interpret the effect of X on y . What can you say about the causal effect of X on y given these estimated conditional mean functions?