Om Shete

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Assignment No: 2

Q.1 Discuss different ensembles techniques.

⇒

- A ensemble is machine learning model that combines the predictions from two or more models.
- The models that contribute for the ensembles are called as ensemble members.
- They may be of the same type or of different types.
- Different Ensemble techniques :-

i) Bagging :

⇒ Bagging involves training multiple instances of the same learning algorithm on different subsets of the data and then averaging the predictions or taking a majority vote.

- Random forest is a widely used bagging technique where multiple decision trees are trained, and the final prediction is based on the majority vote or average.
- Advantages :
  i) Reduces variance
  ii) Helps prevent overfitting
- Disadvantage :

i) Computationally expensive as it requires training multiple models.

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

2) Boosting

⇒ Boosting works by training models sequentially, with each new models focusing on the errors made by the previous models.

- The predictions are combined in a weighted manner, where the models that perform better are given higher importance.
- For eg., AdaBoost focuses on misclassified instances and adjusts weights accordingly for the next learner.
- Advantages :
  i) Reduces both bias and variance.
  ii) Effective for improving the performance of weak models.
- Disadvantages :
  i) Prone to overfitting
  ii) Computationally intensive.

3) Voting

⇒ In voting ensemble, multiple models are trained, and their predictions are combined by taking vote.

- There are two types of voting :-

a) Hard voting : Majority vote is taken for classification.

b) Soft voting : Probabilities or confidences scores are averaged to make the final prediction.

- For e.g.,
A simple voting classifiers might combine logistic regression, decision trees and SVMs and take the majority vote of their predictions

- Disadvantages:
i) Requires the base model to be resonably good.

- Advantages:

i) Easy to implement and interpret.
ii) Reduces the risk of choosing the wrong model.

Q.2 Explain following terms :

i) Weak learners and Strong learners

⇒
- **Weak learners**

⇒ A weak learner is a machine learning model that performs slightly better than random guessing, meaning its accuracy is marginally above 50% for binary classification problems.
- Individually, weak learners have low predictive accuracy.
- Weak learners often underfit the data and fail to capture the complexity of the underlying patterns.
- For e.g., A decision stump is often considered a weak learner.

- **Strong learners**

⇒ A strong learner is a model that has high predictive accuracy and is capable of making highly accurate position on its own.
- Strong learners have low bias and low error rates.
- Strong learners are often more complex models like deep decision trees, neural networks or gradient-boosted models.
- Strong learners are capable of fitting the training data well while also generalizing

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

for new, unseen data.
- For e.g., A fully grown decision tree or a well-tuned deep learning model can be considered strong learners.

2) Meta learning
⇒

- Meta learning, often referred to as "learning to learn", involves algorithms that improve their learning processes over time by using past experiences or learning patterns.
- It focuses on learning at a higher level than conventional machine learning by adjusting the learning process itself based on previous outcomes.
- Characteristics :

i) The system can adapt quickly to new tasks by leveraging prior knowledge ~~for~~ patterns learned from other tasks.
ii) Meta-learning models can achieve good performance with fewer data points or less computational effort by learning how to learn effectively.

- Types of Meta learning

i) Model-based Meta learning: Involve learning an internal model that can quickly adapt

THADOMAL SHAHANI TSEC ENGINEERING COLLEGE

to new tasks with limited fine-tunning.

ii) Metric-based Meta learning: The algorithm learns a similarity metric that allows it to generalize well to new tasks with limited fine-tunning.

iii) Optimization-based Meta learning: Focuses on improving the optimization process itself, such as by learning better updates rules for gradient descent.

