Stacking (MSBA 7027)

Zhengli Wang

Faculty of Business and Economics
The University of Hong Kong
2023

Stacking

Combine the predictions of several base learners.

1. base learners (e.g. single RF or GBM) trained using training data,
2. a combiner or meta algorithm, a.k.a. *super learner*, trained to make a final prediction based on predictions of base learners.

Tend to outperform any individual base learners

Note:
The base learners can be very diverse
Each base learner can be a very complicated method (e.g. RF)

Stacking

The Super Learner Algorithm: Overview

Three Phases

1. Set up the ensemble
2. Train the ensemble
3. Predict on new data

Stacking

The Super Learner Algorithm: 1. Set up the ensemble

- Specify a list of L base learners
  - Each with specific model hyper-parameters

- Specify a meta learning algorithm.
  - Can be any algorithms discussed

Stacking

The Super Learner Algorithm: 2. Train the ensemble

- Train each of the L base learners on the training set.

- Perform *k*-fold CV on each base learner, collect the CV predictions from each
  - same *k*-folds must be used for all base learners.
  - These predicted values represent $p_1,\ldots,p_L$ (details see next slide)

Stacking

The Super Learner Algorithm: 2. Train the ensemble

- The n cross-validated predicted values from each of the L algorithms can be combined to form a new n×L feature matrix
  - This matrix, along with the original response vector (y), are called the "level-one" data.

$$n\left\{\begin{bmatrix} \\ p_1 \\ \\ \end{bmatrix} \cdots \begin{bmatrix} \\ p_L \\ \\ \end{bmatrix} \begin{bmatrix} \\ y \\ \\ \end{bmatrix} \to n\left\{\overbrace{\begin{bmatrix} & & \\ & Z & \\ & & \end{bmatrix}}^{L} \begin{bmatrix} \\ y \\ \\ \end{bmatrix}\right.\right.$$

(n = number of rows in the training set)

- Train the meta learning algorithm on the level-one data (y=f(Z)).
- The "ensemble model": L base learning models + meta learning model is then be used to generate predictions on new data.

(This slide is intended to be empty)

Stacking

The Super Learner Algorithm: 3. Predict on new data

- To generate ensemble predictions
  - 1) generate predictions from the base learners.
  - 2) Feed those predictions into the meta learner to generate the ensemble prediction.

Stacking

The Super Learner Algorithm: Note

- Stacking never does worse than selecting the single best base learner on the **training data**
  - but not necessarily the **validation** or **test data**

- Performance is great when:
  - stacking base learners that have **high variability & uncorrelated** predicted values

- The more similar the predicted values are between the base learners, the less advantage there is to combining them

End