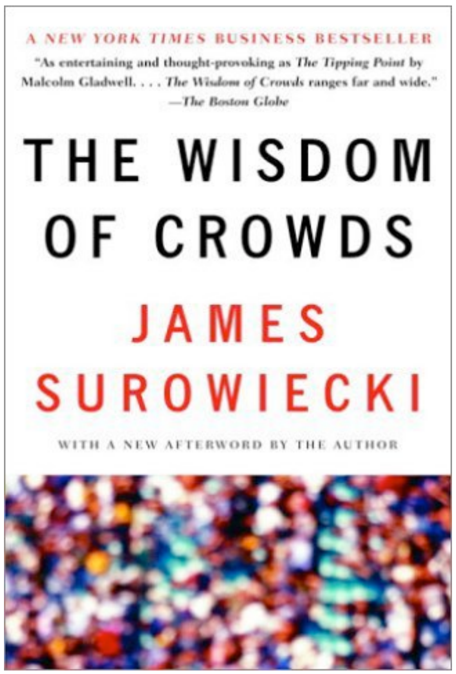
Introduction to Data Science

INTRO TO ENSEMBLES AND STACKING

BRIAN D'ALESSANDRO

WISDOM OF CROWDS

This concept can be applied to Machine Learning.
This is called Ensemble Learning



Conditions for This to Work

Diversity
Each person should have private information

Independence
Peoples opinions aren't determined by opinions of others

Aggregation
Some mechanism exists for turning individual opinions into a collective decision.

TYPES OF ENSEMBLE METHODS

Stacking

Taking a weighted combination of the predictions of a total of *S* different classifiers.

Bagging

Generating multiple classifiers from the same data by resampling, and aggregating the multiple classifications into a single prediction

