A *Practical* Tour of Ensemble (Machine) Learning

Nima Hejazi [1] Evan Muzzall [2]

[1] Division of Biostatistics, University of California, Berkeley

[2] D-Lab, University of California, Berkeley

slides: https://goo.gl/wWa9QC

PUBLIC DOMAIN

Ensemble Learning – What?

> In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone.
>
> - Wikipedia, November 2016

Ensemble Learning – Why?

- Ensemble methods outperform individual (base) learning algorithms.
- By combining a set of individual learning algorithms using a *metalearning* algorithm, ensemble methods can approximate complex functional relationships.
- When the true functional relationship is not in the set of base learning algorithms, ensemble methods approximate the true function well.
- *n.b.*, ensemble methods can, even asymptotically, perform only as well as the best weighted combination of the candidate learners.

Ensemble Learning – How?

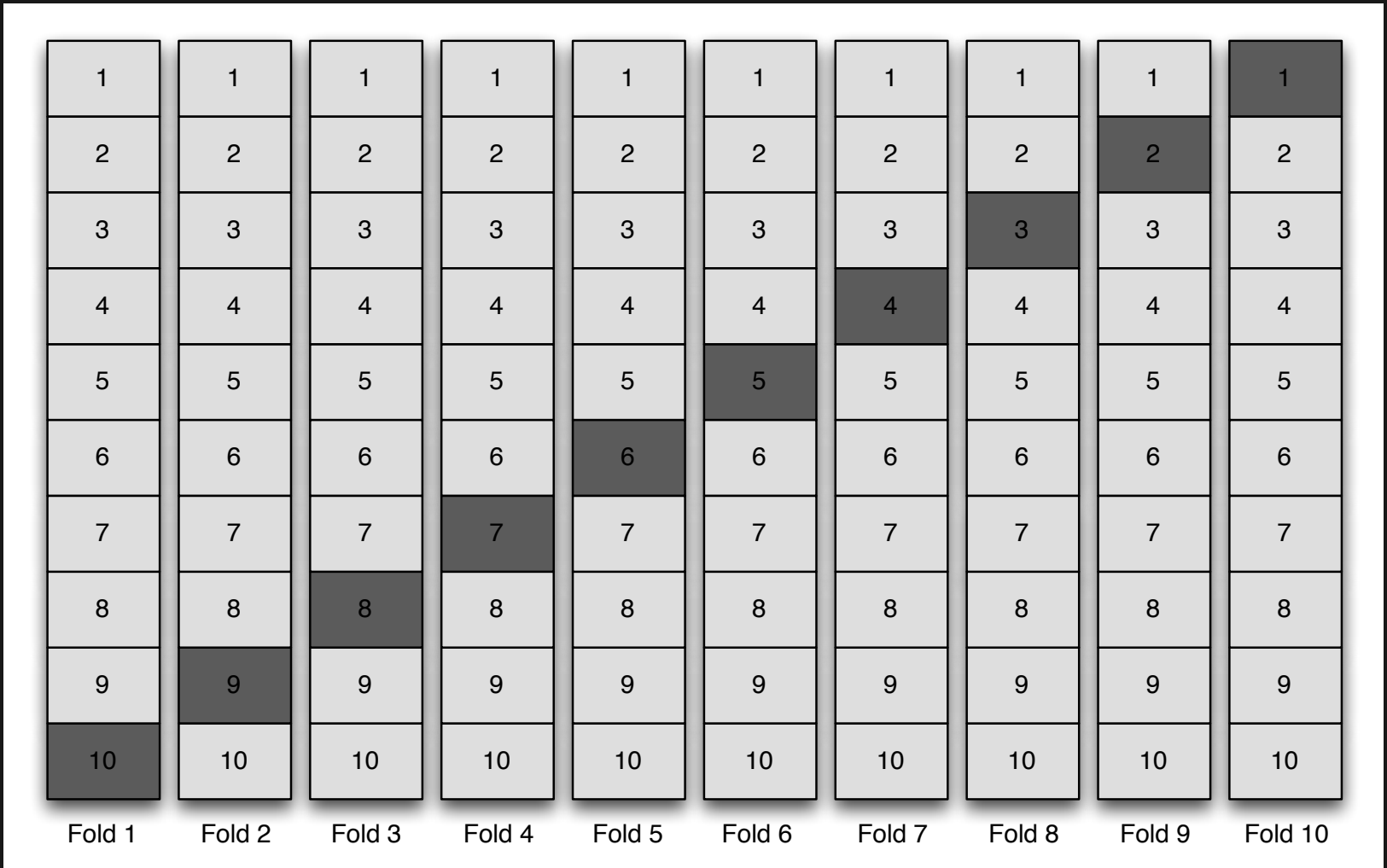
Common strategies for performing ensemble learning:

- **Bagging** – reduces variance and increases accuracy; robust against outliers; often used with decision trees (*i.e.*, Random Forest).
- **Boosting** – reduces variance and increases accuracy; not robust against outliers or noise; accomodates any loss function.
- **Stacking** – used in combining "strong" learners; requires a *metalearning* algorithm to combine the set of learners.

