Q27

Rahul Atre

2023-11-06

Q27 [Nonlinear Modeling]

Consider the two curves, $\hat{g_1}$, $\hat{g_2}$, defined by

$$\hat{g_1} = \arg min_g (\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int [g^3(x)] dx),$$

$$\hat{g_2} = \arg min_g (\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int [g^4(x)] dx),$$

where $g^{(m)}$ represents the mth derivative of g.

Note: The following two curves are known as smoothing splines. According to wikipedia.org, Smoothing splines are function estimates, obtained from a set of noisy observations y_i of the target $f(x_i)$ to balance a goodness of fit measure with a derivative based measure of $\hat{f}(x)$.

1. As $\lambda \to \infty$, will $\hat{g_1}$ or $\hat{g_2}$ have the smaller training RSS?

Examining both functions, we can see that $\hat{g_1}$ and $\hat{g_2}$ both have an integral for the respective derivative of g(x) (3rd and 4th). As we know from calculus, the derivative and integral are opposites and cancel, so we are left with a lower derivative power.

In the case of $\hat{g_1}$, as $\lambda \to \infty$, $g^3(x) \to 0$. Since the third derivative becomes negligible, the function $\hat{g_1}$ becomes quadratic.

Similarly for $\hat{g_2}$, as $\lambda \to \infty$, $g^4(x) \to 0$. Since the forth derivative becomes negligible, the function $\hat{g_2}$ becomes cubic.

As we know, a cubic function is better at modelling the training data than a quadratic due to more flexibility. Therefore, $\hat{g_2}$ will have a smaller training RSS.

2. As $\lambda \to \infty$, will $\hat{g_1}$ or $\hat{g_2}$ have the smaller test RSS?

From a similar previous question (q12), there is not enough information to make a decisive statement on whether a quadratic or cubic regression would have smaller test RSS, if we don't know the true relationship of X and Y. Though a cubic regression models the training data better from overfitting, there is no proof that the same accuracy would follow for the test data.

3. For $\lambda = 0$, will $\hat{g_1}$ or $\hat{g_2}$ have the smaller training and test RSS?

If $\lambda = 0$, then $\lambda \int [g^3(x)] dx = \lambda \int [g^4(x)] dx = 0$. Consequently, both $\hat{g}_1$ and $\hat{g}_2$ will have the same training and test RSS.

References:

- https://en.wikipedia.org/wiki/Smoothing_spline