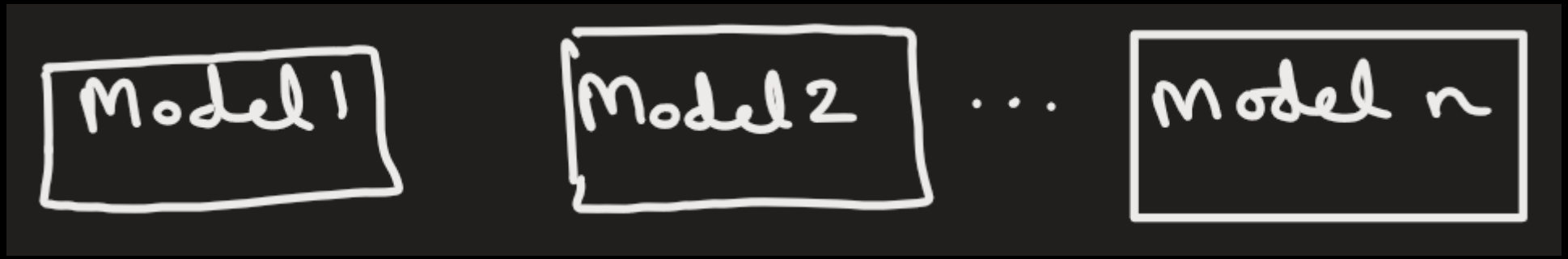
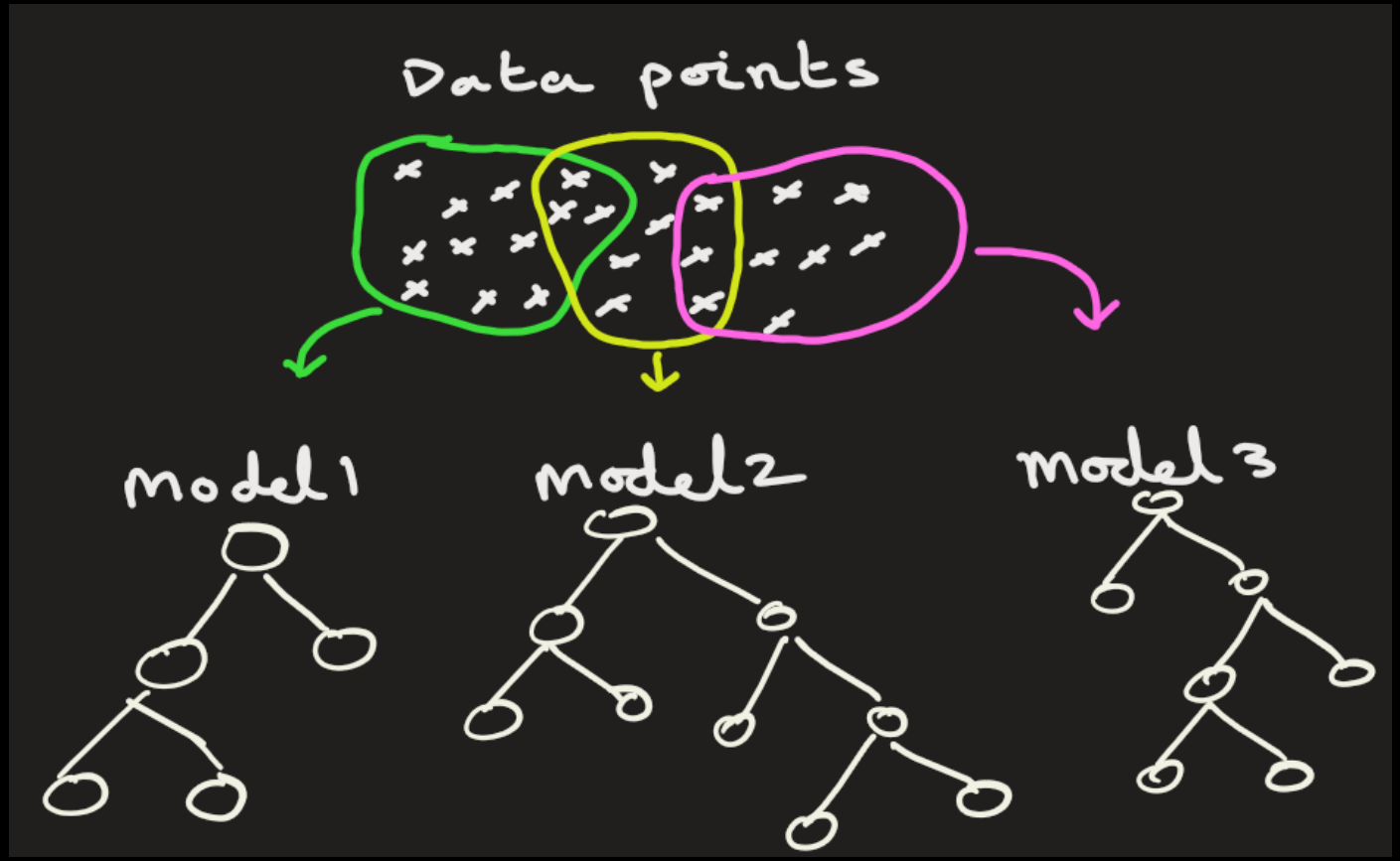
Ensemble Learning

