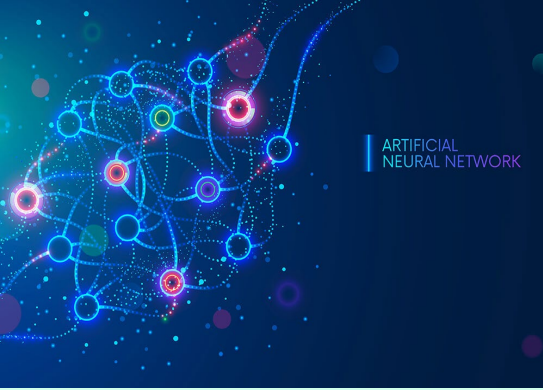




Ensemble Learning

Dr. Seena Joseph

Ensemble Learning

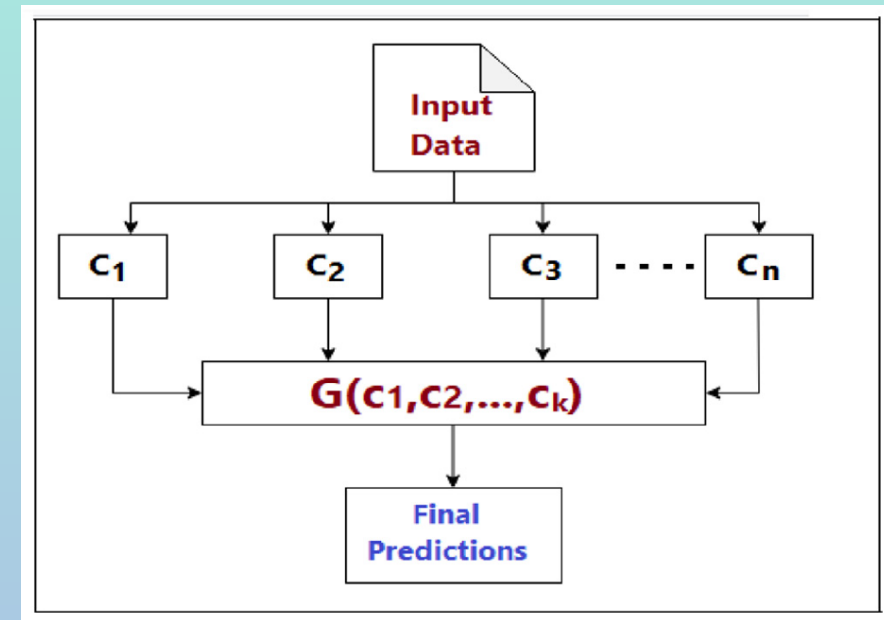
- Learning outcomes
 - Understanding the concept of ensemble learning
 - Application of ensemble learning

Ensemble Learning

- Ensemble learning
 - A learning methodology that combines several baseline models to build a bigger single yet more powerful model than its constituents.
 - Ensemble model can reduce the overfitting risks
 - Empirically, ensembles tend to yield better results when there is a significant diversity among the models

Ensemble Learning

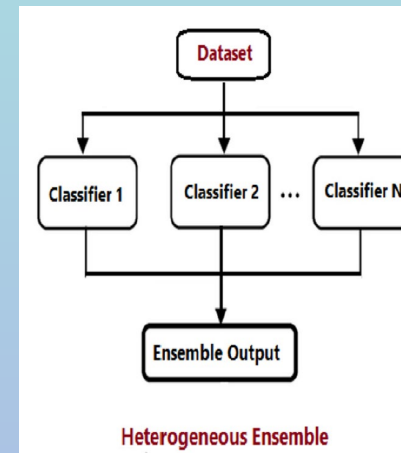
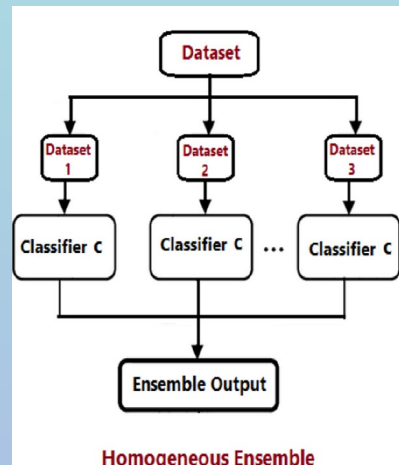
- The general framework of an ensemble learning system is to use an aggregate function, **G** to combine a set **n** of baseline classifiers towards predicting a single output.



