Berkeley Engineering | BerkeleyHaas

PROFESSIONAL CERTIFICATE IN MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Office Hour #21 with Matilde D’Amelio

August 18, 2022 at 9 pm UTC

Ensemble Techniques

Ensemble models is a machine learning approach to combine multiple other models in the prediction process.

Those models are referred to as **base estimators**. It is a solution to overcome the following technical challenges of building a single estimator:

- **High variance:** The model is very sensitive to the provided inputs to the learned features.
- **Low accuracy:** One model or one algorithm to fit the entire training data might not be good enough to meet expectations.
- **Features noise and bias:** The model relies heavily on one or a few features while making a prediction.

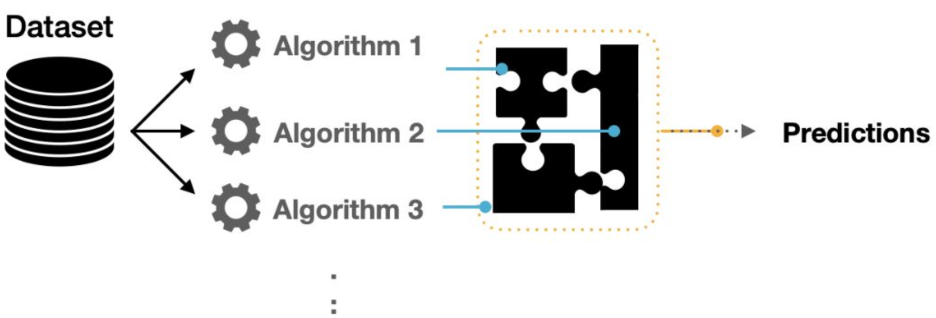


Figure 1: Diversifying the model predictions using multiple algorithms.

Ensemble Techniques

Building ensemble models is not only focused on the variance of the algorithm used.

For instance, we could build multiple ensemble models where each model is learning a specific pattern specialized in predicting one aspect. Those models are called **weak learners** that can be used to obtain a meta-model. In this architecture of ensemble learners, the inputs are passed to each weak learner while collecting their predictions. The combined prediction can be used to build a final ensemble model.

