Ensemble Learning

Before Discussing popular **ensemble** methods such as Bagging, Boosting, let's explain and cover the concept of **Ensemble Learning**.

It's simply combining multiple individual models to produce better predictions than a single model, the group of models used is called **weak learners** will work together to make better predictions than any single model called **strong learner**

The Bias-Variance problem

It's the main reason behind using ensemble methods is to approximate the perfect balance between the bias and variance trade-off, let's quickly define **bias** and **variance** :

- **Bias** is the error that comes from wrong assumptions in the model, **high Bias** models are too simple and miss the actual pattern in the data
- **Variance** is model's sensitivity to changes in the data, **high variance** models are too complex which results in capturing useless noise and fluctuations in the data that often leads to Overfitting

Types of ensemble learning methods

Based on model composition

1. **Homogeneous ensembles** All the **weak learners** are the same type models but trained on different subsets of data or with different parameters such as Random Forests
2. **Heterogeneous ensembles** the base models are of different types for example combining Decision Trees and Logistic Regression and SVM

Based on the learning approach

1. **Parallel methods** Training all base models independently and combine their predictions at the end
2. **Sequential methods** Train models one after another with each new model focusing on the mistakes made by the previous models

Advantages of ensemble learning

- Improve accuracy
- Reduce overfitting by combining multiple models which reduce the high variance
- Robustness to outliers and noisy data, gets canceled by weak learners working together
- Stability

Limitations of ensemble learning

- Increased complexity
- Losing interpretibility
- Higher computational cost (modern computing can handle it)
- Storage requirements(model deployment problem)