Source: Mohammed et al., [1]

Ensemble Learning

- Homogeneous and Heterogeneous Models
- Homogeneous model
 - Same baseline classifier but with each classifier based on different data
 - Same feature selection method same for different training data
 - Easier to implement and less costly compared to heterogeneous models
- Heterogeneous model
 - Different baseline classifiers and each classifier is based on the same data
 - Feature selection method is different



Source: Mohammed et al., [1]

Ensemble Learning

Characteristics of the ensemble models

1. Dependency of baseline models (Parallel and Sequential)
2. Selection of fusion method
3. Heterogeneity of the baseline classifiers

Dependency

Parallel : Base models are trained independently; Base model learned from the same data but with different subset of features or samples; All base model make their prediction independently or simultaneously; Example : Bagging (Random forest); majority voting

Sequential:

Base models are trained sequentially; Each subsequent model learns from the mistakes or residual of the previous ones;

Example: Adaboost; Gradient Boosting; Extreme Gradient Boosting (XGBoost)

Ensemble Learning

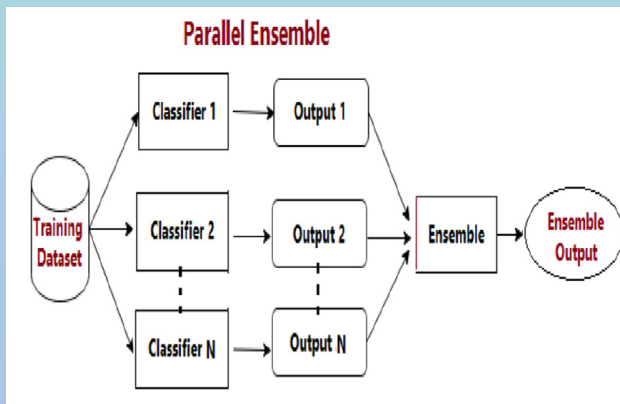
Dependency

Parallel : Base models are trained independently; Base model learned from the same data but with different subset of features or samples; All base model make their prediction independently or simultaneously; Example : Bagging (Random forest); majority voting

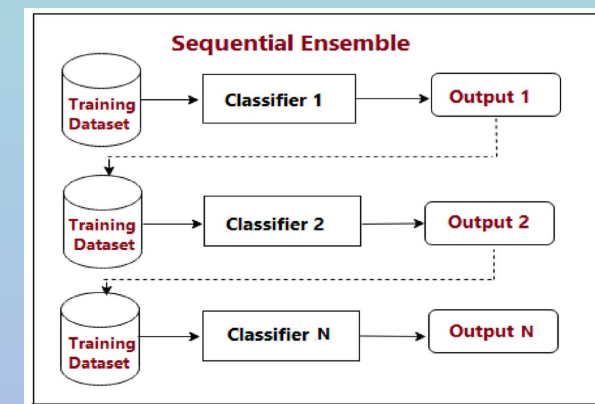
Sequential:

Base models are trained sequentially; Each subsequent model learns from the mistakes or residual of the previous ones;

Example: Adaboost; Gradient Boosting; Extreme Gradient Boosting (XGBoost)



Source: Mohammed et al., [1]



Ensemble Learning

Fusion method

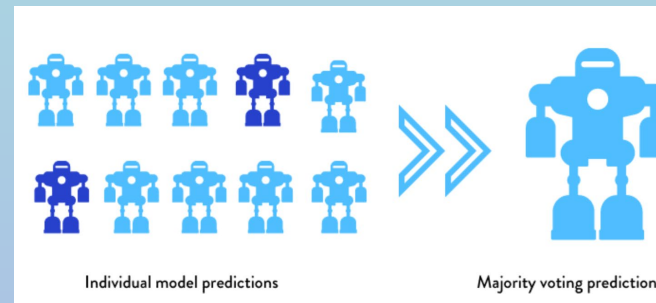
Fusion method is to combine the prediction output from the baseline models;

Two main categories are **voting** and **meta learning**

Voting: Assign a class label to an input instance

Types of **voting**:

- **Max voting** (hard voting or majority voting)
 - Collecting predictions for each class label and predicting the class label with the most votes



Ensemble Learning

- **Averaging voting**
 - Predictions are extracted from the multiple models and average of the predictions is used for the final prediction.

Example: Classification Task

Suppose we have three base models (Model A, Model B, and Model C) trained to classify images of fruits into three categories: apple, banana, and orange. Each model predicts the probabilities of each class for a given image. The predicted probabilities from each model for a particular image are as follows:

- Model A: [0.8, 0.1, 0.1] (80% probability for apple, 10% for banana, and 10% for orange)
- Model B: [0.6, 0.3, 0.1] (60% probability for apple, 30% for banana, and 10% for orange)
- Model C: [0.5, 0.4, 0.1] (50% probability for apple, 40% for banana, and 10% for orange)

To obtain the final prediction using average voting, we calculate the average probability for each class across all models:

- Average probabilities: $[(0.8 + 0.6 + 0.5) / 3, (0.1 + 0.3 + 0.4) / 3, (0.1 + 0.1 + 0.1) / 3]$
= [0.633, 0.267, 0.1]

The class with the highest average probability is chosen as the final prediction. In this case, the highest average probability corresponds to the "apple" class, so the final prediction is "apple."

Ensemble Learning

- **Weighted Averaging voting**
 - Modified version of average voting
 - Weight is given to each baseline model
 - Multiply each prediction by the weight of classifiers to produce a weighted sum and divide it by the sum of the weights of the classifier

- Model A predicts "spam" with a confidence score of 0.8.
- Model B predicts "spam" with a confidence score of 0.7.
- Model C predicts "spam" with a confidence score of 0.6.

