Intermediate Supervised Learnig

by : Dhea Fajriati Anas
Data Fellowship Batch 6

Table of contents

01

Ensemble Learning

What is Ensemble Learning ?

- **Ensemble learning** is a machine learning paradigm where multiple models (often called "weak learners") are trained to solve the same problem and combined to get better results.
- **The main hypothesis** is that when weak models are correctly combined we can obtain more accurate and/or robust models.
- **"Unity is strength".** This old saying expresses pretty well the underlying idea that rules the very powerful "ensemble methods" in machine learning
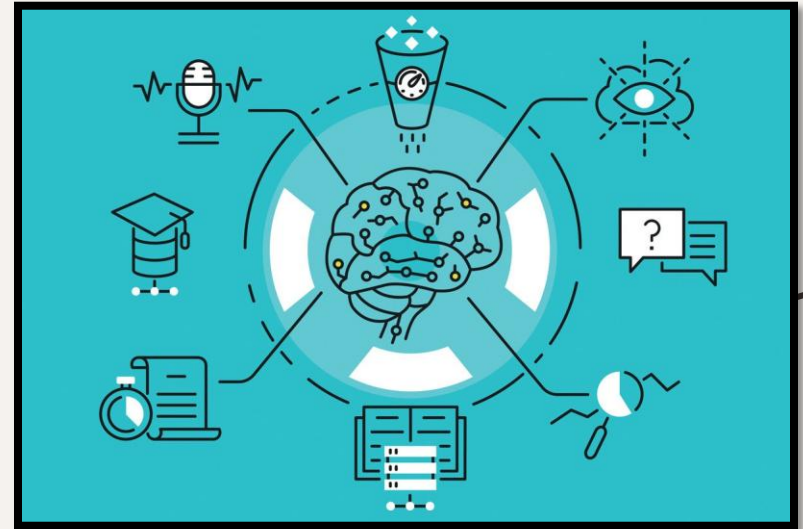
Type of Ensemble Learning

01

Bagging

## 02

Boosting

## 03

Stacking

02

Bagging

What is Bagging?

- **Bagging or Bootstrap Aggregation** is that often considers homogeneous weak learners. Learns them independently from each other in parallel and combines them following some kind of deterministic averaging process.
- **For a regression problem,** the outputs of individual models can literally be averaged to obtain the output of the ensemble model.
- **For classification problem** the class outputted by each model can be seen as a vote and the class that receives the majority of the votes is returned by the ensemble model (this is called **hard-voting)** or the probabilities of each classes returned by all the models, average these probabilities and keep the class with the highest average probability (this is called **soft-voting**).
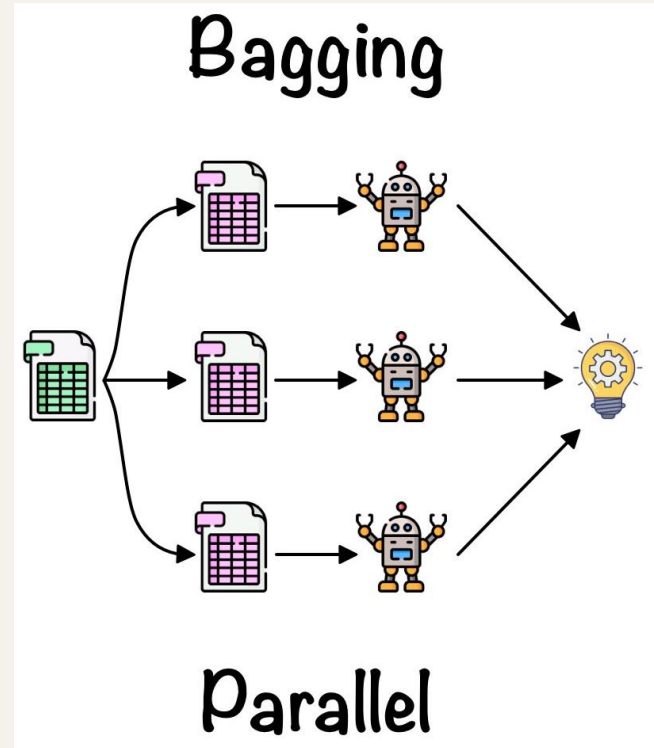
Bagging Algorithms