Training multiple models and combining their predictions.



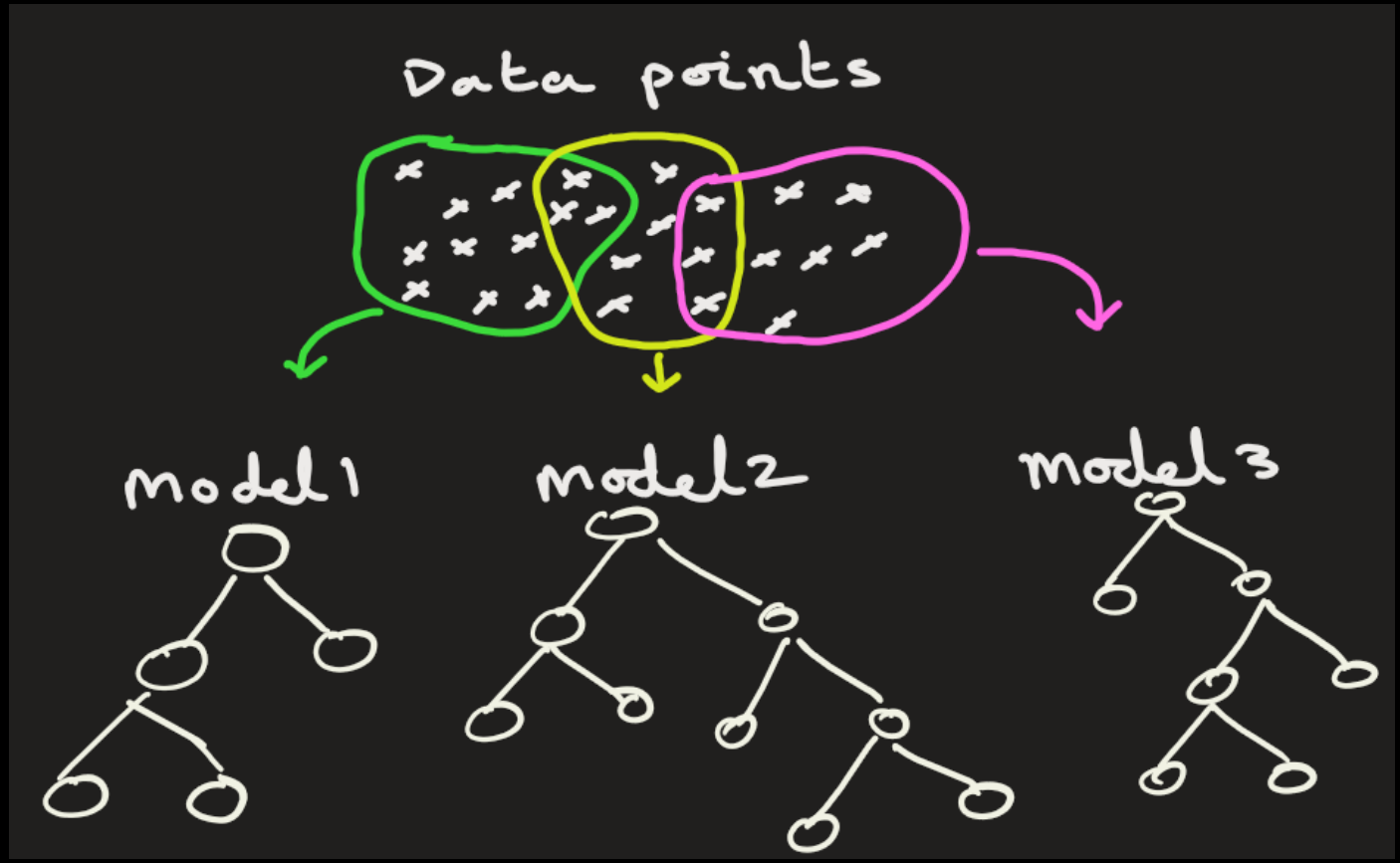
Bagging (Bootstrap Aggregating):