Introduction to Super Learner

- 1996 paper “Stacked Regressions” (L. Breiman) introduced the notion of model stacking using k-fold cross-validation, the precursor to the modern Super Learner algorithm.
- 2007 paper “Super Learner” (van der Laan *et al.*) worked out theoretical details on the optimality of stacking. Before this, the reasons for the superb performance of stacking were unknown.
- The Super Learner algorithm learns the optimal combination of the base learner fits in a manner that is provably asymptotic optimal.

Interlude: Cross-Validation



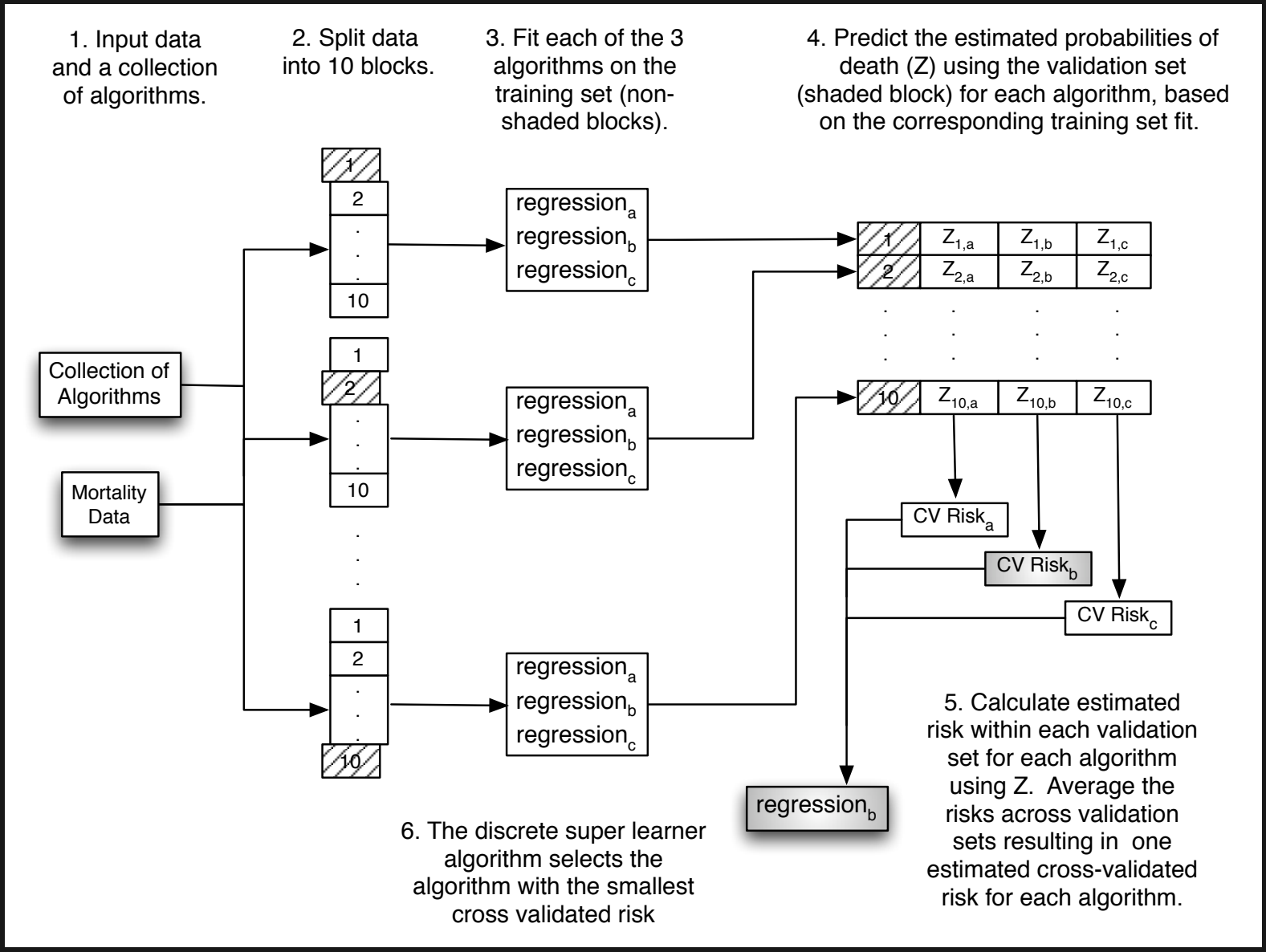
The validation set rotates V times such that each set is used as the validation set once.