- **Random Forest** is another ensemble machine learning algorithm that follows the bagging technique. The base estimators in random forest **are decision trees**.
- The **random forest** approach is a bagging method where **deep trees**, fitted on bootstrap samples, are combined to produce an output with lower variance.
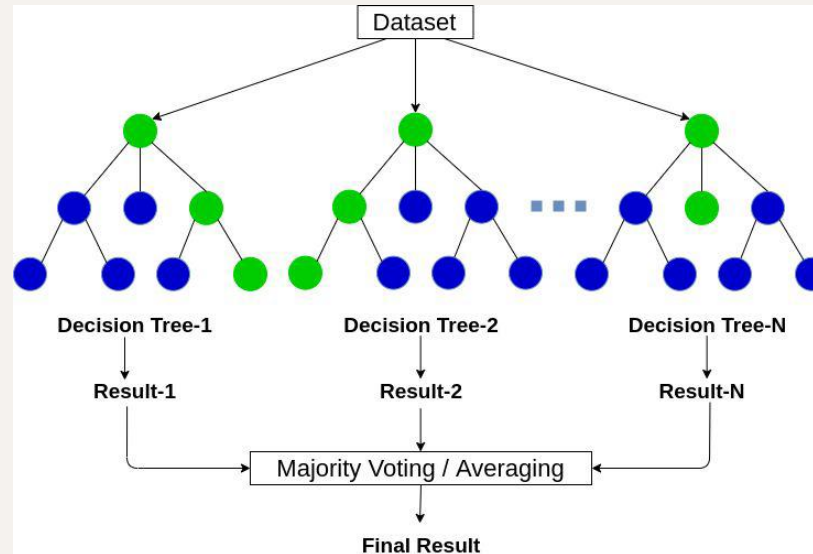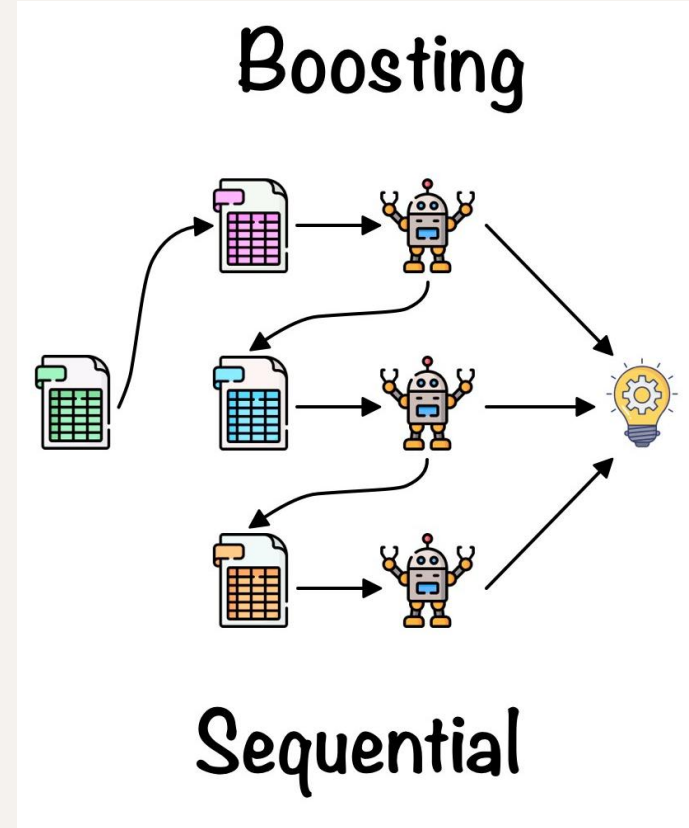
03

Boosting

What is Boosting?

- **Boosting**, that often considers homogeneous weak learners, learns them sequentially in a very adaptative way (a base model depends on the previous ones) and combines them following a deterministic strategy.
- Boosting, like bagging, can be used for **regression and classification problems.**

Example of Boosting Algorithm

Gradient Boosting

Train multiple models sequentially, and for each new model, the model gradually minimizes the loss function using the Gradient Descent method

AdaBoost

Adaptive Boosting as the weights are re-assigned to each instance, with higher weights assigned to incorrectly classified instances. Boosting is used to reduce bias as well as variance for supervised learning.