Train **many models independently** on **different random samples** of the data and **average their predictions**.



How it Works

1. Create multiple datasets using **bootstrapping** (sampling with replacement)
2. Train a **separate model** on each dataset
3. Combine predictions:
   Classification → **majority vote**
   Regression → **average**



Example: Random Forest

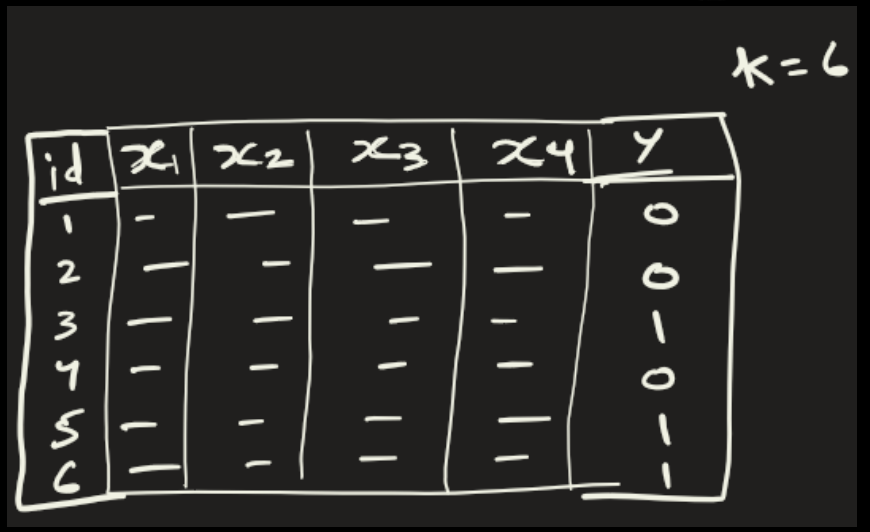
Step 1: Create Random Samples of Data and Features

From a dataset with *k* total records, we repeatedly select:

- **n random data points** (with replacement) → bootstrap sample
- **m random features** out of all available features

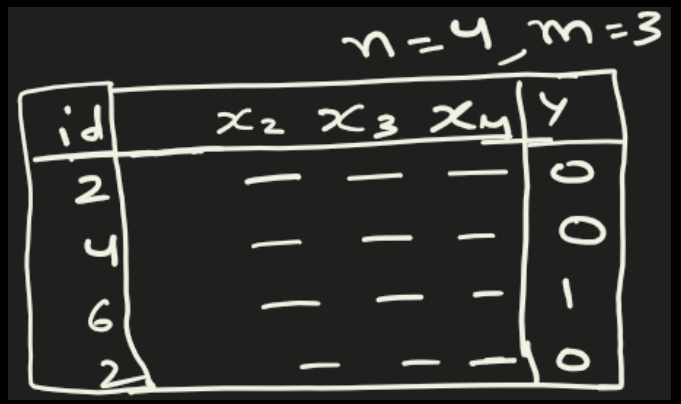
Each sample + feature subset will be used to build one decision tree.