Optimality of Super Learner

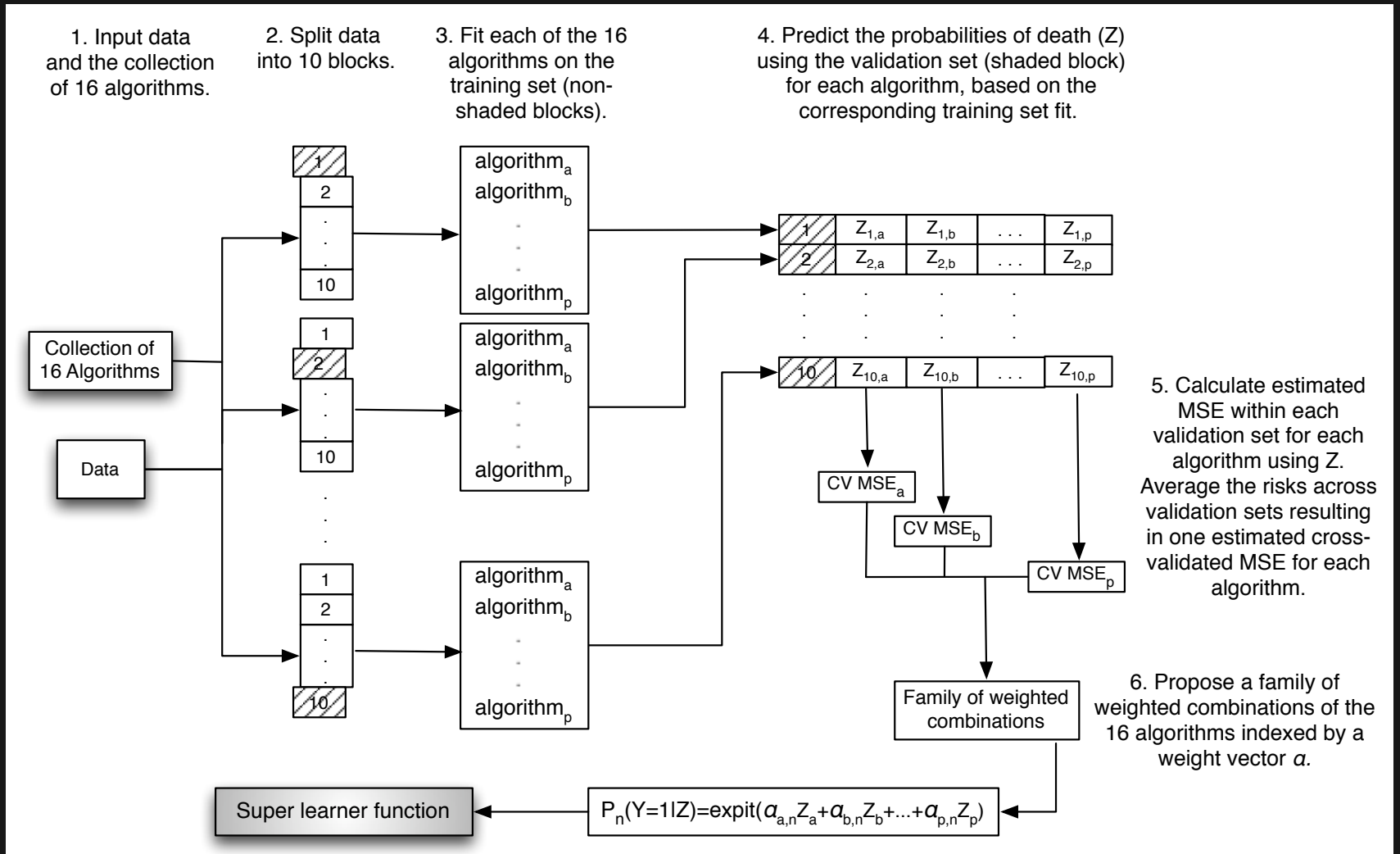
For a random variable $O = (W, A, Y)$, let the **oracle selector** be a rule that picks the algorithm with the lowest cross-validated risk under the *true probability distribution* P_0. The **oracle selector** is unknown because it depends on observed data and the truth.

Asymptotic results prove that in realistic scenarios (where none of the algorithms represent the true relationship), the "discrete super learner" performs *asymptotically as well as* the oracle selector (the best estimator given the algorithms in the collection).

