Non-parametric Regression

- The outcome is a function of the regressor.
- Economic theory usually does not predict functional forms, but often implies constraints.

$$Y_i = f(X_i) + \varepsilon_i$$

Polynomial regression:
- Parametric estimation involves replacing f(x) with a polynomial and estimating the model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_q X_i^q + \varepsilon_i$$

where $q \geq 3$

Kernel regression:
- Non-parametric regression generally involves using some type of smoothing to estimate f(x). The simplest smoothing would be a moving average, where f(x*) is approximated by the average of points around x*.
- A popular non-parametric regression is the Gaussian kernel regression, which essentially is a weighted average of points close to x.

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

$$\text{where } K(a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ \& } h = \text{bandwidth}$$

- K is a kernel, there are many of them, but Gaussian is popular. They integrate to one and are symmetric.
- The bandwidth h is a smoothing parameter. The bigger the h, the more points will get significant weight, this will lead to under fitting. Small h will imply the opposite and lead to overfitting.
- Kernel regression does not require any estimation of parameters (the only parameter is h = bandwidth). Compute f-hat(x) on a finite grid on a relevant support.
- Another non-parametric regression is the locally weighted linear regression, where we are fitting many linear regression to approximate f(x).
- Local weighted regression motivated by taylor's theorem as it suggest that f(x) at a point can be approximated by a polynomial.

$$Y_i = \beta_0(x) + \beta_1(x) X_i + \varepsilon_i \quad \text{for } x \in [X_i - \delta, X_i + \delta]$$

$$\hookrightarrow \beta(x) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} w(x, x_i) (y_i - \beta_0 - \beta_1 x_i)^2 \right\}$$

where $w(x, x_i) = e^{-(x_i - x)^2 / 2\tau^2}$

- Need to estimate parameters each time you need to do an prediction. Would need to estimate many parameters before plotting estimated f(x). Closer points gets more weight.
- As seen by the locally-weighted regression example, "non-parametric regression" does indeed have parameters. However now the parameters are not fixed, but can rather increase in the sample size.

Spline regression:

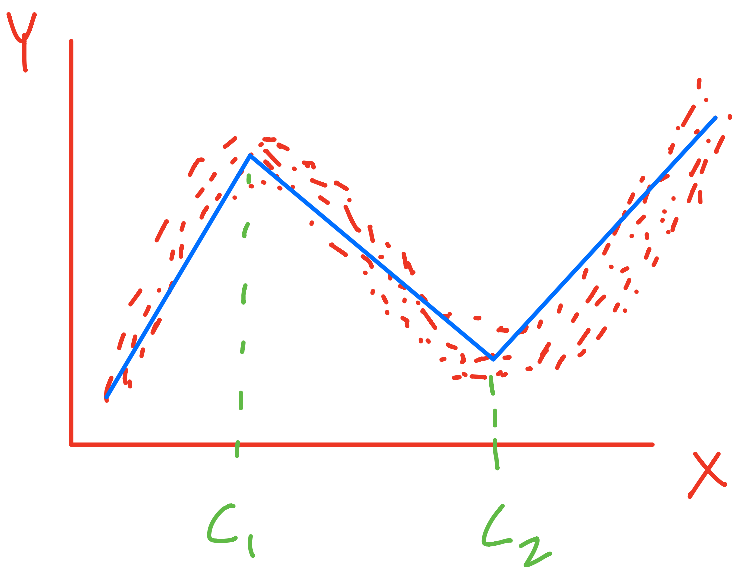
- Idea is to split the data into several portions and then fit a polynomial to each portion. Linear splines would fit the data with a bunch of lines of varying slopes.
- A spline regression is essentially a piecewise polynomial function.
- The point of devisions (point of change in piecewise function) are known as knots.
- A polynomial spline with degree D has the following form:

$$Y = \beta_0 + \sum_{i=1}^{D} \beta_i X^i + \sum_{j=1}^{K} \alpha_j (X - c_j)^D \times I(x > c_j) + \varepsilon$$

↳ Knot j is c_j

- Below is an example of a linear spline regression with 2 knots:

$$Y = b_0 + b_1 X + a_1 I(x > c_1)(x - c_1) + a_2 I(x > c_2)(x - c_2)$$



$\hat{Y}$ is linear spline

↳ c_1 & c_2 are knots

- Spline regression is estimated using the standard minimization of the least square error. Hence we can use the the closed form solution inv(X'X)X'Y for spline regression to estimate the parameters.