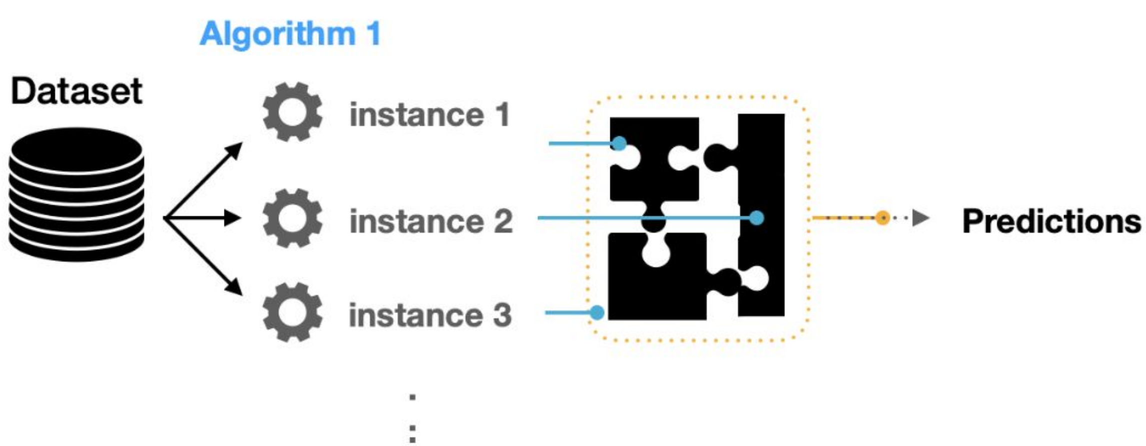
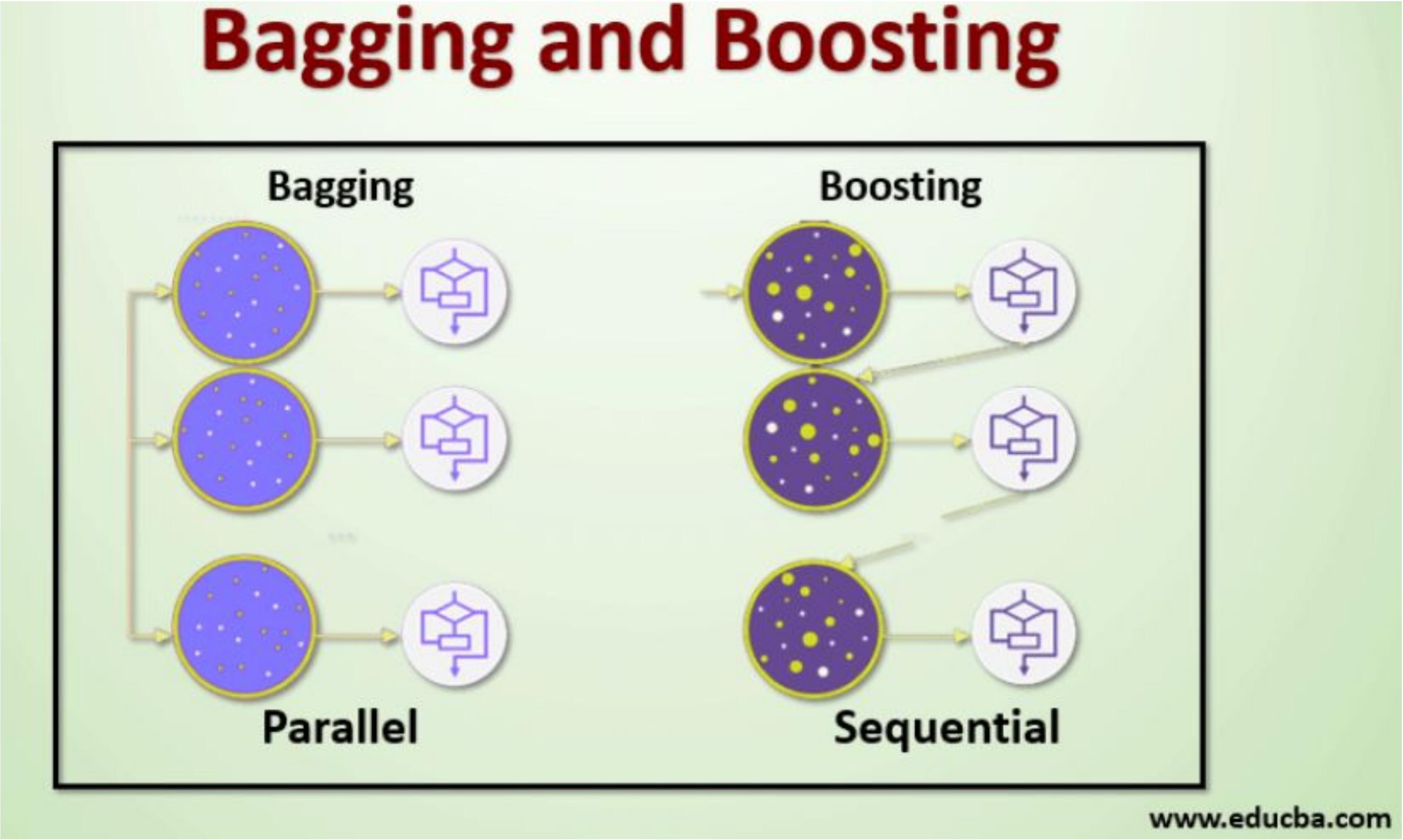


Figure 2: Aggregated predictions using multiple weak learners of the same algorithm.

