ENSEMBLING LEARNING

P SANJUTH REDDY
20BD1A6647
AIML - A
KESHAV MEMORIAL INSTITUTE OF TECHNOLOGY

CONTENT

- Bias Variance Tradeoff
- Ensemble Learning? What and why
- Types Of Ensembling Techniques
- *Bagging*
- *Boosting*
- *Stack Generalization*
- Applications of Ensembling

BIAS VS VARIANCE

What is bias?

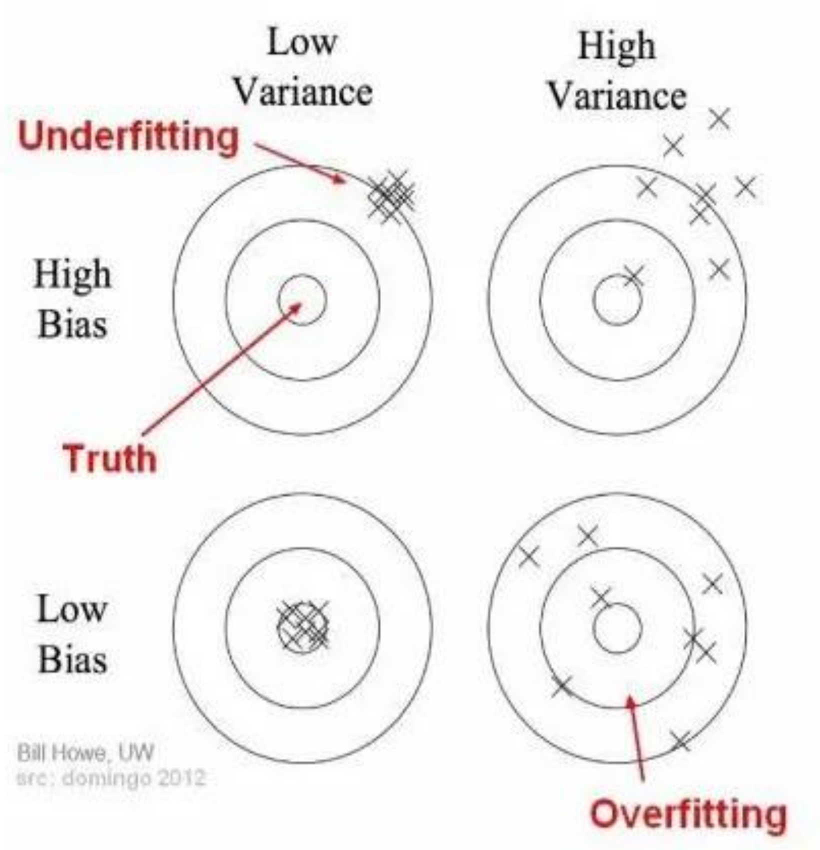
Bias is the **difference between the average prediction of our model and the correct value** which we are trying to predict. Model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on training and test data.

High bias => Underfitting

What is variance?

Variance is the **variability of model prediction for a given data point or a value** which tells us spread of our data. Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn't seen before. As a result, such models perform very well on training data but has high error rates on test data.

High variance => Overfitting


Low Variance
High Variance
Underfitting
High Bias
Truth
Low Bias
Bill Howe, UW
src: domingo 2012
Overfitting

