



Voting Ensemble – Classification (Part 2)

1. Quick Recap – Core Idea

- Ensemble = Combine multiple models to improve performance.
 - Voting Classifier = Special implementation for **classification problems**.
 - Two main strategies:
 1. **Hard Voting** → Majority vote of predicted classes.
 2. **Soft Voting** → Average of predicted probabilities.
-

2. Hard Voting – How it works

- Each classifier gives a **final label** (e.g., Cat/Dog).
- The class that gets the **most votes** becomes the prediction.



Example:

- Logistic Regression → **Cat**
- Random Forest → **Dog**
- SVM → **Dog**

👉 **Final result = Dog** (majority).



Good when models are reasonably accurate.



Weak if some models are very bad (they still get equal weight).

3. Soft Voting – How it works

- Each classifier outputs **probability scores** instead of labels.
- The probabilities are **averaged** (sometimes weighted), and the class with the highest avg probability wins.



Example:

- Logistic Regression → Cat (0.6), Dog (0.4)
- Random Forest → Cat (0.3), Dog (0.7)
- SVM → Cat (0.4), Dog (0.6)

Averaging:

- Cat = $(0.6 + 0.3 + 0.4)/3 = \mathbf{0.43}$
- Dog = $(0.4 + 0.7 + 0.6)/3 = \mathbf{0.57}$



Final result = **Dog**.



More powerful because it uses **confidence** of predictions.



Can give better results when classifiers are **calibrated well**.

4. When to Use Which?

- **Hard Voting:**
 1. When classifiers don't provide reliable probabilities.
 2. Simpler but may ignore confidence.
 - **Soft Voting:**
 1. When models can output **well-calibrated probabilities**.
 2. Usually better accuracy because it considers confidence.
 3. Works best when models are **diverse and balanced**.
-

5. Key Intuition

Think of it as **group decision-making**:

- Hard Voting = “raise your hand if you think it's Cat or Dog” → Majority wins.
 - Soft Voting = “say how confident you are” → Combine confidences → Best-supported choice wins.
-

Final Takeaway:

Voting Ensemble is one of the **simplest but effective ensemble methods**.

- **Hard Voting** = majority rule.
 - **Soft Voting** = average probabilities.
 - Soft is usually preferred if probabilities are available and reliable.
-