

MACHINE LEARNING – UNIT III

NOTES

RGPV Exam-Oriented Notes

 Easy + Detailed + One-Day Before Exam Preparation

UNIT III SYLLABUS

1. Ensemble Learning
 2. Model Combination Schemes
 3. Voting
 4. Error Correcting Output Codes (ECOC)
 5. Bagging
 6. Random Forest Trees
 7. Boosting
 8. AdaBoost
 9. Stacking
-

1. ENSEMBLE LEARNING

Definition

Ensemble Learning is a machine learning technique in which multiple models are combined together to improve prediction accuracy.

Easy Explanation

Instead of depending on one model, Ensemble Learning combines many models and takes better decisions.

Example:

One student predicts marks incorrectly.
5 students together predict more accurately.

Similarly:

Multiple ML models → Better prediction

Working of Ensemble Learning

Training Data
↓
Multiple Models
↓
Combine Predictions
↓
Final Output

Why Ensemble Learning is Used?

Single model may:

- make errors
- overfit data
- give unstable predictions

Combining models:

- ✓ increases accuracy
- ✓ reduces errors
- ✓ improves performance

Types of Ensemble Learning

Method	Meaning
Voting	Majority prediction
Bagging	Parallel learning
Boosting	Sequential learning
Stacking	Meta-learning

Advantages

- ✓ Higher accuracy
- ✓ Reduces overfitting
- ✓ Better generalization
- ✓ Handles complex problems

Disadvantages

- ✗ High computation cost
- ✗ Difficult implementation
- ✗ More memory required

Applications

- Fraud detection
- Recommendation systems
- Medical diagnosis
- Stock prediction

Conclusion

Ensemble Learning combines multiple models to improve prediction accuracy and reduce errors.

2. MODEL COMBINATION SCHEMES

Definition

Model Combination Scheme is a method used to combine outputs of multiple classifiers into one final prediction.

Easy Explanation

Different models give different outputs.

Combination schemes merge all outputs and produce final result.

Types of Combination Schemes

Scheme	Meaning
Majority Voting	Most common prediction selected
Averaging	Average of outputs
Weighted Voting	Important models get more weight
Stacking	Meta-model combines outputs

Example

Model 1 → Yes

Model 2 → Yes

Model 3 → No

Final Output → Yes

Advantages

- Improves reliability
 - Reduces individual model errors
-

Conclusion

Model Combination Schemes help improve overall machine learning performance.

3. VOTING

Definition

Voting is an ensemble technique in which multiple classifiers vote for the final output.

Types of Voting

Type	Meaning
Hard Voting	Majority class selected
Soft Voting	Probability-based voting

Example

Classifier 1 → Cat

Classifier 2 → Dog

Classifier 3 → Cat

Final Prediction → Cat

Advantages

- ✓ Simple method
 - ✓ Improves accuracy
 - ✓ Reduces model bias
-

Disadvantages

- ✗ Weak classifiers reduce performance
-

Applications

- Image classification
 - Spam filtering
 - Sentiment analysis
-

Conclusion

Voting combines predictions from multiple classifiers and selects the most common result.

4. ERROR CORRECTING OUTPUT CODES (ECOC)

Definition

ECOC is a technique used to solve multi-class classification problems using multiple binary classifiers.

Easy Explanation

Complex classification problems are divided into many simple binary classification tasks.

Working

Multi-Class Problem



Binary Classifiers



Combined Output

Example

Suppose classes are:

A, B, C

ECOC converts them into binary code patterns.

Advantages

- ✓ Improves classification accuracy
 - ✓ Reduces classification errors
-

Disadvantages

- ✗ Complex implementation
 - ✗ Requires many classifiers
-

Applications

- Image recognition

- Speech recognition
 - Text classification
-

Conclusion

ECOC converts multi-class problems into binary problems for efficient classification.

5. BAGGING

Full Form

Bagging = Bootstrap Aggregating

Definition

Bagging is an ensemble method in which multiple models are trained independently using random subsets of data.

Easy Explanation

Different datasets are created from original data.

Each model learns separately.

Final prediction is obtained using voting or averaging.

Working

Original Dataset



Random Subsets



Multiple Models



Combine Outputs

Important Point

Bagging reduces:

Variance

Advantages

- ✓ Reduces overfitting
 - ✓ Improves stability
 - ✓ Better accuracy
-

Disadvantages

- ✗ High computation cost
 - ✗ Requires multiple models
-

Applications

- Random Forest
 - Financial prediction
 - Risk analysis
-

Conclusion

Bagging improves performance by combining multiple independently trained models.

6. RANDOM FOREST TREES

Definition

Random Forest is an ensemble learning algorithm that combines multiple decision trees.

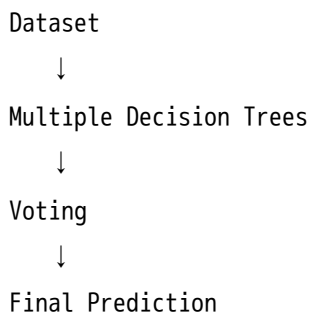
Easy Explanation

Instead of using one decision tree, Random Forest uses many decision trees.