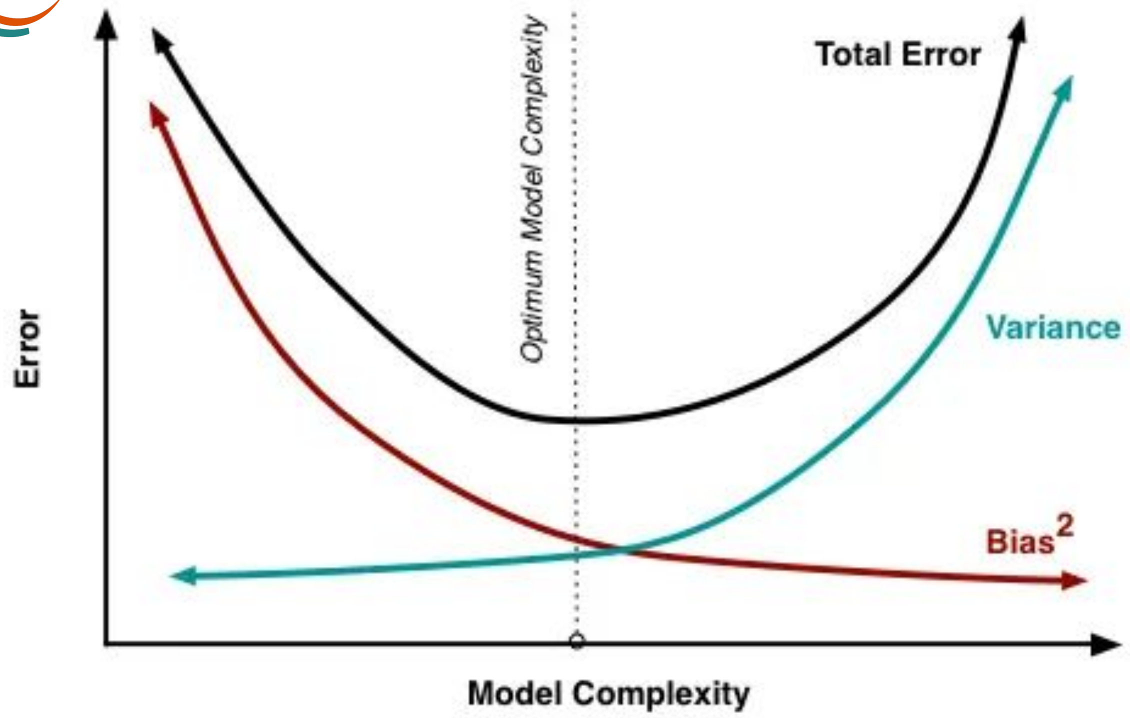
Why is there Bias Variance Tradeoff?

- Model complexity determines the bias-variance tradeoff.
- Simple models have high bias and low variance while
- Complex models have high variance and low bias.
- Striking the right balance is crucial to avoid underfitting and overfitting.

Total Error

To build a good model, we need to find a good balance between bias and variance such that it minimizes the total error.

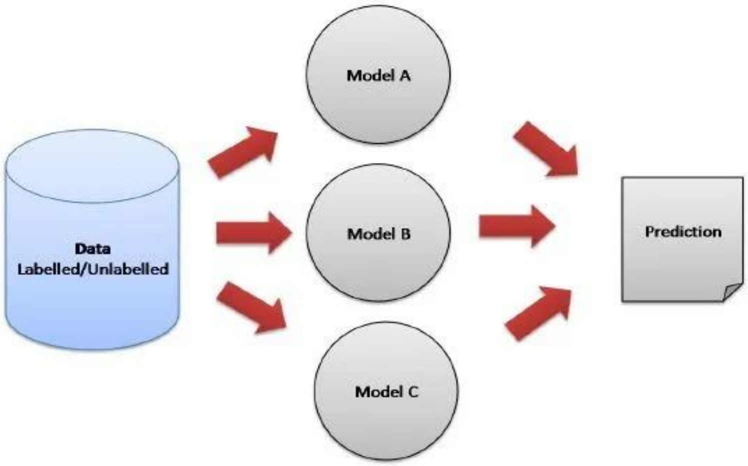
$$\text{Total Error} = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$

