Ensemble Techiques

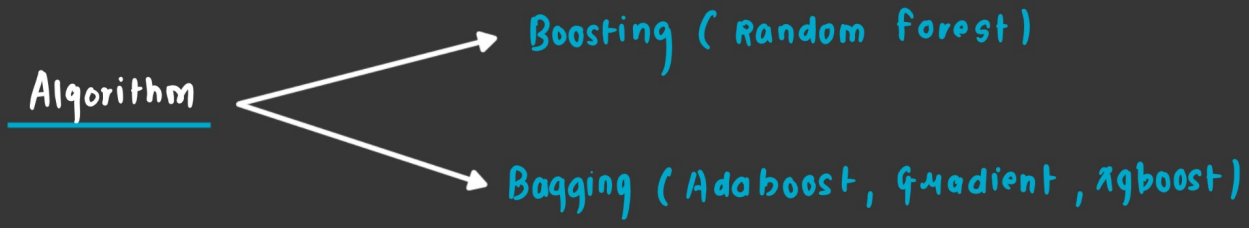
Ensembled technique in ml combines multiple models to improve the overall performance.

Combine weak learners together to make stronger learners.

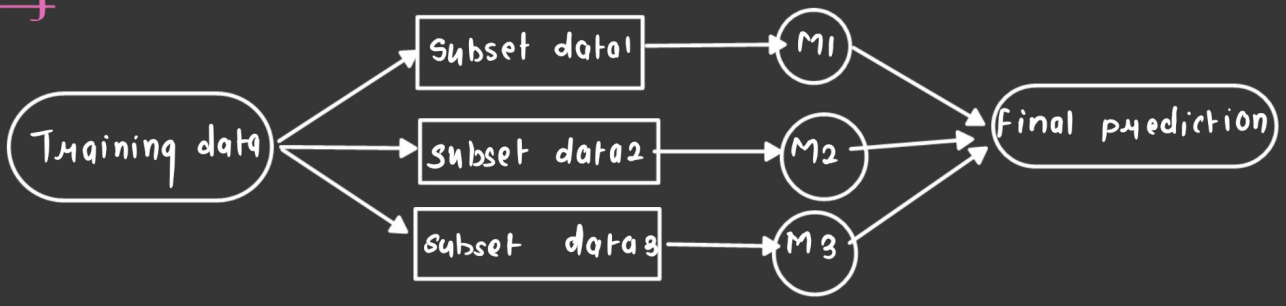
Why use ensemble techniques ?

- > Improve accuracy by aggregating prediction from multiple model
- > Reduction in overfitting.
- > Robustness (Redces the risk of poor accuracy).

Common Algorithms in ensemble techniques



Bagging



Bagging also known as bootstrap aggregation

Models train with different subset of training dataset parallely and independently. then at last we do aggregation and voting for classification.

Algorithm Random forest algorithm

Math behind bagging

Assume dataset ⟶ D

multiple subsets of data ⟶ $D_1, D_2, D_3, \ldots, D_n$

train model ⟶ $M_1, M_2, M_3, \ldots, M_n$

new input ⟶ x,

Prediction ⟶ $\hat{y}$ (Avg of prediction $\hat{y}_1, \hat{y}_2 \ldots$)

$$\hat{y} = \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i$$

Boosting

Model train sequentially, each trying to correct error of the previous one.

(weighted based model)

Algorithm Adaboost, Gradient descent, XG Boost

Math behind Boosting

Initially all sample are given equal weight and M_1 model trained on this weighted data.

Most of data missclassified by M_1 model.

Now M_2 is trained with re-weighted data.

$$\hat{y} = \left(\sum_{i=1}^{n} a_i \cdot \hat{y}_i \right)$$

a_i ⟶ weight