Berkeley Engineering | BerkeleyHaas

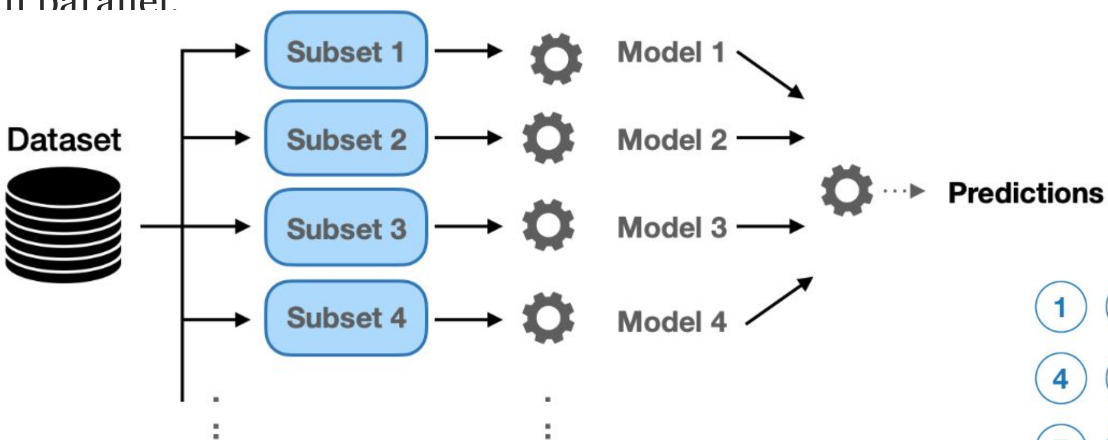
Ensemble Techniques



Bagging

The idea of bagging is based on making the training data available to an **iterative process of learning.** Each model learns the error produced by the previous model using a slightly different subset of the training dataset.
Bagging reduces variance and minimizes overfitting.

Bagging is based on a **bootstrapping sampling technique.** Bootstrapping creates multiple sets of the original training data with replacement. Replacement enables the duplication of sample instances in a set. Each subset has the same equal size and can be used to train models in parallel.



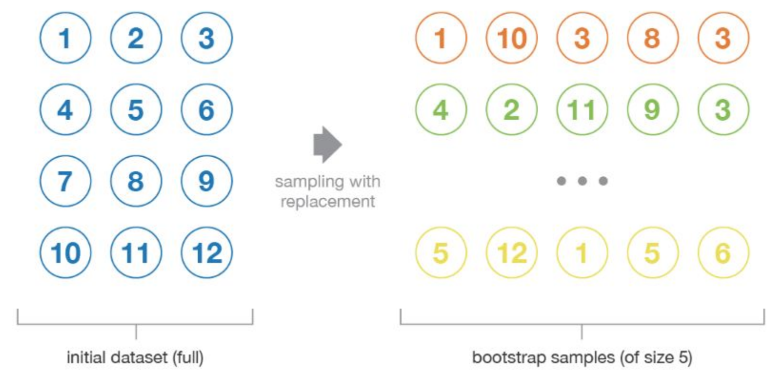
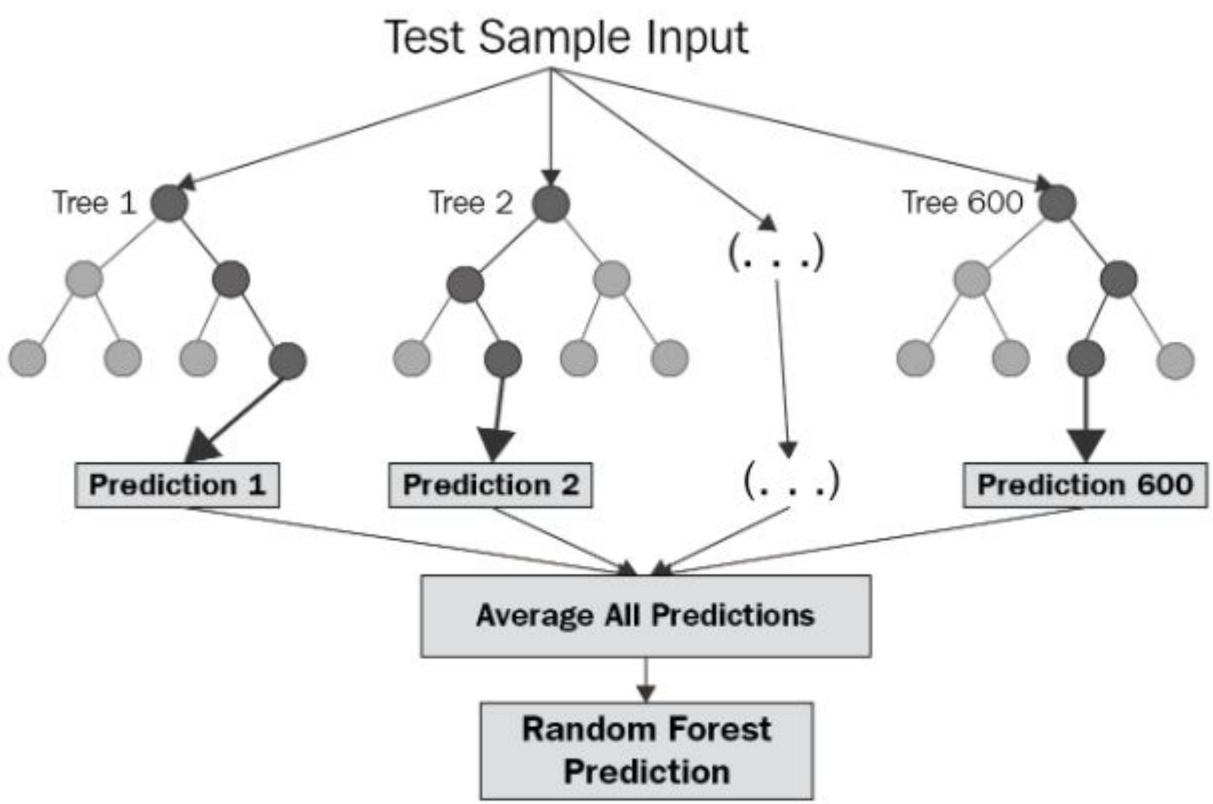