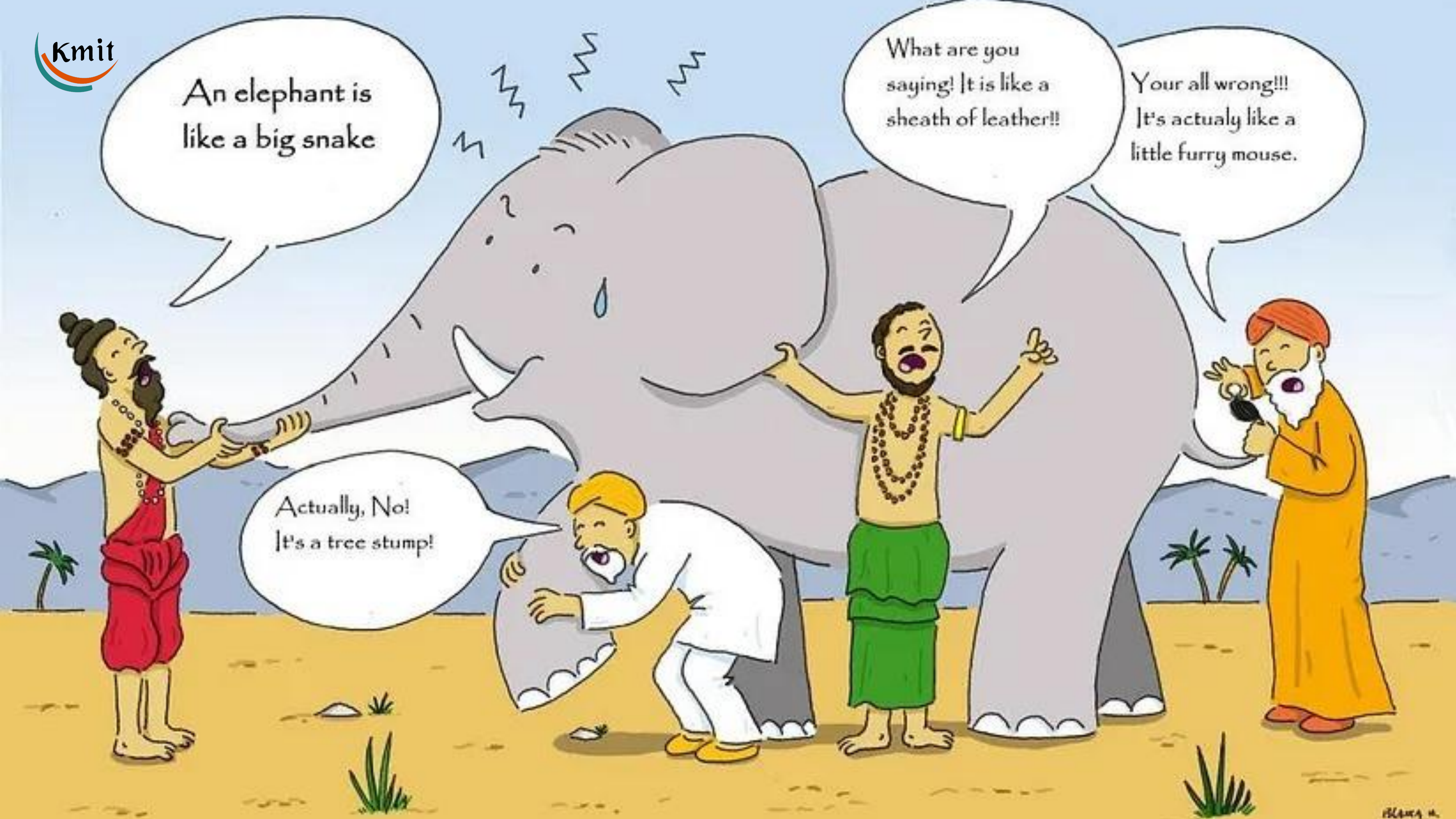


We need to find a good balance between bias and variance such that it minimizes the total error.

