✳ Ensemble Learning – Introduction

1. Introduction

- **Ensemble Learning** = Using **multiple models together** instead of relying on just one.
- Idea: “A group of weak learners can form a strong learner.”
- Improves **accuracy, stability, and robustness**.

2. Wisdom of the Crowd

- Think of asking a difficult question to 100 people.
- Individually, answers vary. But the **average/majority** often comes very close to the truth.
- Same idea in ML: combining multiple models (votes/averages) leads to better predictions.

3. Resource [Book]

- A common reference: "Ensemble Methods in Machine Learning" or ML books covering **Bagging, Boosting, Stacking**.
- These provide the theory + algorithms.

4. Core Idea behind Ensemble Learning

- **Why use multiple models?**
    1. A single model may overfit or underperform.
    2. Different models learn **different aspects/patterns** of the data.
- By combining them:
    1. **Errors cancel out.**
    2. Predictions become more **generalized**.

5. Types of Ensemble Learning

The 3 big categories:

1. **Bagging (Bootstrap Aggregating):**
    1. Train multiple models on random subsets of the data.
    2. Predictions combined (majority vote or average).
    3. Example: **Random Forest**.
2. **Boosting:**
    1. Models are trained sequentially.

2. Each new model focuses on correcting errors made by previous ones.
3. Example: **AdaBoost, Gradient Boosting, XGBoost**.

3. **Stacking:**
    1. Different models are trained.
    2. Their predictions are fed into a "meta-model" which makes the final decision.
    3. More flexible and powerful.

---

6. Why Ensemble Learning Works?

- **Bias-Variance Tradeoff:**
    1. Bagging → reduces variance (stabilizes predictions).
    2. Boosting → reduces bias (corrects weak learners).
- Diversity in models makes the combined prediction more reliable.

---

7. Disadvantage of Ensemble Learning

- Computationally expensive (training multiple models).
- Less interpretable (harder to explain compared to a single tree or regression).

---

8. Benefits of Ensemble Learning

- Higher accuracy.
- Works well even if individual models are weak.
- More robust to noisy data.
- Often used in **Kaggle competitions** because it consistently boosts performance.

---

9. When to use Ensemble Learning?

- When single models underperform.
- When the dataset is complex, high-dimensional, or noisy.
- When maximizing prediction accuracy is more important than interpretability.

➢ Key Takeaway:

Ensemble Learning is like teamwork — weak players alone might lose, but together they can win championships. It leverages **diversity + combination** to build powerful predictive models.