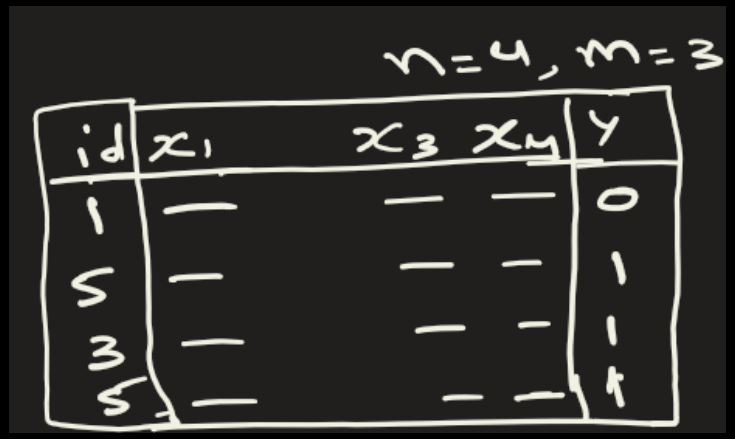
This randomness ensures that every tree is slightly different.



Example: Below we show 3 random samples from k=6 data points with 4 features.

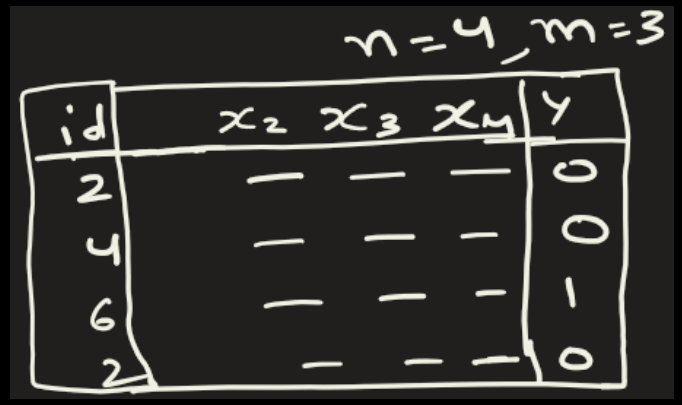
Each sample has n=4 data points and m=3 features.

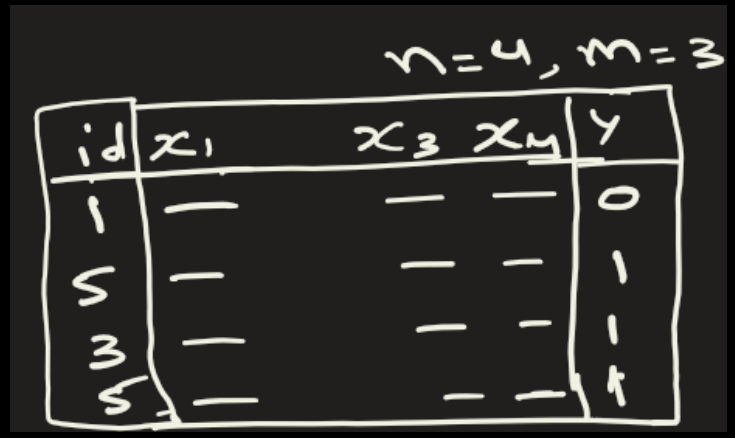


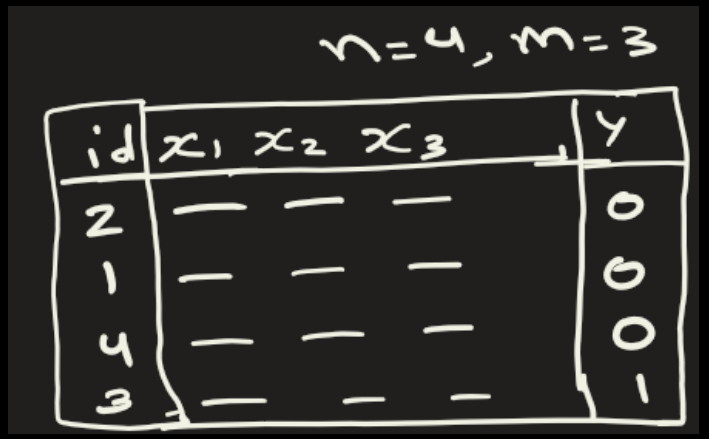


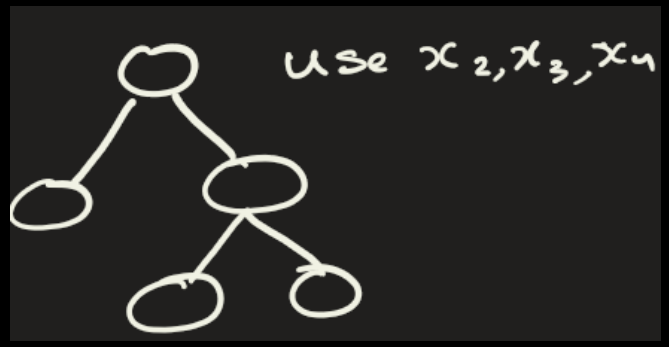
Step 2: Build a Decision Tree for Each Sample