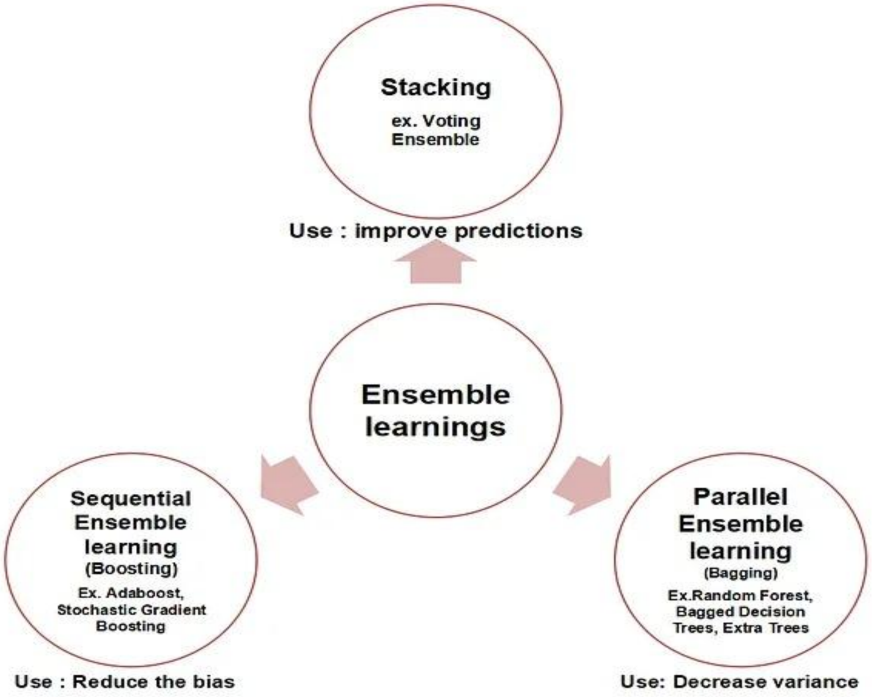
What is Ensemble Learning?

- Ensemble methods combine multiple Machine learning algorithms to improve predictive performance.
- The main principle is to leverage a group of weak learners to create a strong learner, increasing model accuracy.
- Ensemble helps reduce variance and bias, the main causes of differences between actual and predicted values.
- It cannot eliminate noise, which is considered irreducible error in predictions.




An elephant is like a big snake
Actually, No! It's a tree stump!
What are you saying! It is like a sheath of leather!!
Your all wrong!!! It's actualy like a little furry mouse.

TYPES OF ENSEMBLING



BAGGING

Bagging (**B**ootstrap **AGG**regat**ING**) Technique:

- Bagging is an ensemble learning technique used to improve the accuracy and robustness of machine learning models.
- It involves creating multiple copies of a base model, training each copy on different subsets of the original data using bootstrap sampling.
- The final prediction is obtained by averaging (in the case of regression) or voting (in the case of classification) the predictions from each individual model.

Examples:

Random Forest:

- Random Forest is a popular bagging-based ensemble learning method that uses decision trees as the base model.
- It is widely used in classification and regression tasks..

Bagged Ensemble of Neural Networks:

- Bagging can be applied to neural networks, training each network on different subsets of the data and then combining their predictions for improved generalization.