Illustration of the bootstrapping process.

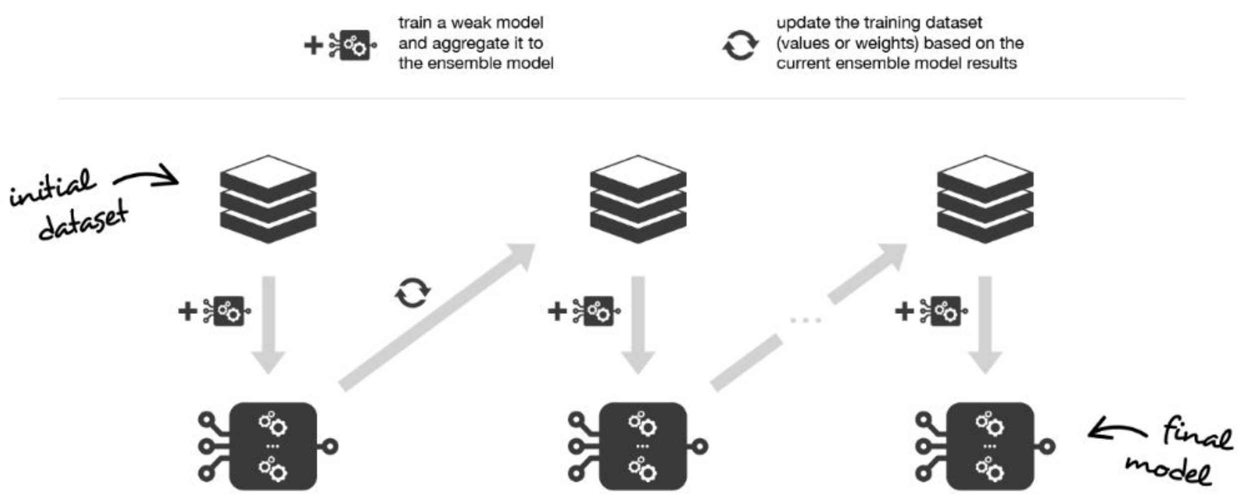
Random Forests (Sample of Bagging)

Uses subset of training samples as well as subset of features to build multiple split trees. Multiple decision trees are built to fit each training set. The distribution of samples/features is typically implemented in a random mode.