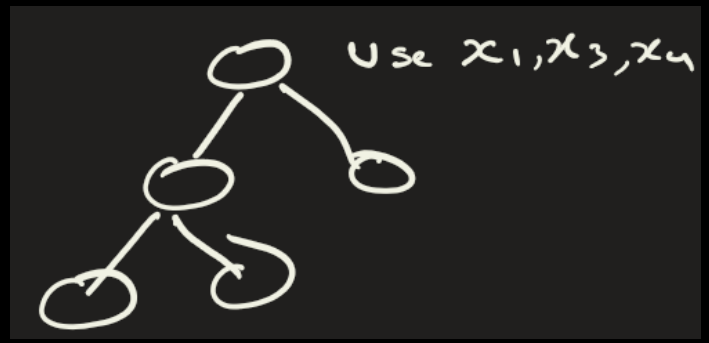
For each bootstrap sample grow a decision tree.
Each decision tree is a model.

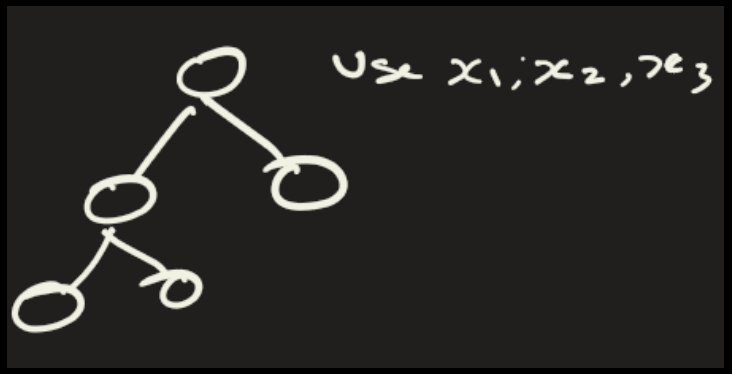








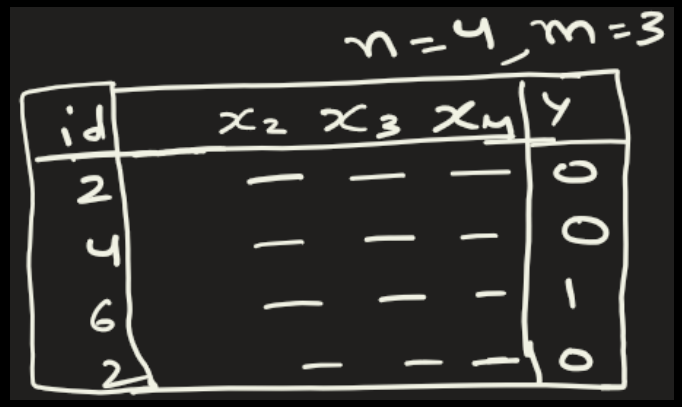


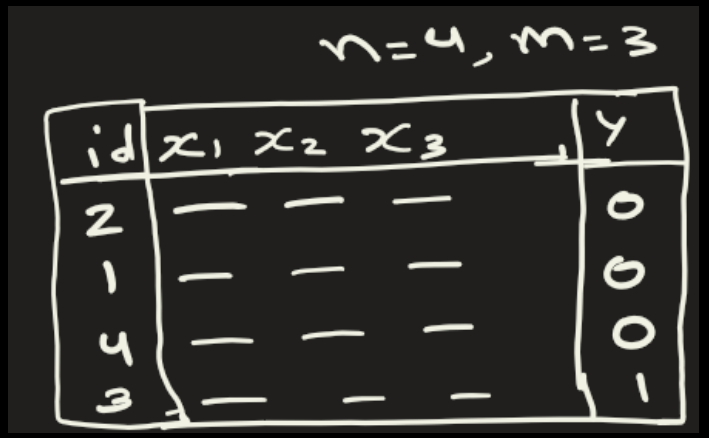