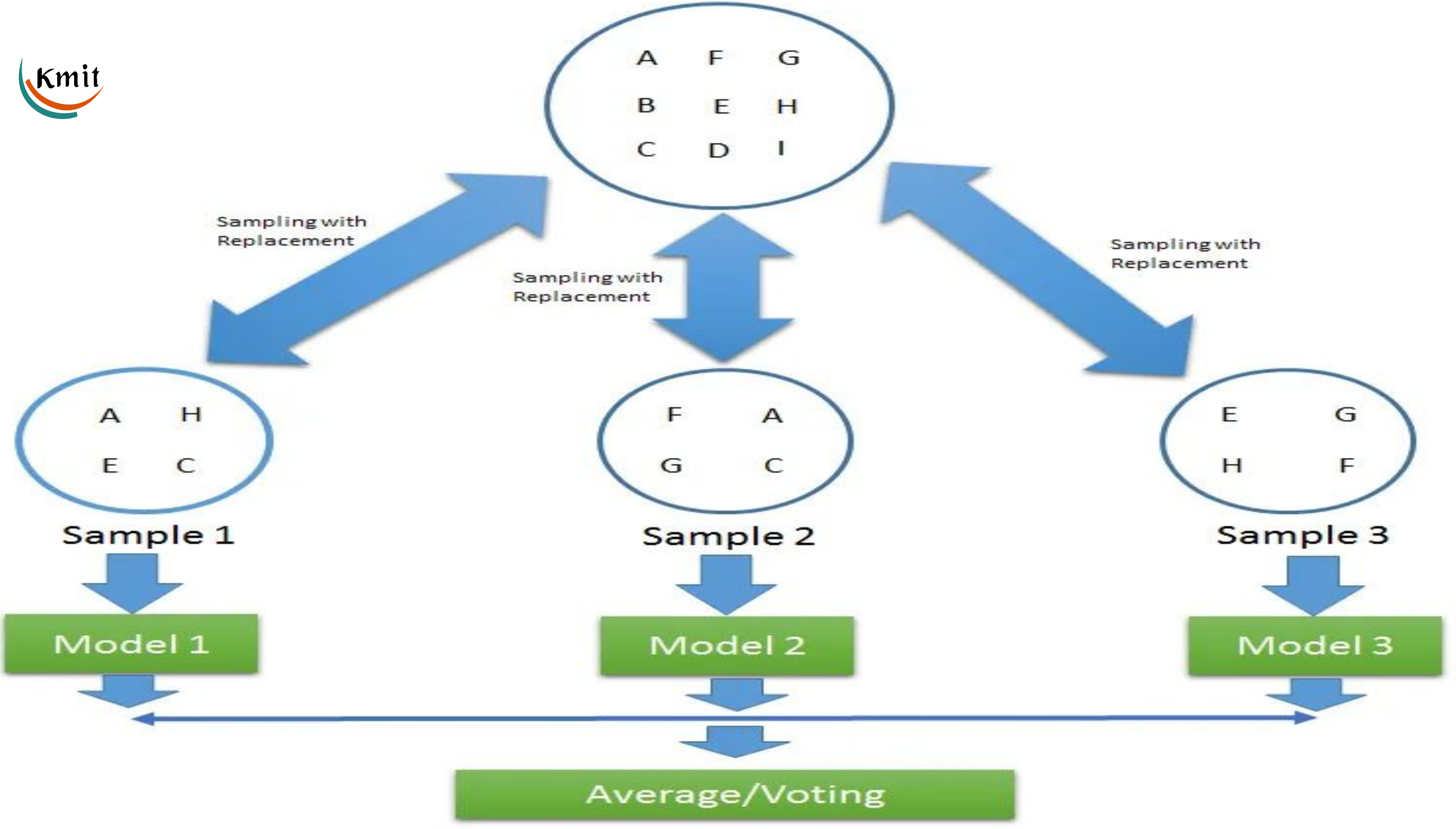
How Bagging Works:

1) Bootstrap Sampling:

- Bagging relies on bootstrap sampling, where random samples of the same size as the original dataset are drawn with replacement.
- This process creates diverse subsets that may contain duplicate instances and leave out some original instances, ensuring variability in the training sets.

2) Parallel Training and Prediction:

- Each base model is trained independently on its corresponding bootstrap sample in parallel, allowing for efficient use of computational resources.
- During prediction, all individual models produce their outputs, and the final prediction is aggregated based on the ensemble method (e.g., averaging for regression or voting for classification).


A F G
B E H
C D I
Sampling with Replacement
Sampling with Replacement
Sampling with Replacement
A H
E C
F A
G C
E G
H F
Sample 1
Sample 2
Sample 3
Model 1
Model 2
Model 3
Average/Voting

Advantages and Benefits:

Reduction of Variance:

- Bagging reduces model variance by combining multiple models, which helps to stabilize predictions and reduce overfitting.
- The averaging or voting process smooths out the noise and inconsistencies present in individual models.

