The Stacking Scheme

Team Proxima Centuari: Danny, James, Kenny, Wenchang
(drafted by James)

March 6, 2018

We propose that the stacked prediction, should we choose to use and combine estimates from N different methods, is

$$\vec{y}_{\text{Stacked}} = \frac{1}{N-1} \sum_{j=1}^{N} \left(1 - \frac{\rho_j}{s}\right) \vec{y}_j \tag{1}$$

where $\vec{y}_j$ is the prediction made with the j-th method, ρ_j is the performance metric for using the j-th method, and

$$s = \rho_1 + \rho_2 + \ldots + \rho_N. \tag{2}$$

Also see the following table:

Estimator	Performance	Prediction
Method 1	ρ_1	$\vec{y}_1$
Method 2	ρ_2	$\vec{y}_2$
$\ldots$	$\ldots$	$\ldots$
Method N	ρ_N	$\vec{y}_N$

We define the performance metric[1] ρ_i in such a way that the smaller the ρ_i, the better the performance. Candidates for ρ_i may include

$$\text{RMSE}, \quad 1 - R^2, \quad \frac{\text{RMSE}}{R^2}, \quad \ldots$$

The combination coefficient

$$1 - \frac{\rho_j}{s}$$

gives more weights to better performing estimators (i.e. the ones with smaller ρ_i), and the whole combination given by (1) is a convex combination.

[1]The metric is computed from simulated validation tests.

Check:[2] Suppose under the rarest situation[3] that

$$\vec{y}_1 = \vec{y}_2 = \ldots = \vec{y}_N = \vec{y}_{\text{Test}}$$

then by (1) and (2),

$$\begin{aligned}
\vec{y}_{\text{Stacked}} &= \frac{1}{N-1}\left(N - \frac{\rho_1 + \rho_2 + \ldots + \rho_N}{s}\right)\vec{y}_{\text{Test}} \\
&= \frac{1}{N-1}\left(N - \frac{s}{s}\right)\vec{y}_{\text{Test}} \\
&= \vec{y}_{\text{Test}}
\end{aligned}$$

which is the desired result.

Example: Say we obtain three different predictions from three different estimators ($N = 3$) as follows:

Estimator	Performance	Prediction
Method 1	ρ_1	$\vec{y}_1$
Method 2	ρ_2	$\vec{y}_2$
Method 3	ρ_3	$\vec{y}_3$

By (1) and (2),

$$\vec{y}_{\text{Stacked}} = \frac{1}{2}\left[\left(1 - \frac{\rho_1}{s}\right)\vec{y}_1 + \left(1 - \frac{\rho_2}{s}\right)\vec{y}_2 + \left(1 - \frac{\rho_3}{s}\right)\vec{y}_3\right]$$

where

$$s = \rho_1 + \rho_2 + \rho_3.$$

[2] Not a proof!

[3] It is possible although improbable.