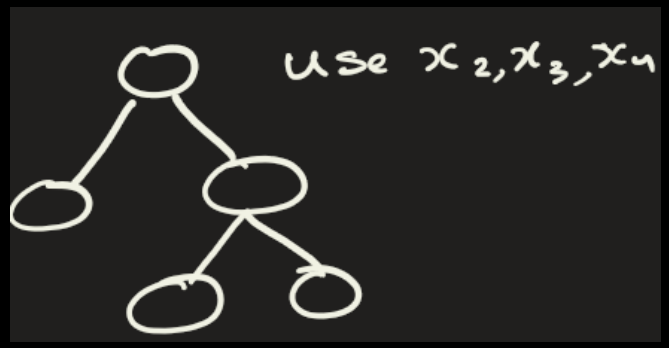
Step 3: Generate Predictions from All Trees

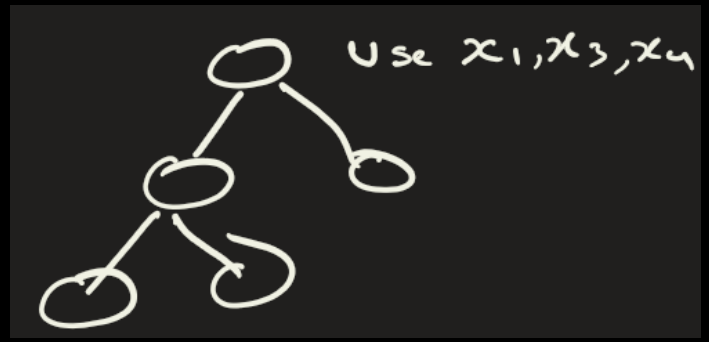
Each decision tree makes its own prediction:

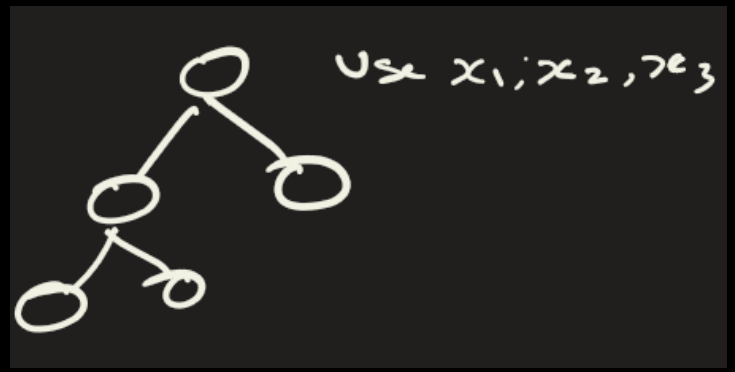
- A class label (for classification)
- A numerical value (for regression)