XGBoost

An improvised version of the gradient boosting algorithm, and the working procedure of both is almost the same.
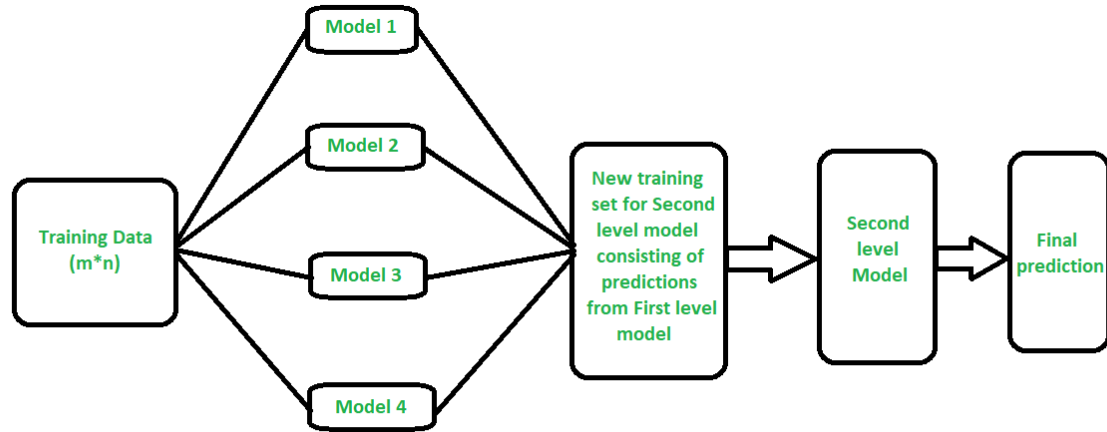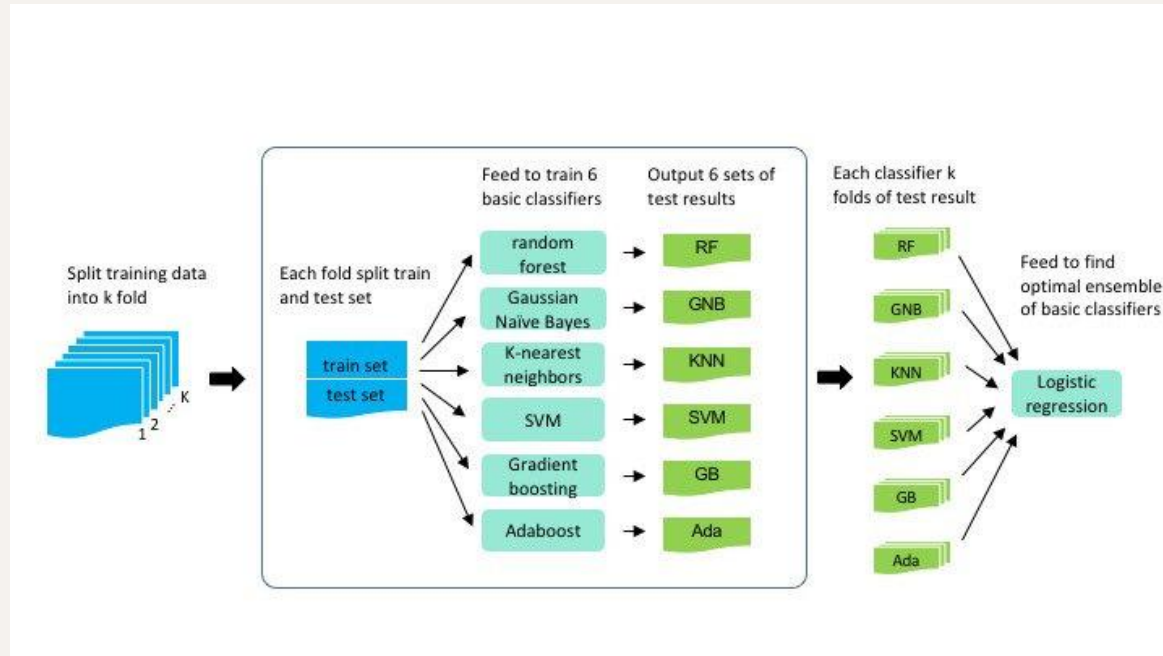
04

Stacking

What is Stacking?

- **Stacking or Stacked Generalization** is a way to ensemble multiple classifications or regression model. There are many ways to ensemble models, the widely known models are **Bagging** or **Boosting**.

Example of Stacking

Thank You !