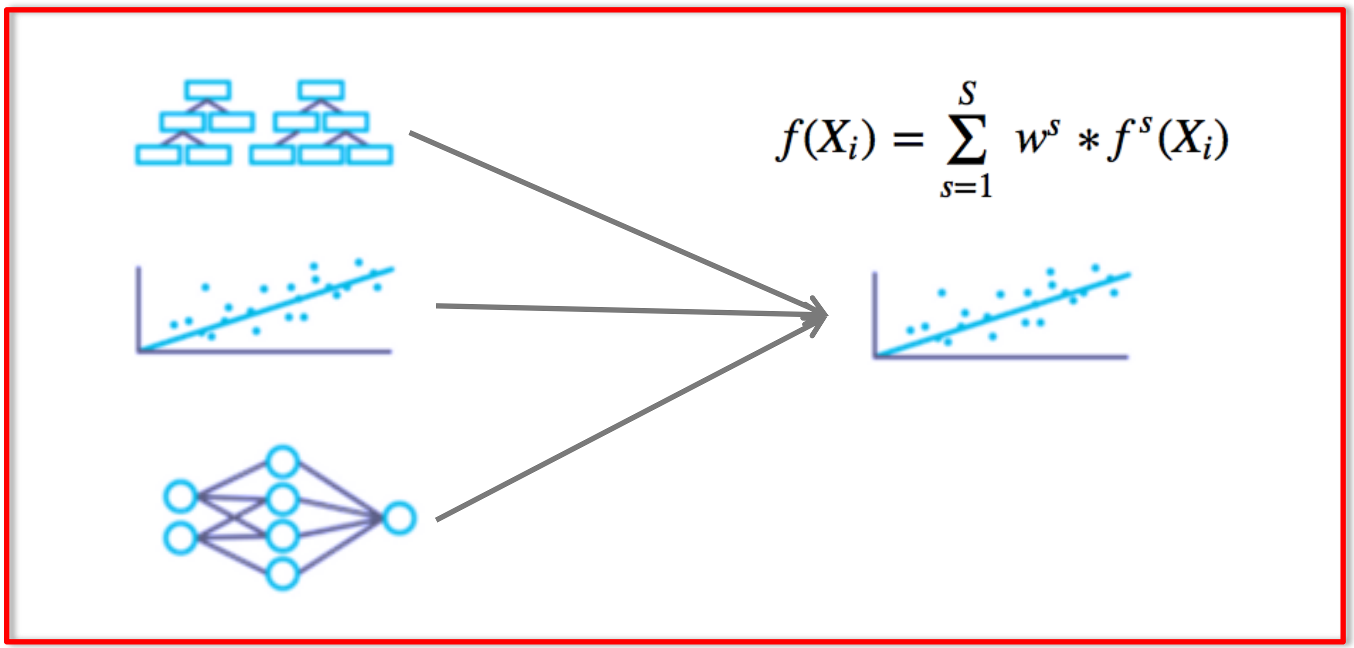
Boosting

Building new classifiers iteratively, using cumulative errors to inform or weight additional classifiers. Aggregating the set of classifiers into a single prediction.

A SIMPLE ENSEMBLE METHOD

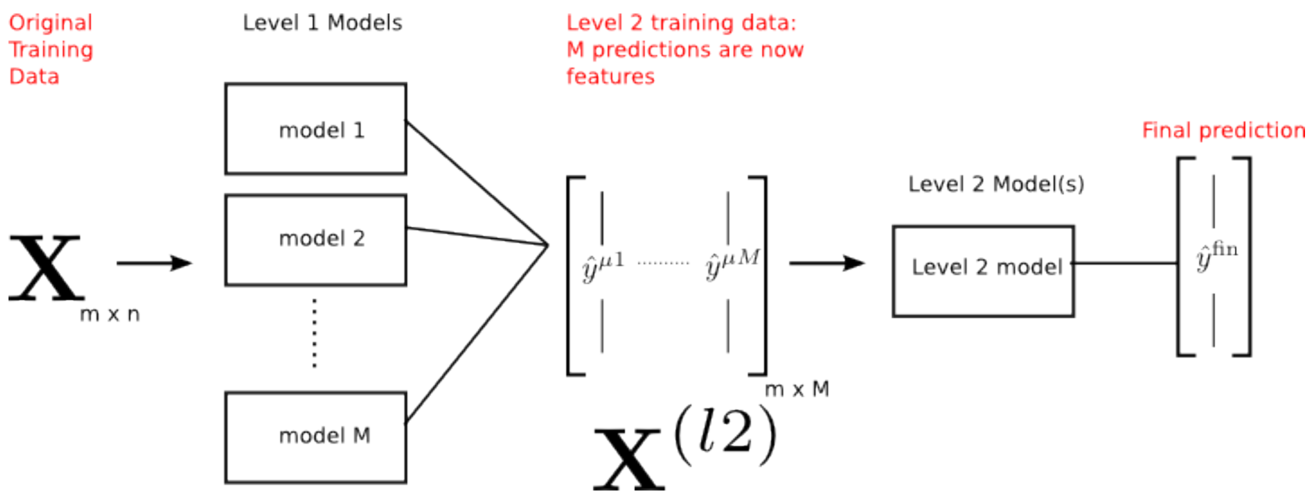
Stacking – Let $f^s(X_i)$ be some prediction on a sample using some arbitrary classifier s. Stacking is the process of taking a weighted combination of the predictions of a total of S different classifiers.

The weights can be a simple average, or learned via a secondary classification/regression process.



BUILDING A STACKED ENSEMBLE

When we wish to learn the weights (using ML), we build the ensemble in two stages.