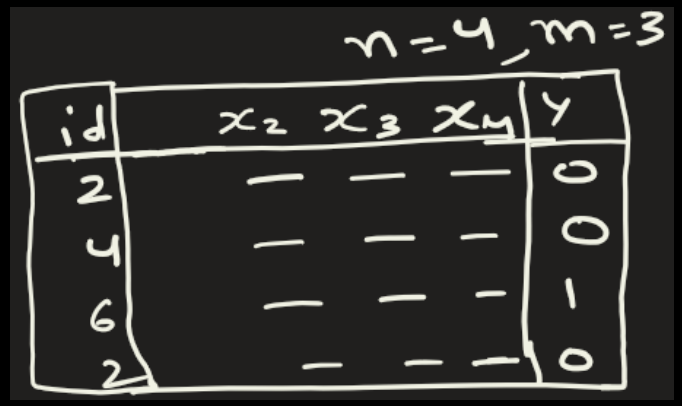


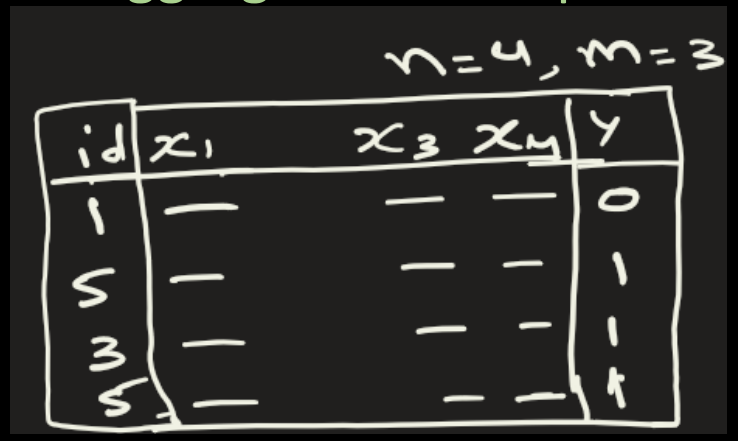


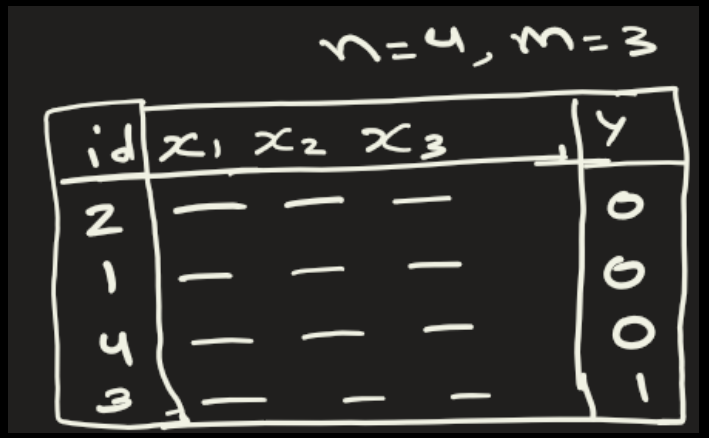
Step 4: Combine All Predictions to Form the Final Output

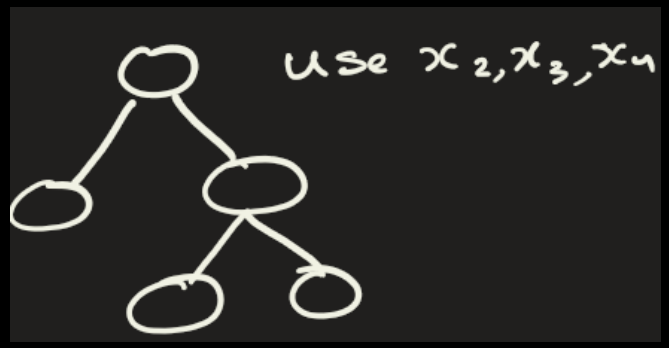
- **Classification:** Use **majority voting** → the class predicted by most trees becomes the final prediction.
- **Regression:** Use **averaging** → take the mean of all tree predictions.

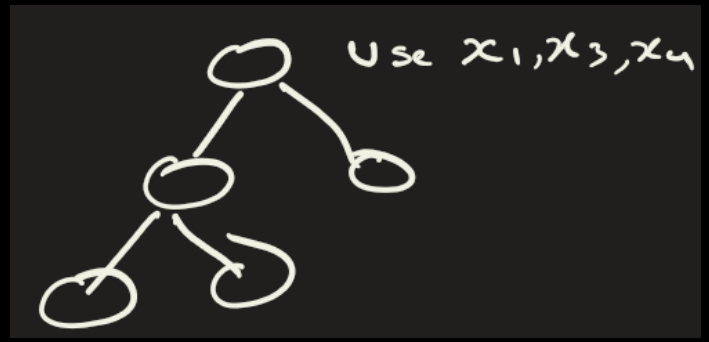
Note: Bootstrapping in step1 and Aggregation in step4 is called **Bagging.**

