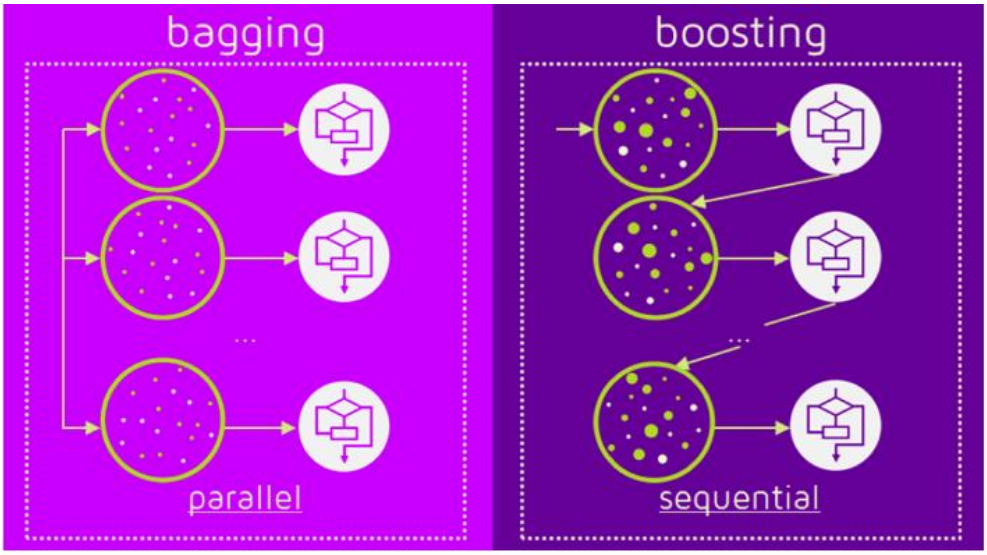
Ensemble methods: Bagging, Boosting and Stacking

Ensemble is a Machine Learning concept in which the idea is to train **multiple models** using the same learning algorithm.

The ensembles take part in a bigger group of methods, called **metaclassifiers,** where a set of hundreds or thousands of learners with a common objective are fused together to solve the problem.



The idea here is to train multiple models, each with the objective to predict or classify a set of results.

The two are the most important terms describing the ensemble (combination) of various models into one more effective model.

- **Bagging** to decrease the model's variance.
- **Boosting** to decreasing the model's bias.

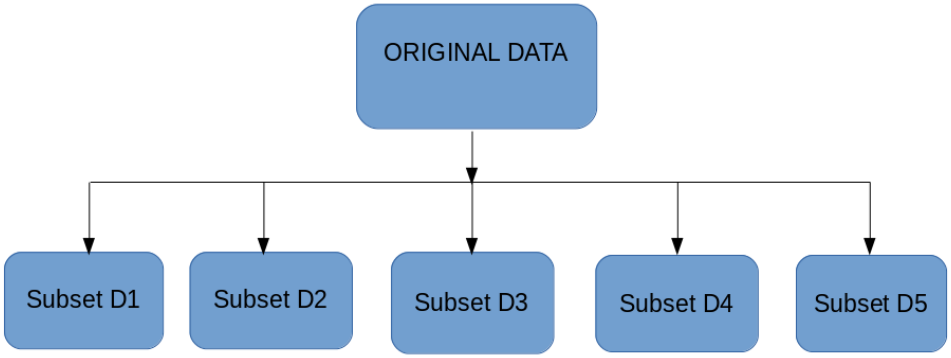
Most of the errors from a model's learning are from three main factors: variance, noise, and bias.

