Super Learners

Machine Learning

Ignacio Sarmiento-Barbieri

Universidad de La Plata

Agenda

1 Superlearners

Agenda

1 Superlearners

Recap

- Queremos predecir y en funcion de observables $(\mathbf{x}_i)$

$$y = f(\mathbf{x}_i) + u \tag{1}$$

- donde la estimación de f implica la que mejor generalize (prediga mejor fuera de muestra):

$$\hat{f} = \underset{f}{\text{argmin}} \left\{ \sum_{i=1}^{n} L(y_i, f(\mathbf{x}_i; \Theta)) \right\} \tag{2}$$

Superlearners: Motivation

- Superlearning is a technique for prediction that involves combining many individual statistical algorithms to create a new, single prediction algorithm that is expected to perform at least as well as any of the individual algorithms.
- The inovation?

Superlearners: Motivation

- Superlearning is a technique for prediction that involves combining many individual statistical algorithms to create a new, single prediction algorithm that is expected to perform at least as well as any of the individual algorithms.
- The inovation?
- The superlearner algorithm "decides" how to combine, or weight, the individual algorithms based upon how well each one minimizes a specified loss function
- The motivation for this type of "ensembling" is that a mix of multiple algorithms may be more optimal for a given data set than any single algorithm.
- For example, a tree based model averaged with a linear model (e.g. random forests and LASSO) could smooth some of the model's edges to improve predictive performance.

Superlearners: Algorithm

Denote the library $\mathcal{L}$ and its cardinality as $V(n)$.

1. Fit each algorithm in $\mathcal{L}$ on the entire data set $X = \{X_i : i = 1, ..., n\}$ to estimate $f_v(X)$ with $v = 1, \ldots, V(n)$

Superlearners: Algorithm

2. Split the data set X into a training and validation sample, according to a K-fold cross-validation scheme:

Superlearners: Algorithm

3. For the kth fold, fit each algorithm in $\mathcal{L}$ on $T(k)$ and save the predictions on the corresponding test , $\hat{f}_{k,T}(X_i)$ with $X \in Te(k)$

Superlearners: Algorithm

4. Bind the predictions from each algorithm together to create a n by V matrix

Superlearners: Algorithm

5. Propose a family of weighted combinations of the candidate estimators indexed by weight-vector α:

Superlearners: Algorithm

6. Determine the α that minimizes the cross-validated risk of the candidate estimator $\sum_{v=1}^{V} \alpha_v \hat{f}(X_i)$ over all allowed α-combinations:

Superlearners: Algorithm

7. Combine $\hat{\alpha}_v$ with $\hat{f}_v(X_i)$ according to the weights found, and create the final super learner fit

Superlearners: Algorithm

Some considerations:

- The super learner theory does not place any restrictions on the family of weighted combinations used for ensembling the algorithms in the library.
- The restriction of the parameter space for α to be the convex combination of the algorithms in the library provides greater stability of the final super learner prediction.

Example: Superlearners

photo from https://www.dailydot.com/parsec/batman-1966-labels-tumblr-twitter-vine/