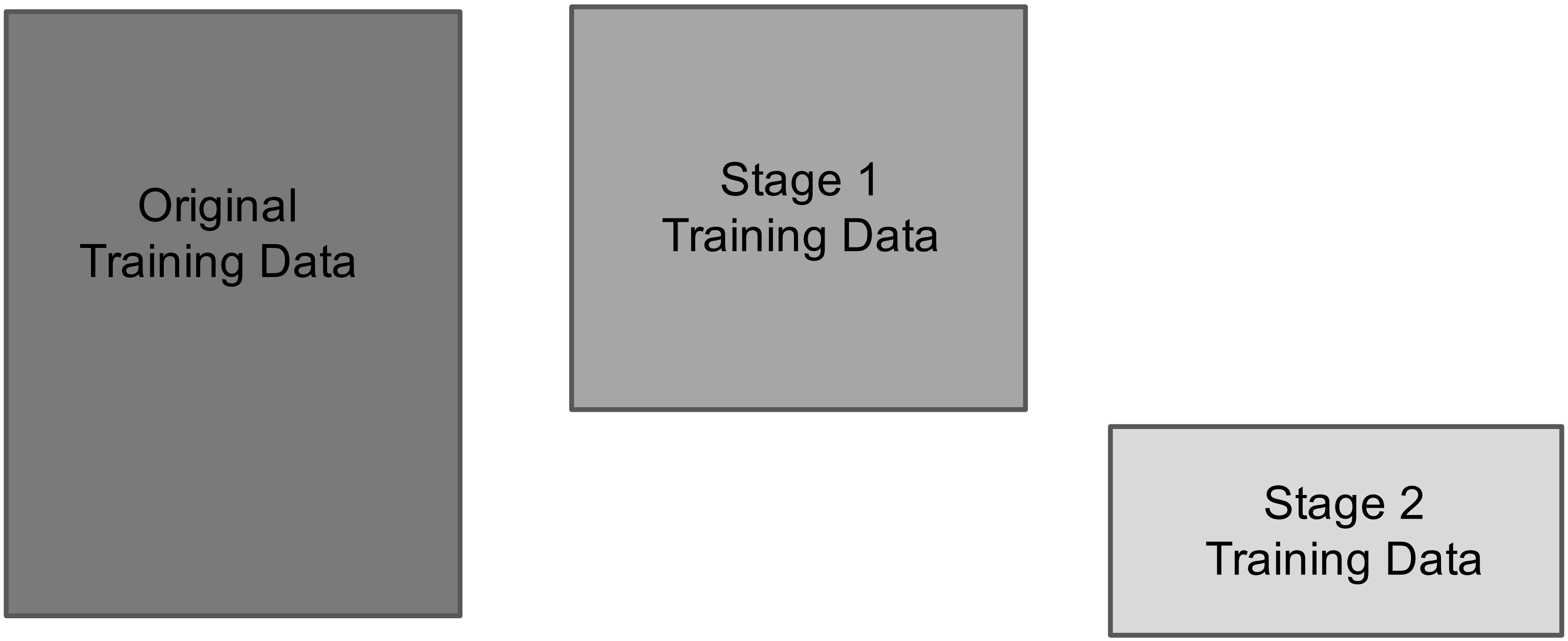
Stage 1: Build M different (and diverse) classifiers on the original training data.

Stage 2: Score each of the M classifiers from stage 1 on stage 2 training data. Use the output of M stage 1 classifiers as features for stage 2 model

Image source: https://www.kdnuggets.com/2017/02/stacking-models-imropved-predictions.html