Bagging

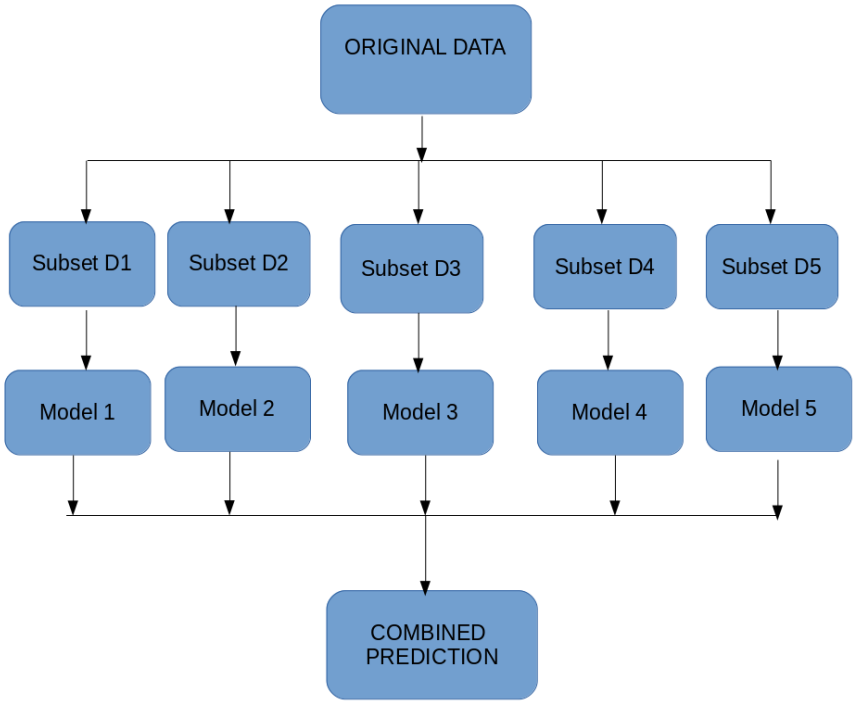
The idea behind bagging is combining the results of multiple models to get a generalized result.

Bootstrapping is a sampling technique in which we create subsets of observations from the original dataset, **with replacement**. The size of the subsets is the same as the size of the original set.

Bagging (or Bootstrap Aggregating) technique uses these subsets (bags) to get a fair idea of the distribution (complete set). The size of subsets created for bagging may be less than the original set.



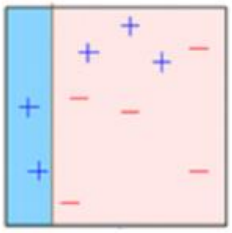
1. Multiple subsets are created from the original dataset, selecting observations with replacement.
2. A base model (weak model) is created on each of these subsets.
3. The models run in parallel and are independent of each other.
4. The final predictions are determined by combining the predictions from all the models.



Boosting is a sequential process, where each subsequent model attempts to correct the errors of the previous model. The succeeding models are dependent on the previous model.

boosting works in the below steps.

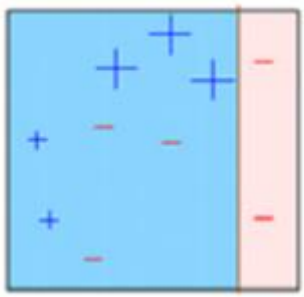
1. A subset is created from the original dataset.
2. Initially, all data points are given equal weights.
3. A base model is created on this subset.
4. This model is used to make predictions on the whole dataset.



5. Errors are calculated using the actual values and predicted values.

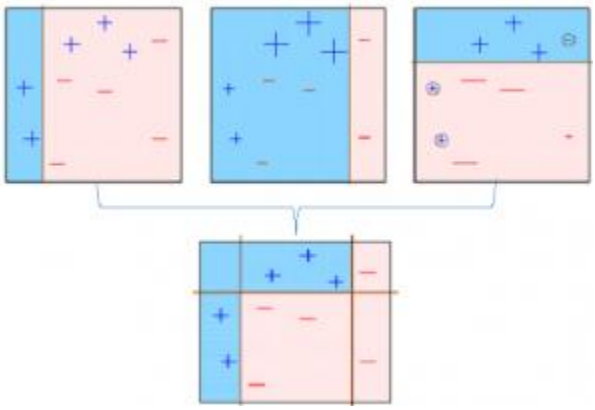
6. The observations which are incorrectly predicted, are given higher weights.

