Assignment24

John Bute

2024-11-09

R Markdown

Suppose that we use some statistical learning method to make a prediction for the response Y for a particular value of the predictor X. Carefully describe how we might estimate the standard deviation of our prediction

The standard deviation of a learning model is in fact calculating the uncertainty of our model. This uncertainty accounts for the model's variability and irreducible error

Now, this irreducible error is part of our prediction error (total error of a single prediction), which is also composed of the model eror, or the error due to the model's inability to perfectly fit to the data:

For some single prediction: = (- Y)^2

Now, the standard deviation is composed of the variance of the model's prediction and variance of its error term (or the inherent noise in data)

Thus, the total prediction variance of given a term X is

$$\text{Var}(\hat{Y} \mid X) = \text{Var}(f(X)) + \sigma^2$$

However, it is complex to estimate the variability of our predictions, especially if complex. Thus, we use bootstrapping, a non-parametric approach that estimates variability without making stron assumptions. We draw multiple samples from the data, with replacement, and for each bootstrapped sample, we fit a model and predictions are made for a value X, and we repeat this as may times.

Thus, we draw M bootstrap samples, train a model and make a prediction using model i, then finally calculate the mean of the bootstrap predictions, before giving an estimation of the standard error

The mean of predictions:

$$\bar{Y} = \frac{1}{M} \sum_{i=1}^{M} \hat{Y}_i$$

The standard error of predictions:

$$\widehat{se}(\hat{Y}) \approx \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}\left(\hat{Y}_i - \bar{Y}\right)^2}$$

Thus, bootstrapping allows us to estimate uncertainty in our predictions by observing how predictions vary across samples of our data.