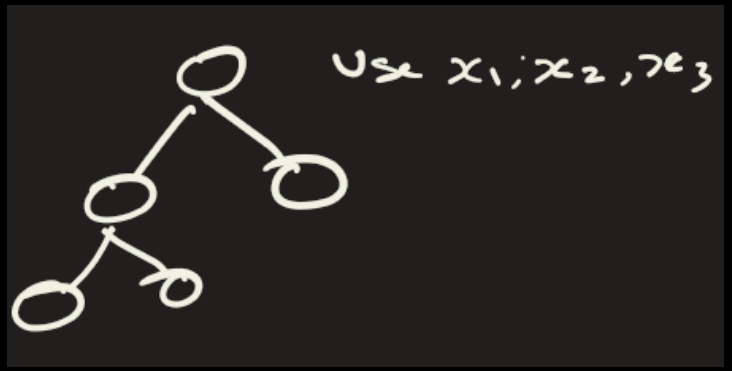










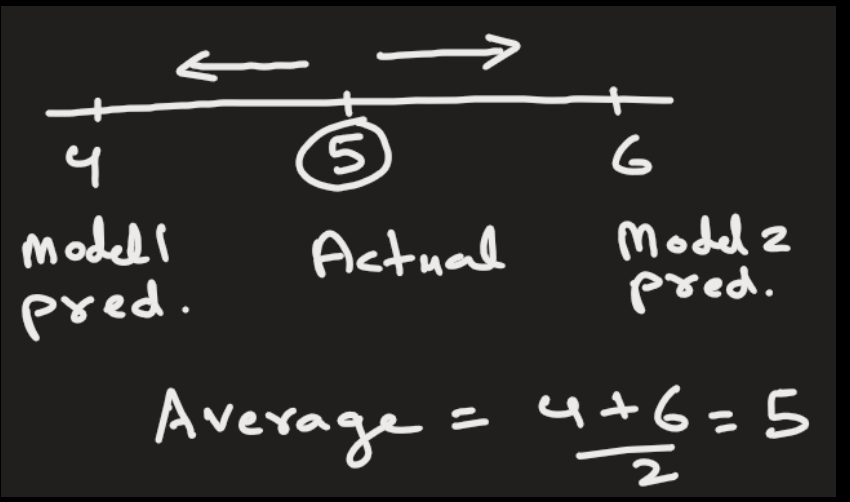


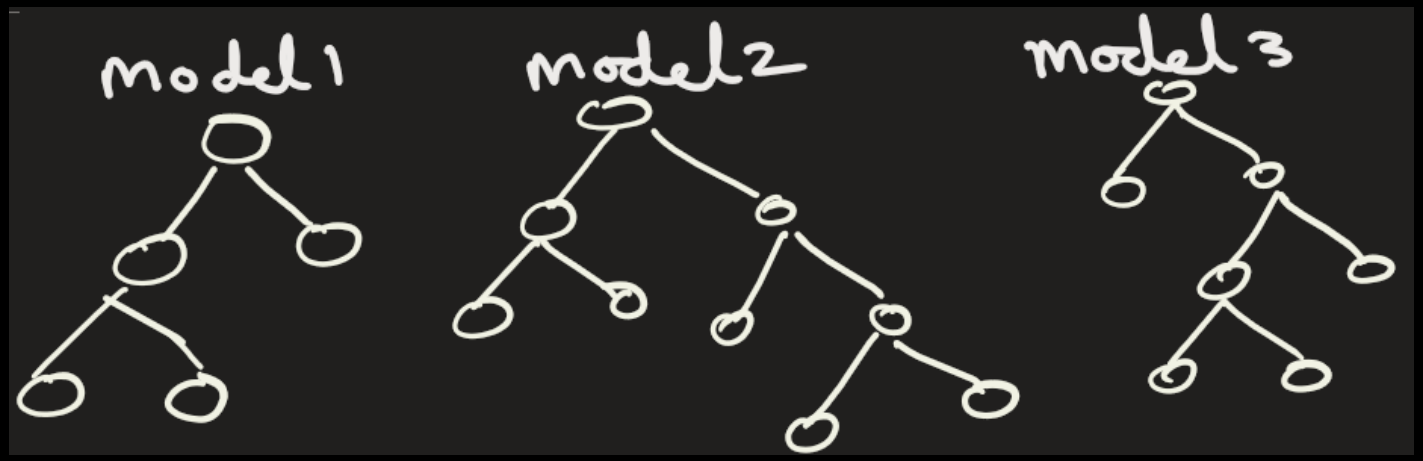
Bagging (Bootstrap Aggregating):

Key Characteristics

- Models are trained **in parallel** because each model is **independent**
- Reduces overfitting

Key idea: Reduce variance by averaging noisy models





Heading Goes Here

Fhdsklf
Fjdsklf
Fjskldf