7. Another model is created, and predictions are made on the dataset. (This model tries to correct the errors from the previous model)



8. Similarly, multiple models are created, each correcting the errors of the previous model.

9. The final model (strong learner) is the weighted mean of all the models (weak learners).



Thus, the boosting algorithm combines several weak learners to form a strong learner. The individual models would not perform well on the entire dataset, but they work well for some part of the dataset. Thus, each model boosts the performance of the ensemble.

Bagging and Boosting decrease the variance of your single estimate as they combine several estimates from different models. So, the result may be a model with **higher stability**.

If the problem is that the single model gets a very low performance, Bagging will rarely get a **better bias**. However, Boosting could generate a combined model with lower errors as it optimizes the advantages and reduces the pitfalls of the single model.

Stacking

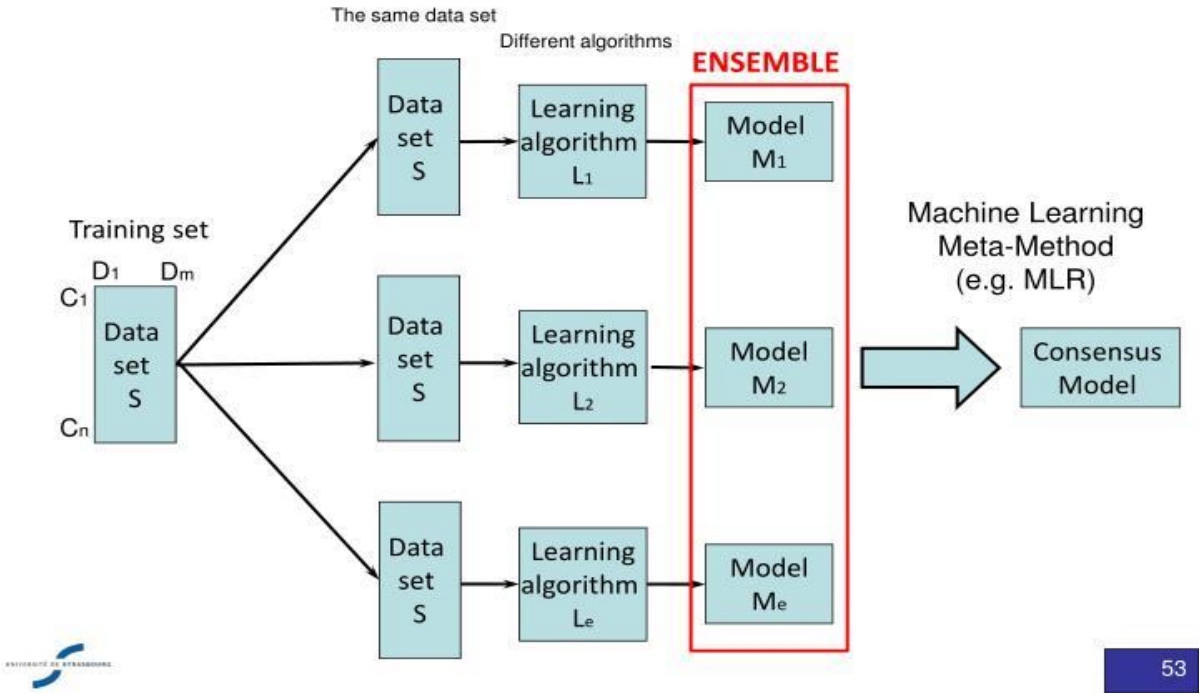
Model averaging is an ensemble technique where multiple sub-models contribute equally to a combined prediction.

Model averaging can be improved by weighting the contributions of each sub-model to the combined prediction by the expected performance of the submodel.

This can be extended further by training an entirely new model to learn how to best combine the contributions from each submodel.

This approach is called stacked generalization, or stacking for short, and can result in better predictive performance than any single contributing model.

Stacking



Common Types of Ensemble Methods

Bagging

- Reduces variance and increases accuracy
- Robust against outliers or noisy data
- Often used with Decision Trees (i.e. Random Forest)

Boosting

- Also reduces variance and increases accuracy
- Not robust against outliers or noisy data
- Flexible – can be used with any loss function

Stacking

- Used to ensemble a diverse group of strong learners
- Involves training a second-level machine learning algorithm called a "metalearner" to learn the optimal combination of the base learners