

# Tutorial: Ensemble Methods

## *Cheat Sheet and Practice Problems*

ES335 - Machine Learning  
IIT Gandhinagar

July 23, 2025

## 1 Summary from Slides

### 1.1 Basic Concept

**Ensemble Learning:** Use multiple models for prediction

- Most winning entries of Kaggle competitions use ensemble learning
- Combines predictions from multiple base learners
- Often outperforms individual models

**Classification Example:**

- Classifier 1: Good
- Classifier 2: Good
- Classifier 3: Bad
- Majority Voting: Predict "Good"

**Regression Example:**

- Regressor 1: 20
- Regressor 2: 30
- Regressor 3: 30
- Averaging: Predict  $\frac{80}{3} = 26.67$

### 1.2 Intuition and Conditions

**Why Ensembles Work:**

- Individual models make different types of errors
- Combining reduces overall error through error averaging
- Diversity among base learners is crucial

**Necessary and Sufficient Conditions:**

- Base learners should be better than random guessing
- Base learners should be diverse (make different errors)
- Independence of base learner errors is ideal

## 1.3 Random Forests

### Key Concepts:

- Ensemble of decision trees
- Uses bootstrap sampling (bagging)
- Random feature selection at each split
- Reduces overfitting compared to single deep trees

### Representation:

- Limited depth decision trees have limited representation
- Random Forests expand representation through combination
- Better generalization than individual trees

## 1.4 Quantitative Perspective

**Error Reduction:** If individual classifiers have error rate  $\epsilon < 0.5$  and are independent, ensemble error decreases exponentially with number of classifiers

### Bias-Variance Trade-off:

- Bagging reduces variance
- Boosting can reduce both bias and variance
- Ensemble typically has lower generalization error

## 2 Practice Problems

### Problem : Basic Ensemble Voting

Three binary classifiers make the following predictions on 5 test samples:

Sample	Classifier 1	Classifier 2	Classifier 3	True Label
1	0	1	1	1
2	1	1	0	1
3	0	0	1	0
4	1	0	0	0
5	1	1	1	1

- a) Calculate individual classifier accuracies b) Calculate ensemble accuracy using majority voting c) Which performs better: individual classifiers or ensemble?

### Problem : Regression Ensemble

Four regression models predict house prices (in \$1000s):

House	Model 1	Model 2	Model 3	Model 4	True Price
A	250	280	240	260	255
B	180	200	170	190	185
C	320	310	330	315	318

- a) Calculate individual model MSEs b) Calculate ensemble MSE using simple averaging c) Try weighted averaging with weights [0.4, 0.3, 0.2, 0.1] - does it improve?

**Problem : Diversity Analysis**

Two classifiers have the following error patterns on 100 samples: - Classifier A: Correct on samples 1-70, wrong on 71-100 - Classifier B: Correct on samples 21-90, wrong on 1-20 and 91-100

- a) What is the individual accuracy of each classifier? b) On how many samples do both classifiers agree? c) What is the ensemble accuracy using majority voting? d) Calculate the diversity between the two classifiers

**Problem : Bootstrap Sampling**

You have a training set with 6 samples: [A, B, C, D, E, F]

- a) Create 3 different bootstrap samples of size 6 b) Which samples might be out-of-bag for each bootstrap? c) How does bootstrap sampling introduce diversity? d) What fraction of original samples are expected to be in each bootstrap sample?

**Problem : Random Forest Construction**

Design a Random Forest with the following specifications: - 5 trees - Maximum depth of 3 - Consider  $\sqrt{p}$  features at each split where  $p = 16$  total features

- a) How many features are considered at each split? b) Draw the structure for one possible tree c) How do you combine predictions for classification vs regression? d) What are the advantages over a single deep tree?