Now, let's assign weights to each model based on some criteria other than relative importance or confidence:

- Model A: Weight = 0.4
- Model B: Weight = 0.3
- Model C: Weight = 0.2

In this example, the sum of weights is not equal to 1:

Sum of weights = $0.4 + 0.3 + 0.2 = 0.9$

To obtain the final prediction using weighted voting, we first multiply each model's confidence score for the "spam" class by its corresponding weight, then normalize the weighted sum by dividing by the sum of weights:

$$\begin{aligned}\text{Final Prediction (for "spam")} &= (0.4 * 0.8) + (0.3 * 0.7) + (0.2 * 0.6) / 0.9 \\ &= (0.32 + 0.21 + 0.12) / 0.9 \\ &= 0.65 / 0.9 \\ &\approx 0.722\end{aligned}$$

Now, let's calculate the final prediction for the "not spam" class:

$$\text{Final Prediction (for "not spam")} = 1 - 0.722 = 0.278$$

So the final prediction using weighted voting is:

- Approximately 72.2% confidence that the email is spam.
- Approximately 27.8% confidence that the email is not spam.

Ensemble Learning

- **Meta-learning (Learning to learn)**

- The meta-learner takes as input the predictions made by the base models on a given dataset and learns to combine these predictions effectively to improve overall performance.
- Stacking is an example of meta learning
 - involves training a model to combine the predictions of several other learning algorithms

Ensemble Learning

- Ensemble methods

- Boosting; Bagging ; Stacking

Boosting - multiple weak learners which are trained sequentially and fused using voting technique to create a strong learner;

Example:

- Adaptive Boosting (AdaBoost)
- Gradient Boosting
- eXtreme Gradient Boosting (XGBoosting)

Ensemble Learning

- Ensemble methods

- Bagging (Bootstrap aggregating)

- Create multiple small subsets of the data
 - Create more diverse prediction by adjusting the stochastic distribution of the training dataset.
 - Uses parallel approach

Eliminate overfitting and perform well on high dimensional data;

Computationally expensive

Example: Random forest

Ensemble Learning

- Ensemble methods

Stacking

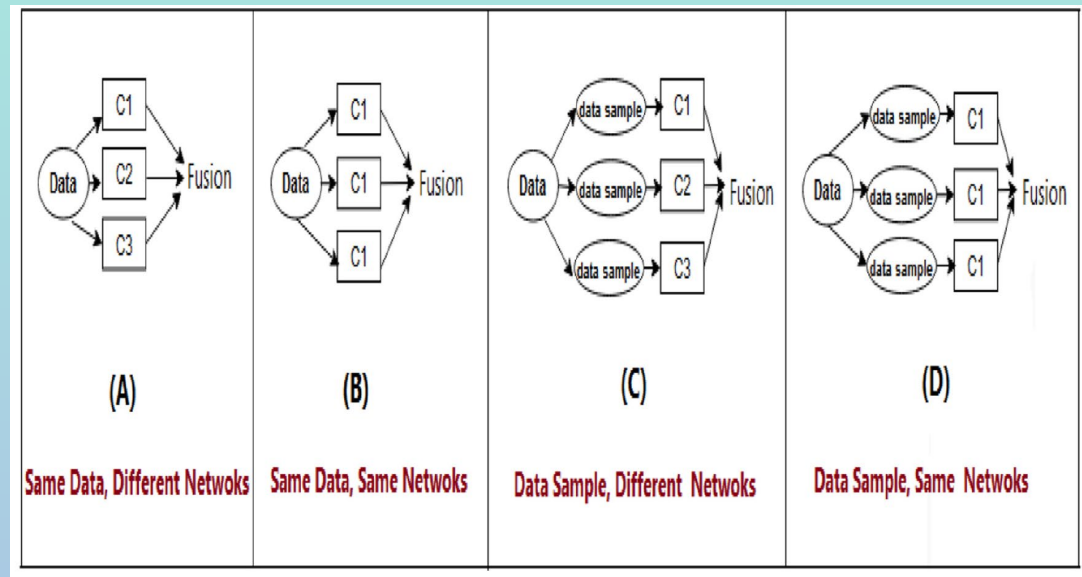
- Combine information from multiple predictive models to generate a new model (meta-model)
- The stacking model contains two or more base models referred to as level 0 models and a meta-model referred to as level 1 models
- Meta model learn how to combine the base model

Computationally complex

Ensemble Learning

Ensemble Deep learning

- A. Applying many different basic models using the same data.
- B. Applying different structures of the same basic model using the same data
- C. Applying many different basic models using many different data samples.
- D. Applying different structures of the same basic model using many different data samples.



Ensemble Learning

Practical Demo

[Ensemble Learning : Ensemble Techniques \(analyticsvidhya.com\)](https://analyticsvidhya.com/en/ensemble-learning/ensemble-techniques/)

References

- [1] A. Mohammed and R. Kora, “A comprehensive review on ensemble deep learning: Opportunities and challenges,” *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, 2023.