Improved Predictive Performance:

- Bagging often leads to improved predictive accuracy compared to a single base model, especially when the base model is unstable or sensitive to the training data.

Robustness to Outliers and Noisy Data:

- By training on multiple diverse subsets, bagging is less susceptible to outliers and noisy data, making it more robust in real-world scenarios.

Parallelizability:

- Bagging allows for parallel training and prediction, making it well-suited for large datasets and distributed computing environments.

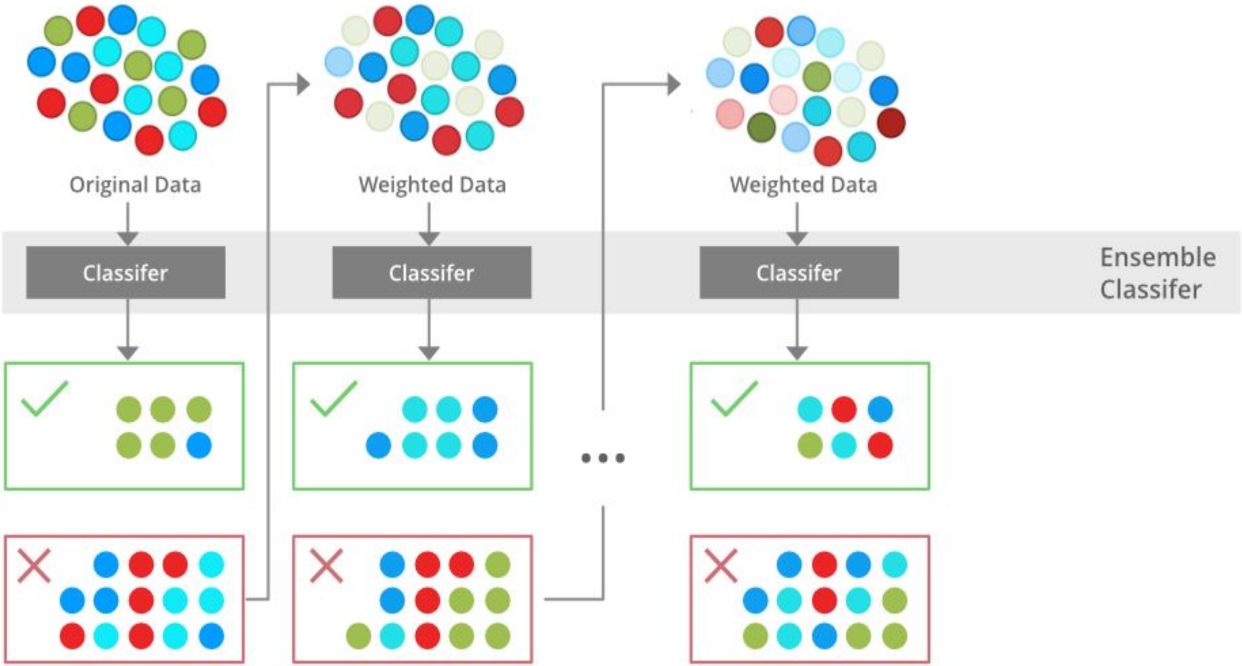
BOOSTING

Boosting Technique in Ensemble Learning:

- Boosting is an ensemble learning technique that aims to improve the performance of weak learners by combining them into a strong learner.
- Unlike bagging, which focuses on reducing variance, boosting focuses on reducing bias and improving model accuracy.

Iterative Training Process:

- Boosting involves iteratively training weak learners in a **sequential manner.**
- Each weak learner is trained to correct the errors made by the previous models, giving more weight to the misclassified instances.