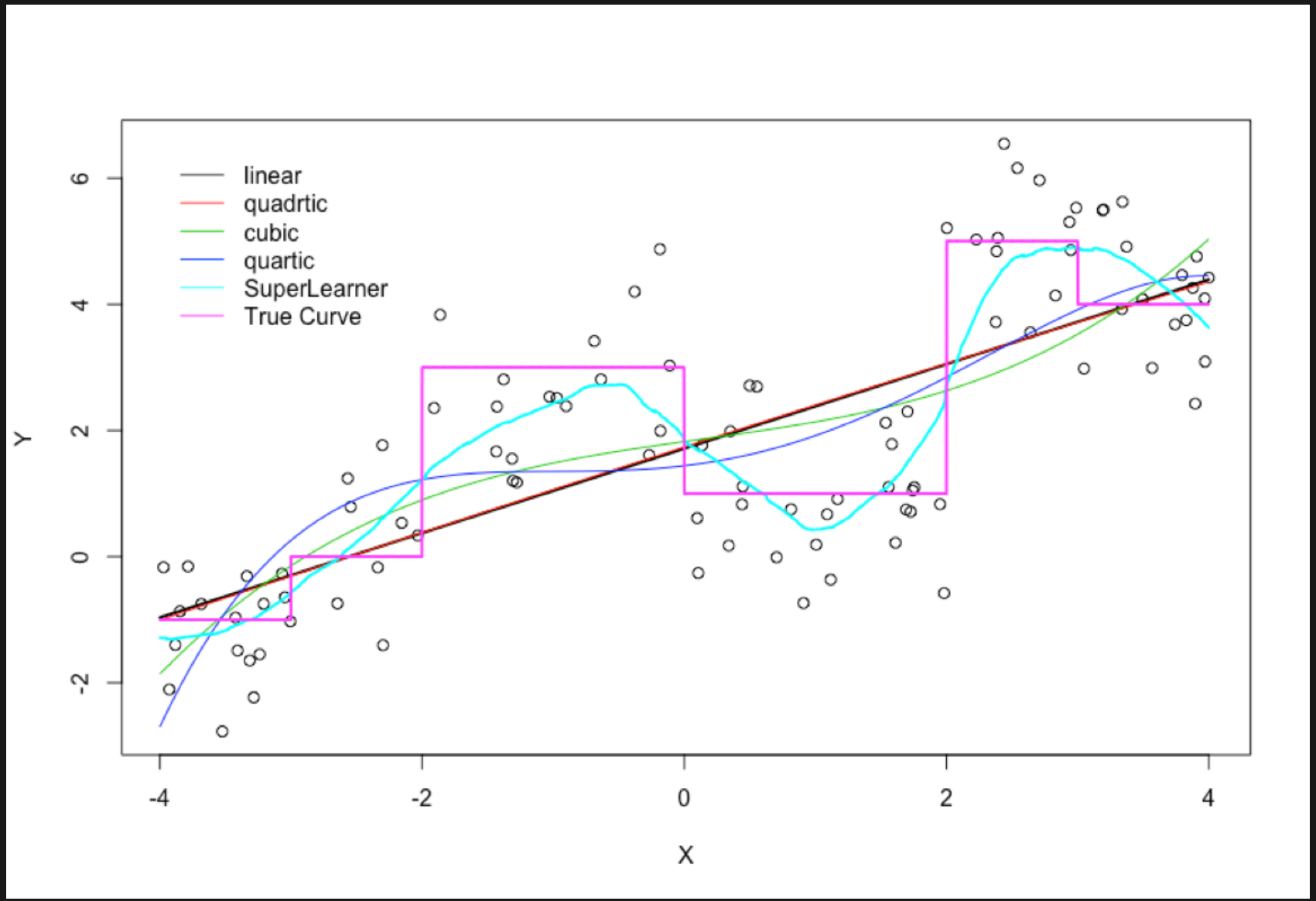
The Discrete Super Learner



The Super Learner Algorithm



Ensembles with Super Learner



R Package: “SuperLearner”

- Implements the Super Learner prediction method (stacking) and contains a library of prediction algorithms to be used in the Super Learner.
- Provides a clean interface to numerous algorithms in R and defines a consistent API for extensibility.

R Package: “h2oEnsemble”

Extension to the “h2o” R package that allows the user to train an ensemble in the H2O cluster using any of the supervised machine learning algorithms in H2O.

- Uses data-distributed and parallelized Java-based algorithms for the ensemble.
- All training and data processing are performed in the high-performance H2O cluster.
- Supports regression and binary classification.

Summary

1. Ensemble methods combine individual learning algorithms to approximate complex relationships.

2. Super Learning (stacking) represents an optimal system for combining individual learning algorithms into an ensemble learner.

3. The “SuperLearner” R package provides a well-maintained implementation of the the Super Learner algorithm.

4. The “h2oEnsemble” R package provides access to a range of ensemble methods, developed by `H2O.ai`.

Slides: http://goo.gl/wWa9QC

GitHub: nhejazi/talk-h2oSL-THW-2016