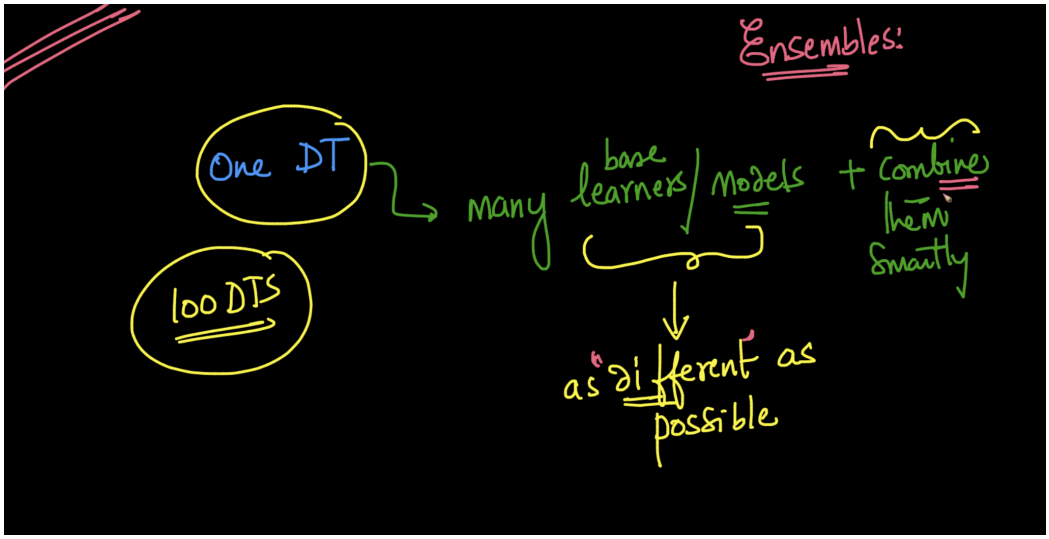
What is Ensemble Learning?

Ensemble Learning: **Training multiple base learners** or models that are as different as possible and **combining them smartly**



What is bagging?

Bagging is short for Bootstrap Aggregation where:

- Bootstrapping -> randomly creating samples of data out of a population with replacement
- Aggregation -> Clubbing predictions of each model to get the outcome

What are Random Forests?

A bagging ensemble which contains Decision Trees as the base learners

1. **The building block of RFs**
    - Sample m data points with replacement to get D_m'

- Train K different models for the K different datasets
- After training we cross-validate each model
- Now, we do Aggregation
    - We use a majority vote for the Classification
    - We use Mean/Median for Regression

2. Randomness in RFs

1. Row-sampling: For each base learner, we randomly select a subset of training data
2. Column-sampling: For each base learner we can select a subset of the columns
3. Depth tuning: By adjusting the maximum depth, you control how deep each tree in the ensemble can grow. Deeper trees may capture more complex relationships in the data but also risk overfitting.

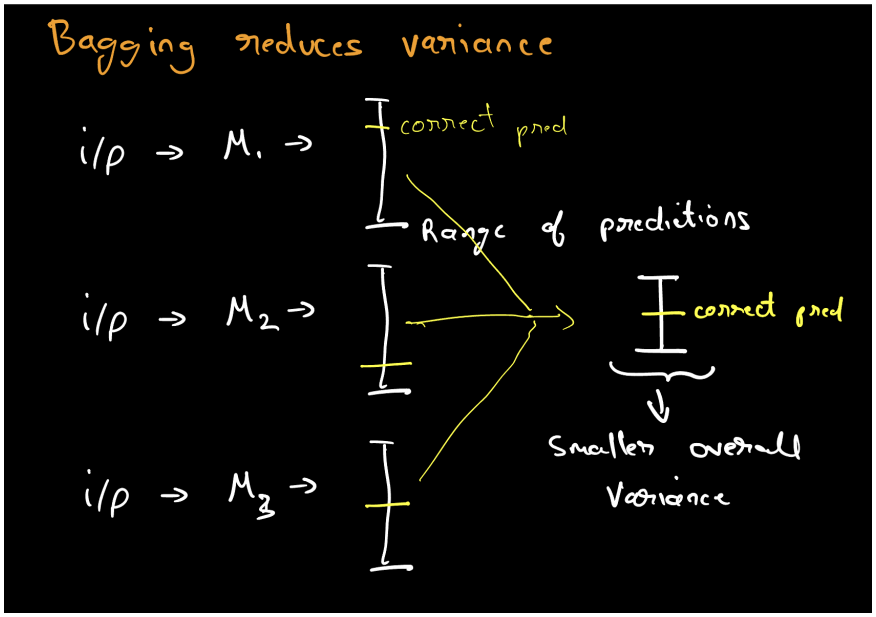
How to Perform Validation in Random Forests?

- Using Out of Bag (OOB) samples -> Samples in the training set which are not selected even once while bootstrapping

What is the Bias/Variance tradeoff in RFs?

- RFs can exhibit bias if the individual DTs are overly simplistic or if the algorithm is not tuned properly.
    - Bias can also arise if the number of trees in the forest is insufficient to capture the complexity of data
- More DTs generally leads to better generalization performance but there is a point at which adding additional trees does not significantly decrease variance but increases computational costs.
- Finding the optimal balance between bias and variance in RFs involves tuning hyperparameters such as the number of trees, the depth of trees and the size of the row/column sampling

How does bagging reduce error?



- Error = Bias^2 + Variance + Irreducible error.
- Since bagging reduces the variance through the use of bootstrap sampling and aggregation
    - Bootstrap sampling creates multiple random samples from the original training data to be used to train multiple models
    - After training multiple models on bootstrap samples, bagging aggregates their predictions to produce a final ensemble prediction
- Bagging keeps the overall bias mostly the same and the error decreases

Feature importance in Random Forests

- Compute the feature importance of a feature in each Decision Tree (Gini reduction and Information gain)
- Take the average of these values.

What if some base learners don't have the feature?

- The importance of that feature will be considered 0 in that Base learner

Training Random Forests

- Base learners can be trivially parallelized. I.e. Each base learner can be built in parallel.
- Each model is trained independently
- The time complexity thus becomes O(k * max_depth of tree)

OOB Score: The performance of the ensemble on the OOB sample is called the OOB score

Grid and Random Search

1. Grid search is systematic and ensures that every combination of hyperparameters is explored, making it thorough
    a. This is potentially computationally expensive, especially with a large search space.
2. Unlike grid search, random search does not systematically explore every possible combination of hyperparameters. Instead, it randomly samples hyperparameter values from the search space.
    a. While random search may not guarantee to find the optimal combination of hyperparameters, it can be more computationally efficient compared to grid search
    1. More number of base learners are required.