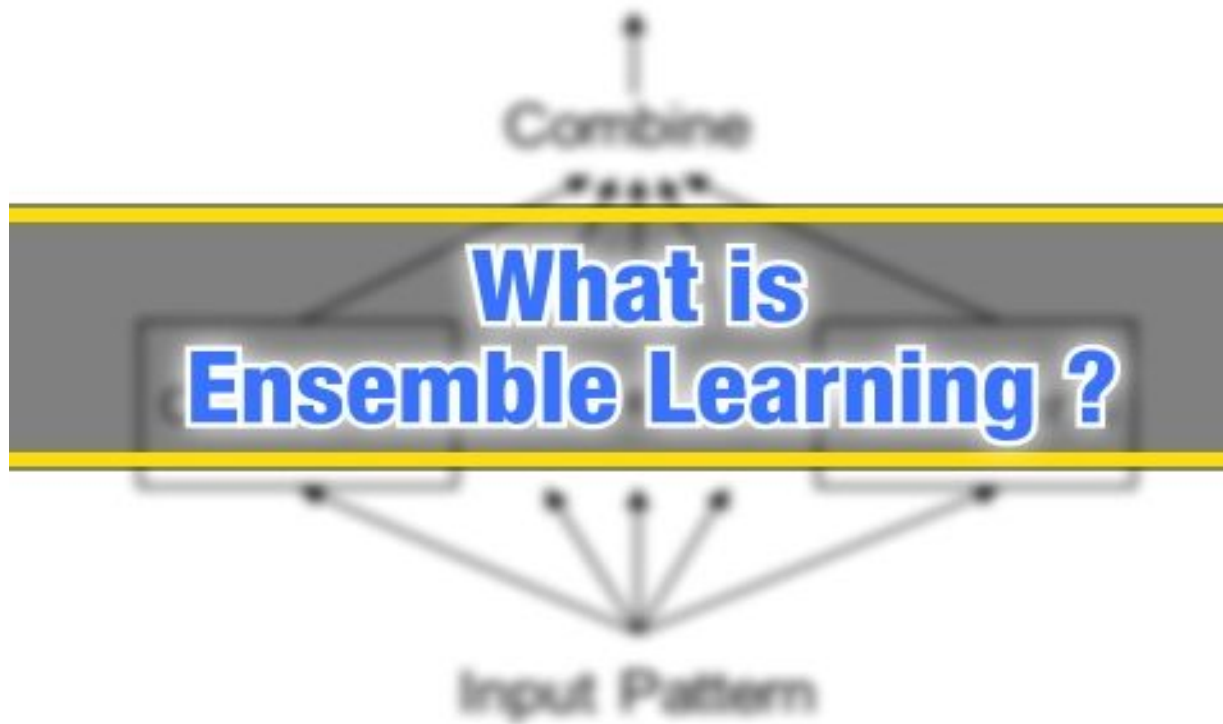# Ensemble Learning Techniques

Agenda

- Ensemble Learning
- Boosting
- Gradient Boosting and XGBoost
- Overfitting/Underfitting
- How to address Overfitting/Underfitting