**Problem : Ensemble Error Analysis**

Given  $n$  independent binary classifiers, each with error rate  $\epsilon = 0.3$ :

- a) What is the ensemble error rate for  $n = 3$  using majority voting? b) Calculate for  $n = 5$  and  $n = 7$  c) How does ensemble error change as  $n$  increases? d) What happens if  $\epsilon > 0.5$ ?

**Problem : Feature Importance in Random Forests**

A Random Forest is trained on a dataset with features [Age, Income, Education, Location]: After training, the feature importance scores are: - Age: 0.4 - Income: 0.3 - Education: 0.2 - Location: 0.1

- a) Interpret these importance scores b) How are feature importances calculated in Random Forests? c) Which features would you select for a simpler model? d) How does random feature selection affect importance calculation?

**Problem : Bias-Variance in Ensembles**

Compare bias and variance for: - Single decision tree (depth 10) - Random Forest (100 trees, depth 5) - Bagged trees (100 trees, depth 10)

- a) Which has highest bias? Why? b) Which has highest variance? Why? c) How does bootstrap sampling affect bias and variance? d) What is the bias-variance trade-off in each method?

**Problem : Ensemble Combination Methods**

You have 5 binary classifiers with the following probability outputs for a sample: [0.8, 0.3, 0.6, 0.9, 0.4]

- Calculate the ensemble prediction using: a) Simple majority voting (threshold 0.5) b) Average probability c) Weighted average with weights [0.3, 0.1, 0.2, 0.3, 0.1] d) Maximum probability e) Which methods agree and which disagree?

### Problem : Out-of-Bag Error Estimation

In a Random Forest with 100 trees:

- a) What fraction of samples are out-of-bag for each tree on average? b) How is out-of-bag error calculated? c) Why is OOB error a good estimate of generalization error? d) Compare OOB error with k-fold cross-validation - advantages and disadvantages?

### Problem : Ensemble Diversity Measures

Three classifiers make predictions on 8 samples:

Sample	C1	C2	C3
1	1	1	1
2	1	1	0
3	1	0	1
4	1	0	0
5	0	1	1
6	0	1	0
7	0	0	1
8	0	0	0

- a) Calculate pairwise disagreement between classifiers b) Calculate Q-statistic for each pair c) Which pair is most diverse? d) How does diversity relate to ensemble performance?

### Problem : Stacking vs Voting

Design a two-level ensemble: - Level 1: Train 4 base models (SVM, Random Forest, Logistic Regression, KNN) - Level 2: Meta-learner combines Level 1 predictions

- a) What are the inputs to the meta-learner? b) How do you prevent overfitting in stacking? c) Compare stacking with simple voting d) When might stacking perform worse than voting?

### Problem : Computational Complexity

Compare computational costs: - Training: Single tree vs Random Forest (100 trees) - Prediction: Single tree vs Random Forest - Memory: Single tree vs Random Forest

- a) How does ensemble size affect training time? b) Can ensemble prediction be parallelized? c) What are the memory requirements for storing 100 trees? d) When might single models be preferred despite lower accuracy?

### Problem : Real-world Ensemble Design

You're building an ensemble for credit card fraud detection with: - Highly imbalanced data (0.1% fraud) - Need real-time predictions ( $\leq 100\text{ms}$ ) - Interpretability required for regulatory compliance

- a) What base models would you choose and why? b) How would you handle class imbalance in each base model? c) What ensemble combination method is most appropriate? d) How would you ensure interpretability? e) What constraints does the timing requirement impose?

## Problem : Advanced Challenge

Prove mathematically that for binary classification with  $n$  independent classifiers (each with error rate  $\epsilon < 0.5$ ), the ensemble error using majority voting is:

$$P(\text{ensemble error}) = \sum_{k=\lceil n/2 \rceil}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k}$$

- a) Derive this formula step by step b) Show that this error decreases as  $n$  increases when  $\epsilon < 0.5$  c) What happens when  $\epsilon = 0.5$ ? d) Extend to the case where classifiers have different error rates