🌳 Bagging (Bootstrap Aggregating) – Introduction

---

1. Core Idea

- Bagging is short for **Bootstrap Aggregating**.
- It's an **ensemble method** that combines multiple models to improve accuracy and stability.
- Works by:
    1. Taking **random subsets** of the training data (with replacement → bootstrap).
    2. Training a separate model on each subset.
    3. Combining predictions:
        - **Classification** → majority vote.
        - **Regression** → average.

---

2. Why Bagging?

- Single models (like decision trees) can be very **unstable** → small changes in data cause big differences in predictions.
- Bagging **reduces variance** by averaging results of many models → making predictions more **stable and accurate**.

---

3. Process (Step by Step)

Imagine you have a dataset with 100 rows.

1. **Bootstrap Sampling:**
    1. Randomly pick 100 rows *with replacement* → some rows repeat, some are left out.
    2. Do this multiple times to create several training sets.
2. **Model Training:**
    1. Train a model (say a decision tree) on each sample.
    2. Each model learns differently because each sample is slightly different.
3. **Aggregation:**
    - ❖ Combine outputs of all models:
        - Classification → **majority voting**.
        - Regression → **averaging**.

---

4. Example

Suppose we want to classify whether an email is spam:

- Model 1 says: Spam
- Model 2 says: Not Spam
- Model 3 says: Spam
- Final (Majority Vote) → **Spam**

---

5. Advantages of Bagging

1. Reduces **variance** → avoids overfitting.
2. Works well with **high-variance models** (e.g., Decision Trees).
3. Easy to implement.

---

6. Disadvantages

❌ Not effective if the base model already has **low variance** (like linear regression).

❌ More computationally expensive (training many models).

---

7. Intuition (Easy Analogy)

Think of **guessing a jar's number of candies**:

- If you ask one person → they might be far off.
- If you ask 100 people and take the average → the result is usually much closer to the truth.

That's bagging → "wisdom of the crowd" in action.

❖ Final Takeaway:

- Bagging = **Bootstrap sampling + Aggregation**.
- Reduces variance & increases stability.
- Often used with decision trees → leads to **Random Forest** (which is bagging + feature randomness).
