Regression Analysis Tutoring5

Seung Bong Jung

Seoul National University

November 1, 2021

Polynomial models

- Not every function can be approximated by linear function. For better approximation, one may use a high-order polynomial function.
- pth order polynomial model in predictor:

$$Y = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \epsilon$$

- One can fit a polynomial model by the multiple linear regression technique treating x_j for $1 \leq j \leq p$ as separate predictors.
- A second-order polynomial model in two predictors:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon$$

- Suppose we're considering $(n-1)$th order polynomial with n data points $\{(x_i, Y_i) : 1 \leq i \leq n\}$.
- As the order of polynomial increases, the matrix

$$\mathbf{X}^t\mathbf{X} = (\sum_{i=1}^{n} x_i^j x_j^k)_{0 \leq j,k \leq p}$$

  becomes ill-conditioned. This is because the smallest eigenvalue of the matrix gets closer to zero. Then the parameter estimates would be unstable numerically as well as statistically.
- To overcome such problem, using orthogonal polynomials may be useful.

Orthogonal polynomials

- Suppose we find a set of jth order polynomial P_j for $0 \leq j \leq p$ such that, for some α_j's depending on β_j's,

$$\beta_0 + \beta_1 x + \cdots + \beta_p x^p \overset{x}{\equiv} \alpha_0 P_0(x) + \alpha_1 P_1(x) + \cdots + \alpha_p P_p(x) \quad (1)$$

$$\sum_{i=1}^{n} P_j(x_i) P_k(x_i) = 0 \text{ for } 0 \leq j \neq k \leq p, P_0(x) \equiv 1 \quad (2)$$

- Then, the least squares estimator $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)^t$ is given by

$$\hat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i) Y_i}{\sum_{i=1}^{n} P_j(x_i)^2}, 1 \leq j \leq p, \qquad \hat{\alpha}_0 = \bar{Y}$$

- Note that higher-order polynomial function is not always successful.

Spline function

- Spline function: Let $\tau_1 < \tau_2 < \cdots \tau_K$ be preselected points in the range of x, called knots. We call f a spline function of order l if f is a polynomial of order l in each interval $[\tau_j, \tau_{j+1}]$ and has $(l-1)$ continuous derivative at the knots.
- Representation of spline functions:

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_l x^l + \sum_{j=1}^{K} \beta_{l+j}(x - \tau_j)_+^l,$$

  where $a_+ = \max\{0, a\}$.
- Note we need total $(l+1)(K+1) - K \cdot l = K + l + 1$ parameters to determine spline function.
- Higher K does not imply better estimation(approximation).

Local Approximation by Kernel Function

- One may approximate a function f at a point x by the local average f in the interval $[x-h, x+h]$ for a small $h > 0$:

$$\begin{aligned}\tilde{f}(x) &:= \arg\min_{\alpha} \int_{x-h}^{x+h} (f(u) - \alpha)^2 du \\ &= \int (2h)^{-1} I_{[-1,1]}(\frac{u-x}{h}) f(u) du\end{aligned}$$

- More generally, one may approximate $f(x)$ by a weighted local average of f:

$$\begin{aligned}\tilde{f}(x) &:= \arg\min_{\alpha} \int (f(u) - \alpha)^2 K(\frac{u-x}{h}) du \\ &= (\int K(\frac{u-x}{h}) du)^{-1} \int K(\frac{u-x}{h}) f(u) du\end{aligned}$$

Kernel Regression

- Taking smaller $h > 0$ does not imply better estimation of $m(x) := \mathsf{E}(Y|x)$.
- One can estimate $m(x)$ by

$$\hat{m}(x) = \arg\min_{\alpha} \sum_{i=1}^{n} (Y_i - \alpha)^2 K(\frac{x_i - x}{h})$$
$$= (\sum_{i=1}^{n} K(\frac{x_i - x}{h}))^{-1} \sum_{i=1}^{n} K(\frac{x_i - x}{h}) Y_i$$

- h is called the bandwidth, smoothing parameter or tuning parameter, and K is called the kernel.

Bandwidth selection

- Too small $h > 0$ leads too small number of observations in the kernel-weighted least squares estimation, and thus generates overfitting again.
- But, taking too large h would give over-smoothed estimates that lose important structure of the regression function.
- One useful technique to choose $h > 0$ is a cross-validation criterion. For example,

$$h_{\mathsf{CV}} := \arg\min_{h>0} \sum_{i=1}^{n} (Y_i - \hat{m}_{-i}(x_i, h))^2,$$

where $\hat{m}_{-i}(\cdot, h)$ denotes the kernel estimate constructed from the dataset with the ith pair (x_i, Y_i) being deleted and the bandwidth h being used.