- For e.g.,

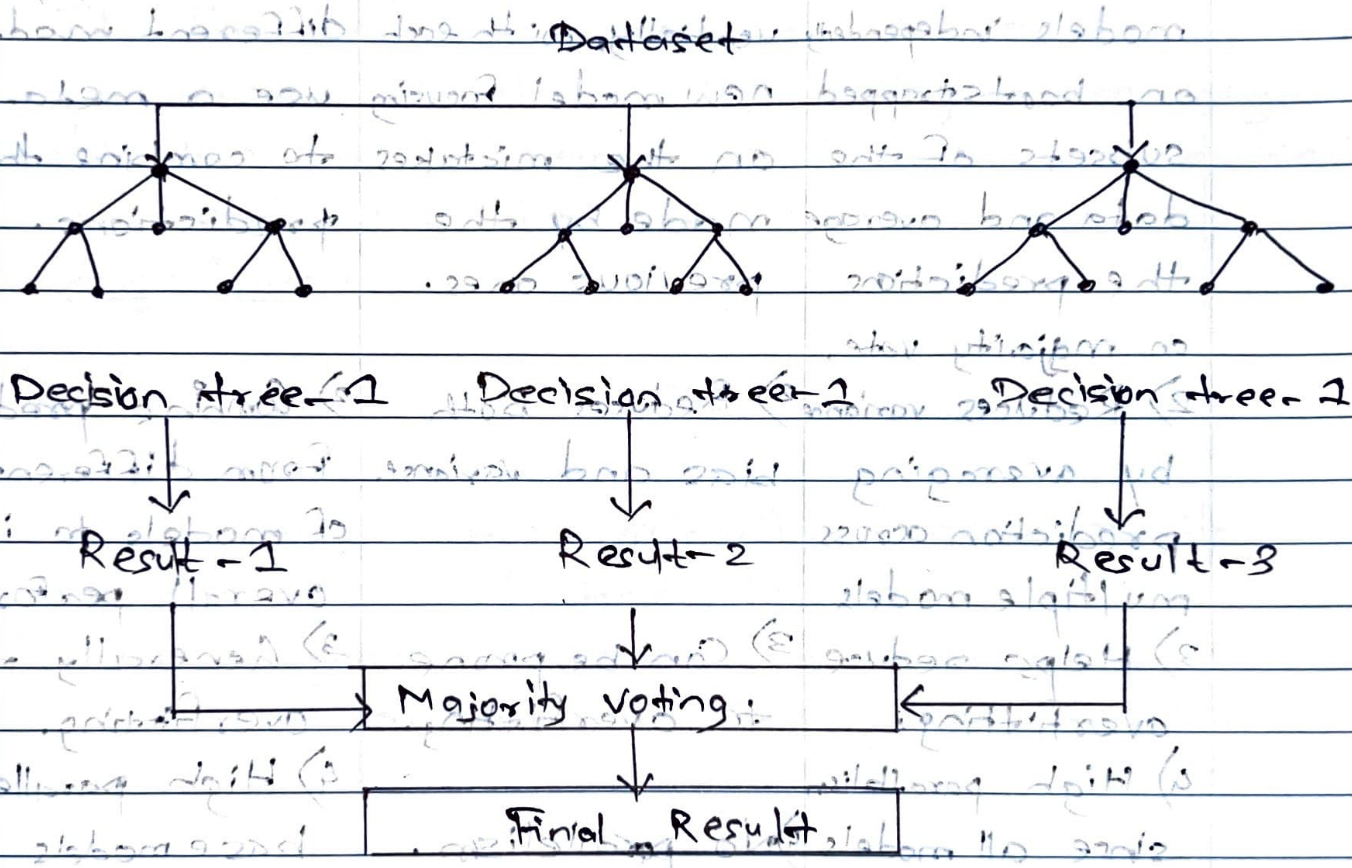
Meta-learning is commonly used in few-shot learning scenarios, where a model is trained to adapt to new tasks with only a few training examples.

3) Random Forest

=>

- Random forest is a supervised machine learning algorithm that is used widely in classification and Regression problems.
- It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

- Random Forest algorithm can handle the dataset containing continuous variables in case of regression and categorical variables in case of classification.
- It performs better result for classification problem.



- Advantages:-

1) It is robust to noise and works well with large datasets and a high number of features.
2) Random Forest can be used to assess the relative importance of features for the prediction.

Q.3. Compare Bagging, Boosting and stacking techniques.

→

Bagging	Boosting	Stacking
1) Train multiple models independently on bootstrapped subsets of the data and average the predictions on majority vote.	1) Train models sequentially, with each new model focusing on the mistakes made by the previous ones.	1) Train multiple different models and use a meta-learner to combine their predictions.
2) Reduces variance by averaging prediction across multiple models	2) Reduces both bias and variance.	2) Combine prediction from different types of models to improve overall performance.
3) Helps reduce overfitting.	3) Can be prone to overfitting.	3) Generally reduces overfitting.
4) High parallelism since all models are trained independently.	4) No parallelism.	4) High parallelism, base models are trained independently
5) Models are relatively strong individually	5) Models are weak individually.	5) Base models can Be both strong or weak.
6) For e.g., Random Forest, Bagged decision tree	6) For e.g. AdaBoost, XGBoost.	6) For e.g., stacking with different base learners.

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Q.4 Explain AdaBoost Algorithm
=)

- AdaBoost (Adaptive Boosting) algorithm is one of the most popular boosting techniques that combines multiple weak learners to create strong learners.
- It focuses on improving the performance of weak classifier by giving more weight to the data points that are misclassified in each iteration.
- Algorithm:

1) Weight Initialization:

- At every start, schooling instances is assigned an identical weights.
- These weights determine the importance of every example in the getting to know method.

2) Model Training:

- A weak learner is skilled at the dataset, with the aim of minimising class errors.
- A weak learner is usually easy model, which includes a selection stump or a small neural network.

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

3) Weighted Error Calculation:

- After the vulnerable learner is skilled, its miles are used to make predictions at the education dataset.
- The weighted mistakes are then calculated by means of summing up the weights of the misclassified items.

4) Model weight calculation:

- The weight of susceptible learner is calculated primarily biased on their performance in classifying the training data.
- Models that perform properly are assigned higher weights, indicating that they're more reliable.

5) Update instance weight:

- The example weights are updated to offer more weight to the misclassified samples from the previous step.
- The adjustment focuses on the studying method at the times that the present day model struggles with.

6) Repeat :

- Steps 2 through 5 are repeated for a predefined variety of iterations or till a distinctive overall performance threshold is met.

7) Final Model Creation :

- The very last study model is created by means of combining the weighted outputs of all weak newcomers.

(A)

27/9/24