Original Data
Weighted Data
Weighted Data
Classifer
Classifer
Classifer
Ensemble Classifer

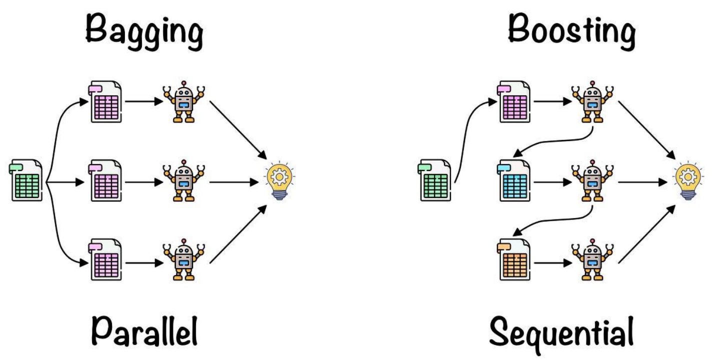
How Boosting Works:

Weighted Training Data:

- Initially, all data points are assigned equal weights.
- During each iteration, the misclassified data points from the previous model are given higher weights to emphasize their importance in the next training.

Sequential Model Training:

- Weak learners (e.g., decision trees with limited depth) are trained one after another, with each model focusing on the difficult-to-classify instances.
- The weak learners are combined into a strong learner through weighted majority voting or weighted averaging.



Advantages and Benefits:

Improved Model Accuracy:

- Boosting is known for its ability to significantly improve model accuracy by reducing bias and handling complex relationships in the data.

Adaptive Learning:

- The iterative nature of boosting allows the model to adapt to difficult instances, giving more attention to previously misclassified data points.

Reduced Underfitting (bias):

- Boosting is less prone to Underfitting than individual strong learners, as the weak learners are constrained in complexity and combined to create a robust model.

State-of-the-Art Performance:

- Boosting algorithms like AdaBoost and Gradient Boosting Machines (GBM) have consistently achieved top performance in various machine learning competitions and real-world applications.

Examples and Use Cases:

AdaBoost (Adaptive Boosting):

- AdaBoost is one of the earliest and most popular boosting algorithms used for binary classification tasks.
- It assigns higher weights to misclassified instances in each iteration, allowing subsequent models to focus on correcting those mistakes.

Gradient Boosting Machines (GBM):

- GBM is a widely used boosting algorithm that builds weak learners sequentially, optimizing the model using gradient descent methods.
- It is known for its high predictive accuracy and ability to handle large-scale datasets.

XGBoost:

- XGBoost is an optimized version of gradient boosting that leverages advanced regularization techniques and parallel computation for even better performance.

STACKING

Introduction to Stacking:

Stacking in Ensemble Learning:

- Stacking, also known as Stacked Generalization, is an advanced ensemble learning technique that combines the predictions of multiple models using a meta-model.
- Unlike bagging and boosting, which focus on combining multiple models at the same level, stacking involves a two-level architecture: base models and a meta-model.

Two-Level Architecture:

- Base Models: Several diverse base models are trained on the original data, each capturing different patterns and aspects of the problem.
- Meta-Model: The predictions from the base models become the input features for a higher-level model (the meta-model), which learns to make the final prediction.

How Stacking Works:

Base Model Training:

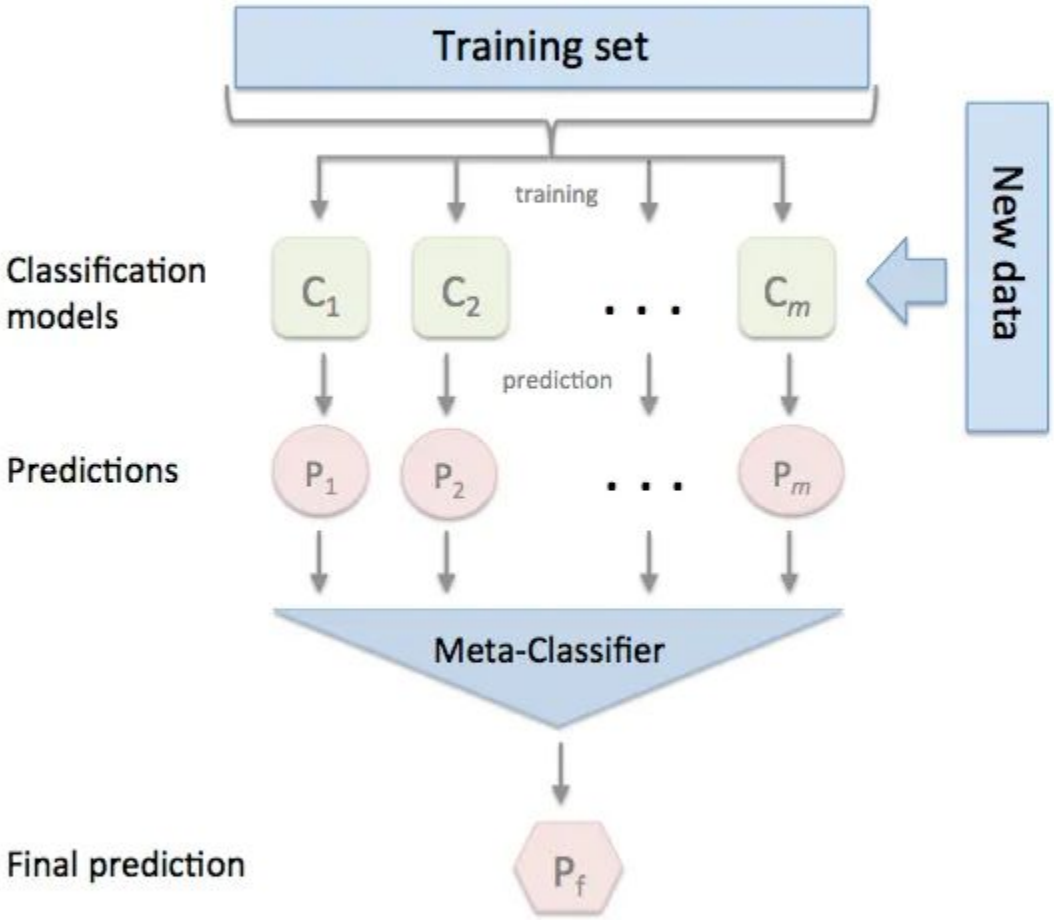
- Multiple base models of different types are trained on the training data. These can include decision trees, support vector machines, neural networks, etc.

Generating Meta-Features:

- The base models make predictions on the validation data and the original test data.
- The predictions from the base models are then used as meta-features, which become the input for the meta-model.

Training the Meta-Model:

- The meta-model (e.g., linear regression, random forest) is trained on the meta-features along with the actual target values from the validation data.
- The meta-model learns to combine the base models' predictions to make the final prediction.



Training set
training
New data
Classification models
C_1
C_2
. . .
C_m
prediction
Predictions
P_1
P_2
. . .
P_m
Meta-Classifier
Final prediction
P_f

Applications of Ensemble Learning in Real-World:

Medical Diagnosis:

- Ensemble learning is used for diagnosing medical conditions, where multiple models combine their predictions to provide more accurate and reliable diagnoses.

Credit Risk Assessment:

- In the financial industry, ensembles are employed to assess credit risk by combining models that analyze various credit-related factors.

Image and Speech Recognition:

- Ensembles are used in computer vision and speech recognition tasks to improve accuracy and handle complex patterns in visual and auditory data.

Fraud Detection:

- In the banking and e-commerce sectors, ensembling is used to detect fraudulent transactions by combining multiple anomaly detection models.

…….. ***And Many More***

Further Reading :

- https://en.wikipedia.org/wiki/Ensemble_learning
- https://en.wikipedia.org/wiki/Bagging
- https://medium.com/ml-research-lab/ensemble-learning-relation-with-bias-and-variance-431cdc0a3fc9


Thank
You