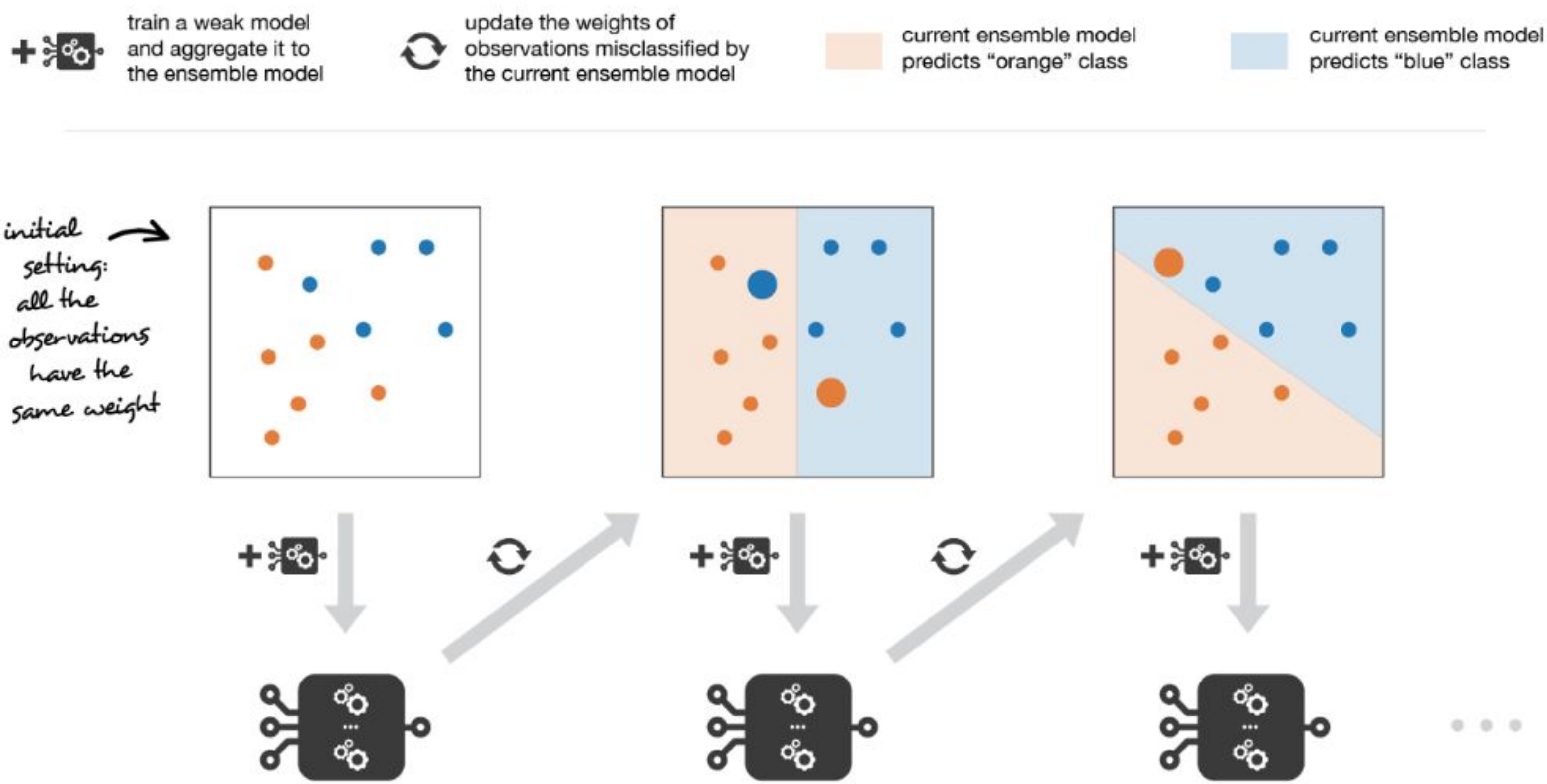
Boosting

Boosting: we build a family of models that are aggregated to obtain a strong learner that performs better.
However, unlike bagging that mainly aims at reducing variance, boosting is a technique that consists in fitting sequentially multiple weak learners in a **very adaptative way**: each model in the sequence is fitted giving more importance to observations in the dataset that were badly handled by the previous models in the sequence. Intuitively, each new model **focus its efforts on the most difficult observations** to fit up to now, so that we obtain, at the end of the process, a strong learner with lower bias



Boosting consists in, iteratively, fitting a weak learner, aggregate it to the ensemble model and "update" the training dataset to better take into account the strengths and weakness of the current ensemble model when fitting the next base model.