Ensemble Learning

- **Ensemble methods** is a machine learning technique that combines several base models in order to produce one optimal predictive model.
- The process of generating models from data is called learning or training and the learned model can be called as hypothesis or learner.
- This type of machine learning algorithm helps in improving the overall performance of the model.
- The learning algorithms which construct a set of classifiers are known as Ensemble methods.
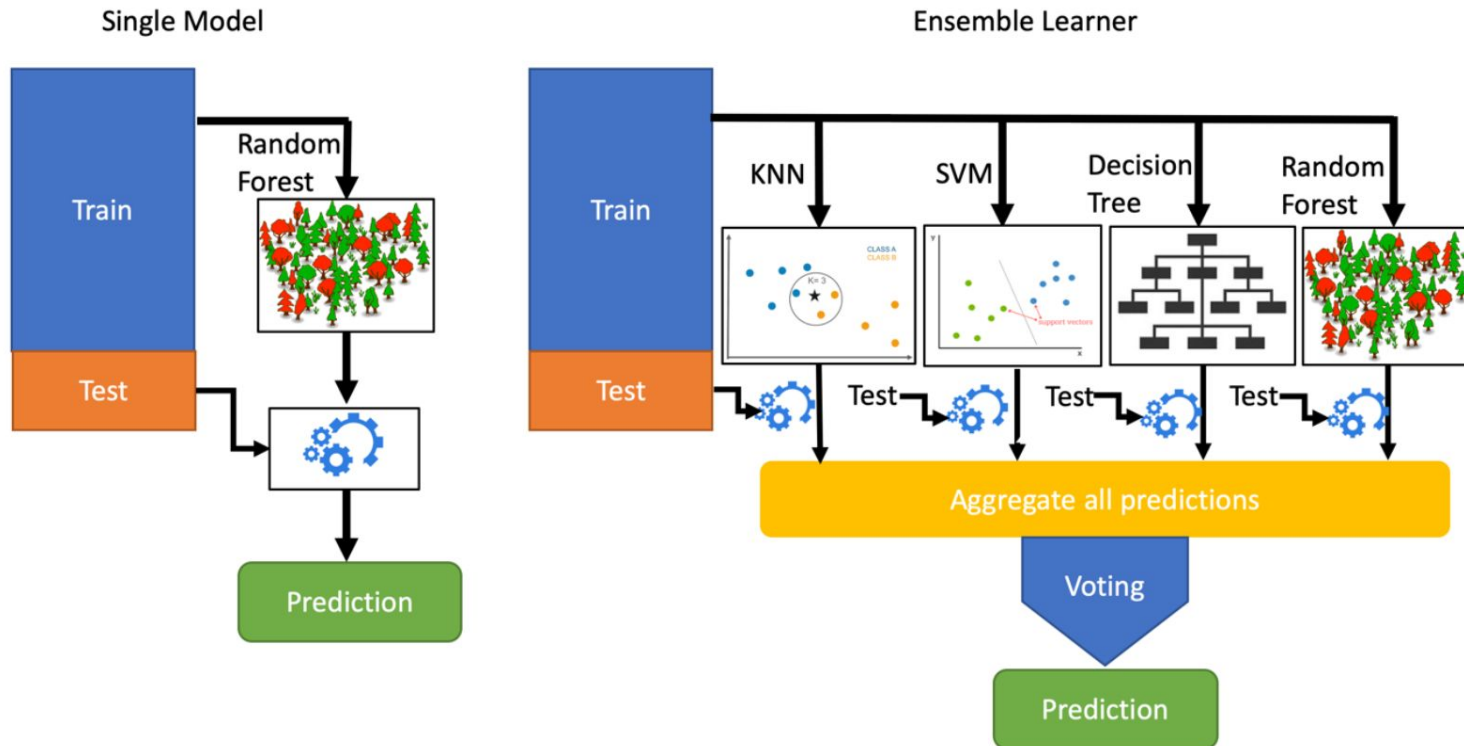
Ensemble Learning
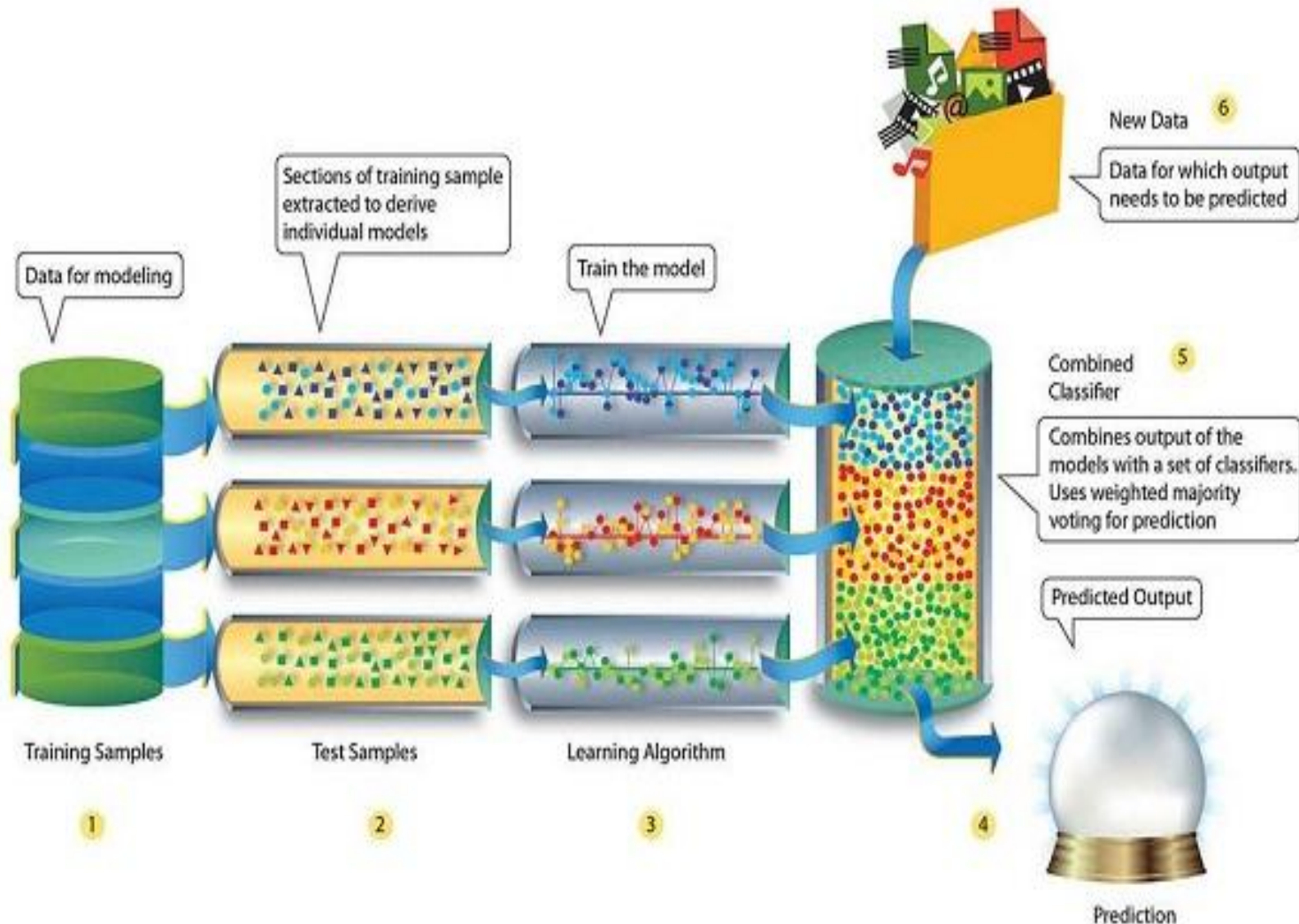
Single Model Prediction vs Ensemble Learner

