Q24

Rahul Atre

2023-11-02

Q24 [Resampling Methods]

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.

Ans: In order to estimate the standard deviation of our prediction, we can use a re-sampling method to draw out different samples from the available data and model the various parameter estimates. This allows us to examine the extent to which each sample differs from one another.

The **bootstrap method** provides an estimate for the variability of a statistical model parameter (in this case the standard deviation), without the need to generate new observations. As mentioned above, it does so by re-sampling the data to simulate the process of obtaining new replicates, i.e., sampling observations of the original dataset with replacement.

The process is repeated M times to obtain bootstrap samples Tr_i^* and parameter estimate $\hat{\alpha}_i^*$, for i = 1, ..., M, for which, we derive a bootstrap estimate

$$\hat{\alpha}^* = \frac{1}{M} \sum_{i=1}^{M} \hat{\alpha}_i^*,$$

with standard error with

$$\hat{se}(\hat{\alpha}^*) = \sqrt{\frac{1}{M-1} \sum_{i=1}^{M} (\hat{\alpha_i^*} - \hat{\alpha^*})^2}$$

The standard error is an estimate of the standard deviation that is generated from the repeated sampling.

From the original datasets observations, we can create M such bootstrap samples and for each sample train the statistical learning method to obtain the predicted response Y. Lastly, we calculate the estimated standard deviation using the obtained M predictions.