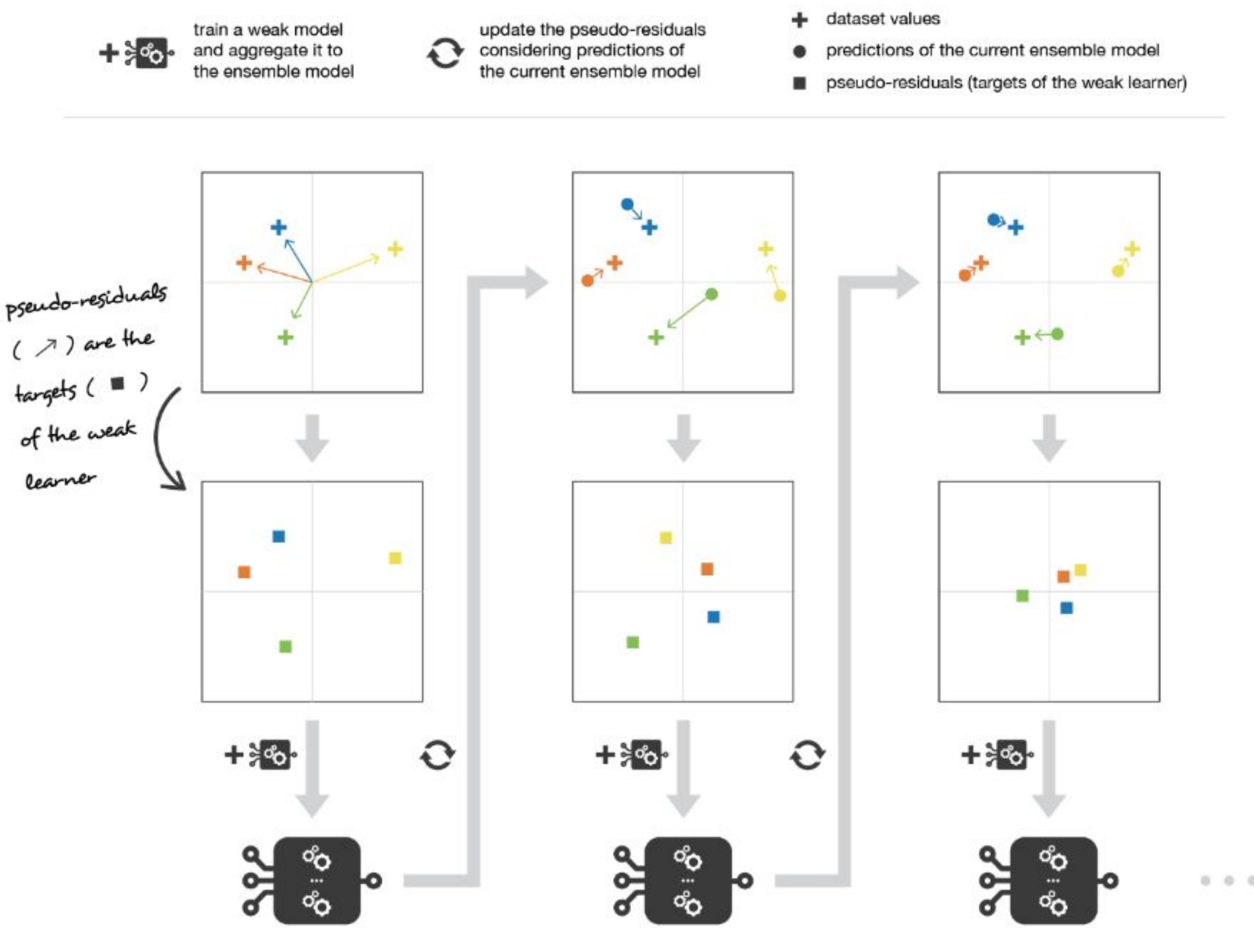
Adaptive Boosting



Adaboost updates weights of the observations at each iteration. Weights of well classified observations decrease relatively to weights of misclassified observations. Models that perform better have higher weights in the final ensemble model.

Berkeley Engineering | BerkeleyHaas

Gradient Boosting



Gradient boosting updates values of the observations at each iteration. Weak learners are trained to fit the pseudo-residuals that indicate in which direction to correct the current ensemble model predictions to lower the error.

Aggregating Predictors

When we ensemble multiple algorithms to adapt the prediction process to combine multiple models, we need an aggregating method. Three main techniques can be used:

- **Max Voting:** The final prediction in this technique is made based on majority voting for classification problems.
- **Averaging:** Typically used for regression problems where predictions are averaged. The probability can be used as well, for instance, in averaging the final classification.
- **Weighted Average:** Sometimes, we need to give weights to some models/algorithms when producing the final predictions.

Berkeley Engineering | BerkeleyHaas

QUESTIONS?