Why Ensemble Methods?

- A diverse set of models in comparison to single models are likely to make better decisions.
- A decision tree basically works on several rules and provides a predictive output, where the rules are the nodes and their decisions will be their children and the leaf nodes will constitute the ultimate decision. The example of a decision tree below about a bank loan decision.
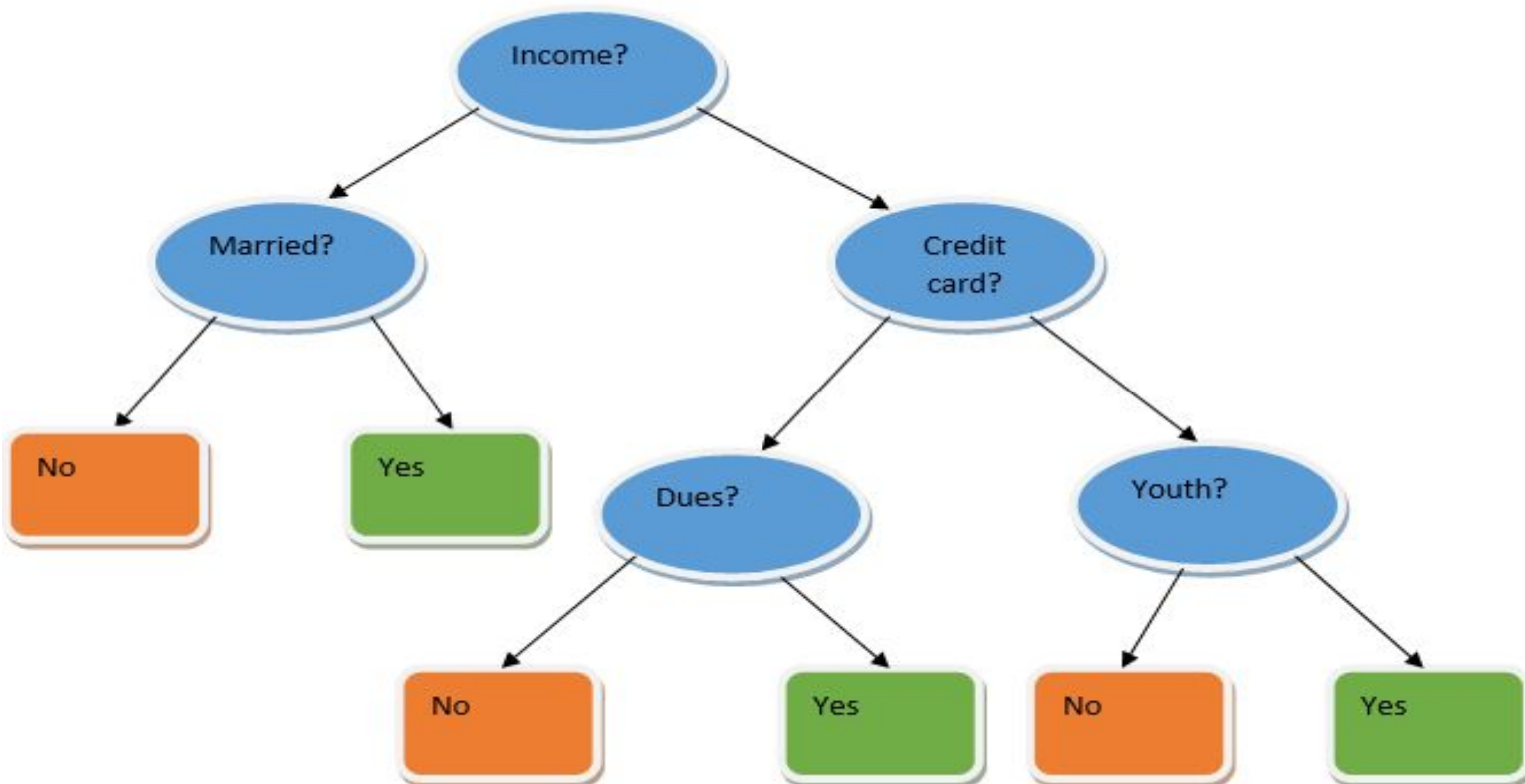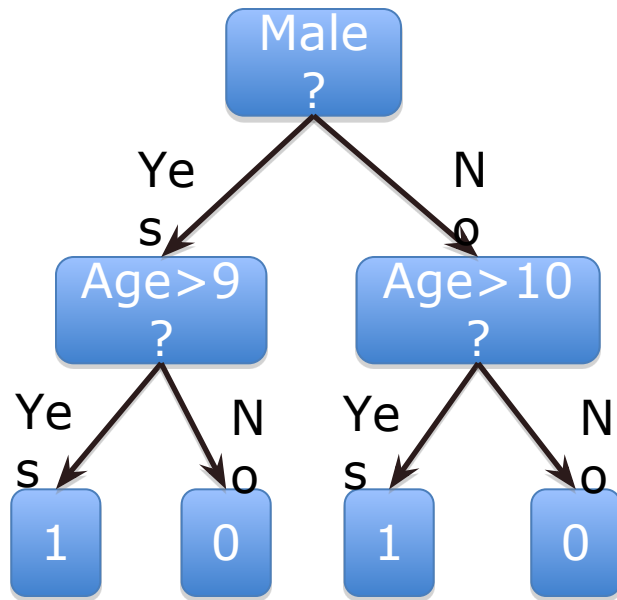
One classifier is not enough!