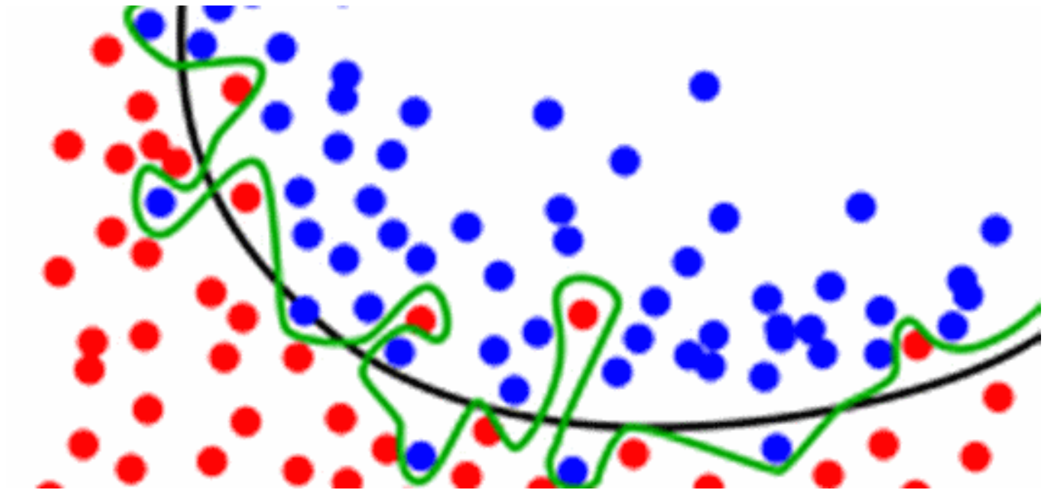
BEING CAREFUL ABOUT GENERALIZATION

In order for learned stacking to generalize appropriately, the stage 2 model needs to be built from data that is separate from the stage 1 models. Otherwise the system risks overfitting. Note that this is best accomplished when sample sizes are large.



Further reading: https://bradleyboehmke.github.io/HOML/stacking.html

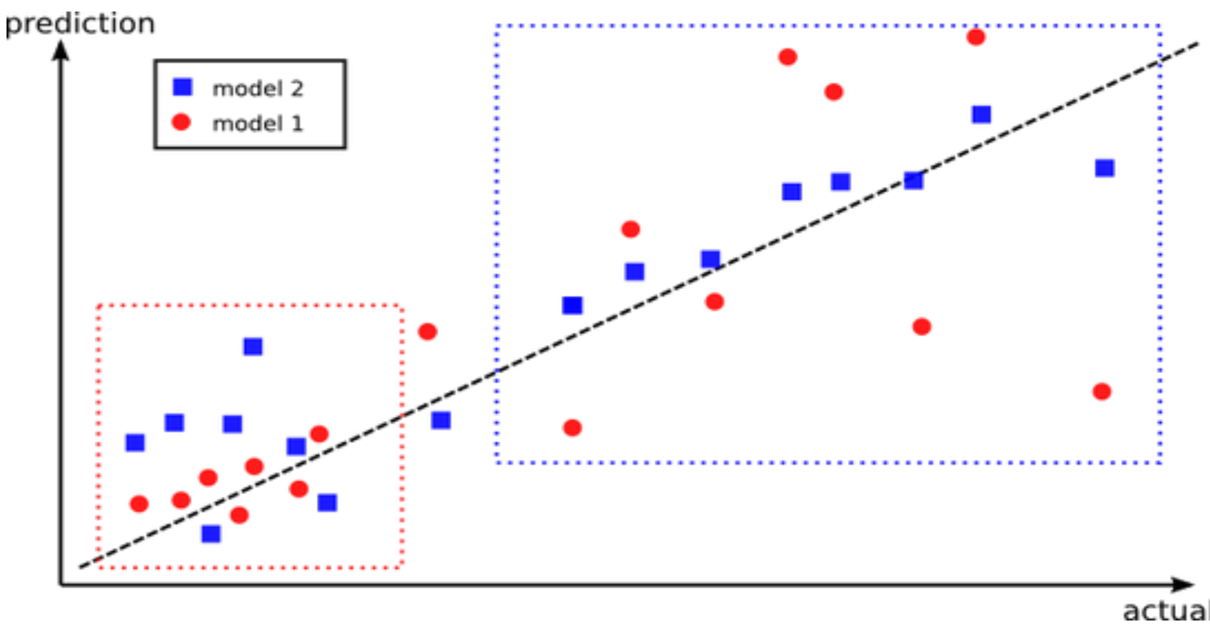
WHY DOES STACKING WORK?



Case 1: In the case of simple averaging (or learned simple models), the weighted average model de-noises individual models that are potentially overfit around the decision boundaries

Image source: https://mlwave.com/kaggle-ensembling-guide/

WHY DOES STACKING WORK?



Model 1 performs better here ***Model 2 performs better here***

Case 2: Different models perform better on different parts of the input space X. I.e., more complex models will do better on regions of X with higher support. More biased models may then be better where there is lower support. Stacking learns the best of both worlds.

Image source: https://www.kdnuggets.com/2017/02/stacking-models-imropved-predictions.html