Measuring Importance of Inputs for Bagged Predictors

Stephen Vardeman
Analytics Iowa LLC
ISU Statistics and IMSE

Average error across a bootstrap sample

An idea of Breiman (phrased originally for random forests, but relevant to any bagged predictor) is this. For every bootstrap sample $\mathbf{T}_b^*$ and predictor $\hat{f}^{*b}$ based on the corresponding remainder $\mathbf{T} - \mathbf{T}_b^*$, one can compute a bth average error across the corresponding OOB sample, say

$$\overline{\text{err}}_b = \frac{1}{\#\left[i \middle| \begin{array}{c}\text{case } i \text{ is not in the}\\ \text{bootstrap sample } b\end{array} \right]} \sum_{\substack{i \text{ s.t. case } i \text{ is not in the}\\ \text{the bootstrap sample } b}} L\left(\hat{f}^{*b}\left(\mathbf{x}_i\right), y_i \right)$$

Increase from permuting values of *j*th input

Then in the OOB sample randomly permute among cases the values of the jth coordinate of the input vectors, producing, say, input vectors $\tilde{\mathbf{x}}_i^j$. One can then define

$$\widetilde{\mathrm{err}}_b^j = \frac{1}{\#\left[i \middle| \begin{array}{c} \text{case } i \text{ is not in the} \\ \text{bootstrap sample } b \end{array} \right]} \sum_{\substack{i \text{ s.t. case } i \text{ is not in the} \\ \text{the bootstrap sample } b}} L\left(\hat{f}^{*b}\left(\tilde{\mathbf{x}}_i^j\right), y_i \right)$$

and take the difference

$$I_b^j = \widetilde{\mathrm{err}}_b^j - \overline{\mathrm{err}}_b$$

as an indicator of the importance of variable j to prediction for the bth bootstrap sample.

Bagged predictor variable importance

These can then be averaged across the B bootstrap samples to produce

$$I^j = \frac{1}{B} \sum_{b=1}^{B} I_b^j$$

as a bagged predictor variable importance measure for variable j. These can be compared across j. Typically they will be positive and large values are indicative of high variable importance.

When applied to its specially constructed trees, this produces a variable importance measure for a random forest. (What is made is then something different than what was suggested earlier for a predictor that is ultimately an average of tree predictors, that could alternatively be employed for the random forest.)