Each tree gives prediction.

Final prediction is selected using voting.

Working



Features

Feature	Explanation
Based on Bagging	Uses random subsets
Uses Multiple Trees	Improves prediction
Final Output by Voting	Majority result selected

Advantages

- ✓ High accuracy
- ✓ Reduces overfitting
- ✓ Handles large datasets
- ✓ Works for classification and regression

Disadvantages

- ✗ Slow training
- ✗ Large memory requirement

Applications

- Disease prediction
 - Stock market analysis
 - Recommendation systems
-

Conclusion

Random Forest combines many decision trees to produce accurate and stable predictions.

7. BOOSTING

Definition

Boosting is an ensemble technique that trains models sequentially to improve weak learners.

Easy Explanation

Each new model focuses more on previously misclassified examples.

Working

Weak Learner 1



Errors Identified



Weak Learner 2 Improves Errors



Final Strong Model

Important Point

Boosting reduces:

Bias

Advantages

- ✓ High prediction accuracy
- ✓ Converts weak learner into strong learner

Disadvantages

- ✗ Sensitive to noise
- ✗ Slow training process

Applications

- Face detection
- Spam filtering
- Fraud detection

Conclusion

Boosting improves model accuracy by correcting previous mistakes sequentially.

8. ADABOOST

Full Form

AdaBoost = Adaptive Boosting

Definition

AdaBoost is a boosting algorithm that combines weak classifiers into one strong classifier.

Easy Explanation

Initially all data points have equal importance.

Wrong predictions get higher weights.

Next model focuses more on difficult examples.

Working

1. Train weak classifier.
 2. Find wrong predictions.
 3. Increase weights of wrong examples.
 4. Train next classifier.
 5. Combine all classifiers.
-

Diagram

Weak Learner 1 → Errors

Weak Learner 2 → Correct Errors

Weak Learner 3 → Better Prediction

Advantages

- ✓ High accuracy
 - ✓ Handles complex classification
-

Disadvantages

- ✗ Sensitive to noisy data
 - ✗ Computationally expensive
-

Applications

- Face recognition
 - Text classification
 - Medical diagnosis
-

Conclusion

AdaBoost improves classification performance by focusing on previously misclassified examples.

9. STACKING

Definition

Stacking is an ensemble learning technique where outputs of multiple models are combined using another model called meta-learner.

Easy Explanation

Different models first make predictions.

Another model learns from those predictions and gives final output.

Working

Base Models
↓
Predictions
↓
Meta Learner
↓
Final Output

Advantages

- ✓ Improves prediction accuracy
 - ✓ Combines strengths of multiple models
-

Disadvantages

✗ Complex implementation

✗ High computation cost

Applications

- Recommendation systems
 - Forecasting systems
 - AI prediction systems
-

Conclusion

Stacking combines outputs of multiple models using a meta-model for better predictions.

MOST IMPORTANT 7-MARK QUESTIONS

1. Explain Ensemble Learning.
 2. Explain Voting method.
 3. Explain Bagging.
 4. Explain Random Forest Trees.
 5. Explain Boosting.
 6. Explain AdaBoost Algorithm.
 7. Explain Stacking method.
 8. Explain ECOC.
-

MOST IMPORTANT 14-MARK QUESTIONS

★ Very Important

1. Explain Ensemble Learning and its types.
 2. Explain Bagging and Random Forest Trees.
 3. Explain Boosting and AdaBoost.
 4. Differentiate Bagging and Boosting.
 5. Explain Voting and Stacking methods.
 6. Explain ECOC with applications.
-

PYQ-BASED EXPECTED QUESTIONS

★ High Probability

- Explain Ensemble Learning.
 - Explain Random Forest Trees.
 - Explain AdaBoost Algorithm.
 - Differentiate Bagging and Boosting.
 - Explain Voting.
-

★ Medium Probability

- Explain Stacking.
 - Explain ECOC.
 - Explain Model Combination Schemes.
-

ONE-NIGHT REVISION

- ✓ Ensemble Learning combines multiple models
- ✓ Voting selects majority prediction
- ✓ Bagging works parallelly

- ✓ Random Forest uses multiple decision trees
 - ✓ Boosting works sequentially
 - ✓ AdaBoost focuses on wrong predictions
 - ✓ Stacking uses meta learner
 - ✓ ECOC converts multi-class into binary problems
 - ✓ Bagging reduces variance
 - ✓ Boosting reduces bias
-

SMART STUDY PLAN

FIRST STUDY

1. Ensemble Learning
 2. Random Forest
 3. AdaBoost
 4. Bagging vs Boosting
-

SECOND PRIORITY

5. Voting
 6. Stacking
 7. ECOC
-

LAST REVISION

8. Applications
 9. Advantages & Disadvantages
-

FINAL EXAM TIP

For 14-mark answers always write:

1. Definition
2. Working
3. Example
4. Advantages
5. Disadvantages
6. Applications
7. Conclusion

This presentation style helps score higher marks in RGPV exams.