What is nonparametric regression?

Jake Spiteri

July 21, 2022

The regression problem

Suppose we have a dataset $(y_i, x_i)_{i=1}^N$, where $y_i \in \mathbb{R}$, $x_i \in \mathbb{R}^p$ for $i = 1, \ldots, N$.

The response variable y is related to the covariate via

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

where $\mathbb{E}(\epsilon_i) = 0$, for $i = 1, \ldots, N$.

We are interested in learning the regression function f.

Linear regression

The linear regression model has the following form

$$f(x_i) = \beta_0 + \sum_{i=1}^{p} \beta_j x_{ij}.$$

We assume that the regression function is *linear*.

We can use basis expansions to introduce non-linearity.

Ridge regression

We minimize the following loss function with respect to $\beta \in \mathbb{R}^p$

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\|\beta\|_2^2,$$

which has solution

$$\hat{\beta} = (X^T X + \lambda I_p)^{-1} X^T y$$

Kernel ridge regression

From ridge regression: Replace data with a d-dimensional feature mapping: $x_i \to \phi(x_i)$,

We have the solution

$$\hat{\beta} = (\Phi_X^T \Phi_X + \lambda I_d)^{-1} \Phi_X^T y = \Phi_X (\Phi_X \Phi_X^T + \lambda I_N)^{-1} y$$

Nonparametric statistics = infinite-dimensional statistics

Kernel ridge regression

Let $\mathcal{H}$ be a reproducing kernel Hilbert space, with associated kernel $k(x, y) = \langle \phi(x), \phi(y) \rangle$. Then the kernel ridge regression problem is:

$$\min_{f \in \mathcal{H}} \{ \frac{1}{N} \sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \|f\|_{\mathcal{H}}^2 \}$$

Representer theorem: The solution to the above problem is of the form $f(x) = \sum_{i=1}^{N} \beta_i k(x_i, x)$.

Nonparametric statistics = optimizing over a large function space

Smoothing splines

We minimize the following over a Sobolev space

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i))^2 + \lambda \int \{f''(t)\}^2 dt.$$

The solution is a natural cubic spline. To me, this is clearly nonparametric.

Cubic splines are *linear smoothers*:

$$f(x) = \sum_{i=1}^{N} \beta_j N_j(x)$$

GAMs

The model has the form

$$f(x) = \sum_{i=1}^{p} f_i(x).$$

- Parametric: $f_i(x) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$
- Nonparametric: f_i is a non-linear smoother (i.e. spline).
- Semi-parametric: A mix of both parametric and nonparametric.

Conclusion

Linear sums of smoother features are not parametric.

We are optimizing over a very large class of functions!