- Performance
  - None of the classifiers is perfect
  - Complementary
    - Examples which are not correctly classified by one classifier may be correctly classified by the other classifiers

- Potential Improvements
  - Utilize the complementary property

What is Ensemble Learning ?

An EXAMPLE

```
        Male?
      /        \
   Yes          No
    /              \
 Age>9?          Age>10?
  /    \          /    \
Yes    No       Yes    No
 /      \        /      \
 1       0       1       0
```

Name	Age	Male?	Height > 55"
Alice	14	0	1
Bob	10	1	1
Carol	13	0	1
Dave	8	1	0
Erin	11	0	0
Frank	9	1	1
Gena	8	0	0

✔
✔
✔
✔
✖
✖
✔

$$x = \begin{bmatrix} age \\ 1_{[gender=male]} \end{bmatrix}$$

$$y = \begin{cases} 1 & height > 55" \\ 0 & height \le 55" \end{cases}$$

Ensembles of Classifiers

Combine the classifiers to improve the performance

Ensembles of Classifiers

– Two ways to combine the classification results from different classifiers to produce the final output

  • Unweighted voting
  • Weighted voting

Example: Weather Forecast

Type of Ensemble methods:

The three most popular methods for combining the predictions from different models are:

- **Bagging** Building multiple models (typically of the same type) from different subsamples of the training dataset.

- **Boosting**. Building multiple models (typically of the same type) each of which learns to fix the prediction errors of a prior model in the chain.

- Building multiple models (typically of differing types) and simple statistics (like calculating the mean) are used to combine predictions.

Bias and Variance

- Bias is an error that occurs due to incorrect assumptions in our algorithm; a high bias indicates our model is too simple/underfit.
- Variance is the error that is caused due to sensitivity of the model to very small fluctuations in the data set; a high variance indicates our model is highly complex/overfit.
- An ideal ML model should have a proper balance between bias and variance.
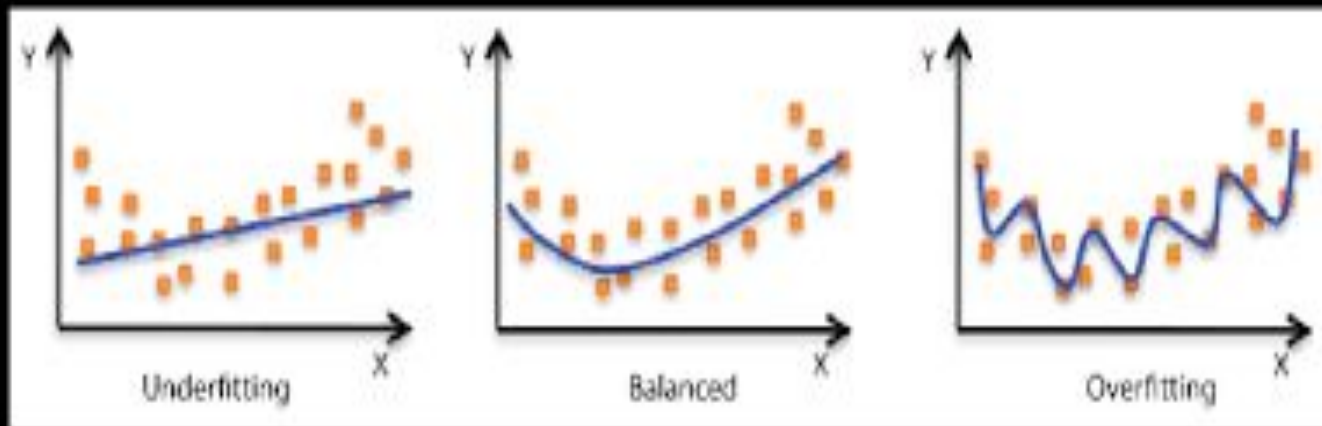
Ensemble methods

- Ensemble methods that minimize variance
  - Bagging
  - Random Forests

- Ensemble methods that minimize bias
  - Functional Gradient Descent
  - Boosting
  - Ensemble Selection

- Q.1 What is Ensemble Learning?

- Q.2 What is the need of ensemble learning in ml?

- Q.3 Why only one classifier is not enough in Machine Learning?

- Q.4 What are the types of Ensemble Methods?