Lesson 6: Decision Trees and Ensemble Learning — Complete Reviewer

1 What is a Decision Tree?

A ***Decision Tree*** helps us to make decisions by mapping out different choices and their possible outcomes. It's used in machine learning for tasks like ***classification and prediction***. In this article, we'll see more about Decision Trees, their types and other core concepts.

1.1 How It Works

- ***Root node*** asks: "Is age > 30?"
- Branch left if yes, right if no
- Each node asks another question
- ***Leaf nodes*** provide the final prediction

1.2 Real Example

Predict if a customer will buy a product based on age, income, and browsing history—the tree learns patterns and ***makes decisions automatically***.

2 Why Decision Trees Matter

- ***Intuitive & Transparent***: Easy to visualize and understand—even non-experts can follow the logic and explain predictions.
- ***Proven Algorithms***: ***ID3, C4.5, and CART*** are industry-standard methods with decades of reliable performance.
- ***Versatile Applications***: Work for classification (spam detection) and regression (house price prediction) across domains.

3 Core Components & Terminology

- ***Root Node***: Starting point representing the whole dataset.
- ***Branches***: Lines connecting nodes showing the flow from one decision to another.
- ***Internal Nodes***: Points where decisions are made based on data features.
- ***Leaf Nodes***: End points of the tree where the ***final decision or prediction*** is made.
- ***Pruning***: ***Removal of subnodes*** at the end of every node.

4 Types of Decision Trees

1. ***Classification Trees***: Used for predicting ***categorical outcomes*** like spam or not spam. These trees split the data based on features to classify data into predefined categories.

2. ***Regression Trees***: Used for predicting ***continuous outcomes*** like predicting house prices. Instead of assigning categories, it provides ***numerical predictions*** based on the input features.

5 How Decision Trees Work (Step-by-Step)

1. ***Start with the Root Node***: It begins with a main question at the root node which is derived from the dataset's features.

2. ***Ask Yes/No Questions***: From the root, the tree asks a series of yes/no questions to split the data into subsets based on specific attributes.

3. ***Branching Based on Answers***: Each question leads to different branches:

   - If the answer is yes, the tree follows one path.
   - If the answer is no, the tree follows another path.

4. ***Continue Splitting***: This branching continues through further decisions helps in ***reducing the data down step-by-step***.

5. ***Reach the Leaf Node***: The process ends when there are no more useful questions to ask leading to the leaf node where the ***final decision or prediction*** is made.

6 Advantages and Disadvantages

6.1 Advantages

- ***Easy to Understand***: Decision Trees are visual which makes it easy to follow the decision-making process.

- ***Versatility***: Can be used for both ***classification and regression*** problems.

- ***No Need for Feature Scaling***: Unlike many machine learning models, it don't require us to ***scale or normalize*** our data.

- ***Handles Non-linear Relationships***: It capture complex, ***non-linear relationships*** between features and outcomes effectively.

- ***Interpretability***: The tree structure is easy to interpret helps in allowing users to understand the reasoning behind each decision.

6.2 Disadvantages

- ***Overfitting***: They can overfit the training data if they are too deep which means they ***memorize the data*** instead of learning general patterns.

- ***Instability***: It can be unstable which means that ***small changes in the data*** may lead to significant differences in the tree structure and predictions.

- ***Bias towards Features with Many Categories***: It can become biased toward features with many distinct values which focuses too much on them and potentially ***missing other important features***.

7 Ensemble Learning

Ensemble learning is a method where we use many small models instead of just one. Each of these models may not be very strong on its own, but when we put their results together, we get a ***better and more accurate answer***. It's like asking a ***group of people for advice*** instead of just one person.

7.1 Types of Ensemble Learning

- ***Bagging (Bootstrap Aggregating)***: Models are ***trained independently*** on different random subsets of training data. Results are combined by ***averaging (regression) or voting (classification)***.
- ***Boosting***: Models are ***trained one after another***. Each new model focuses on ***fixing the errors*** made by the previous ones.

8 How Bagging Works

Bootstrap Aggregating (Bagging) is designed to improve the ***stability and accuracy*** of machine learning algorithms.

1. ***Step 1***: Multiple subsets are created from the original data set with ***equal tuples***, selecting observations ***with replacement***.
2. ***Step 2***: A ***base model*** is created on each of these subsets.
3. ***Step 3***: Each model is learned in ***parallel*** with each training set and independent of each other.
4. ***Step 4***: The final predictions are determined by ***combining the predictions*** from all the models.

9 Bagging and Random Forests

Random Forest is a machine learning algorithm that uses many decision trees to make better predictions.

- ***Bagging***: Trains many trees on random subsets of data, ***reducing variance*** through averaging.
- ***Random Forests***: Adds ***feature randomness***—each split considers only ***random feature subsets***, maximizing diversity among trees.

9.1 How Random Forest Works (Steps)

1. ***Create Many Decision Trees***: The algorithm makes many decision trees each using a random part of the data. Every tree is a bit different.
2. ***Pick Random Features***: When building each tree it doesn't look at all the features at once. It picks a ***few at random*** to decide how to split the data.
3. ***Each Tree Makes a Prediction***: Every tree gives its own answer or prediction based on what it learned.
4. ***Combine the Predictions***: For classification we choose via ***majority voting***; for regression we use the ***average*** of all tree predictions.
5. ***Why It Works Well***: Using random data and features for each tree helps avoid ***overfitting*** and makes the overall prediction more accurate and trustworthy.

10 Real-World Impact: Ensemble Learning in Action

- ***Financial Services***: Random Forests predict ***credit risk*** for millions of loan applicants, reducing defaults and protecting lenders.
- ***Healthcare***: ***Gradient Boosting*** models diagnose diseases from medical images with expert-level accuracy and consistency.
- ***E-Commerce***: Ensemble methods ***personalize recommendations***, boosting sales and customer satisfaction through intelligent predictions.

11 Boosting Algorithm

Boosting combines multiple weak learners to create a strong learner. Weak models are ***trained in series*** such that each next model tries to correct errors of the previous model.

11.1 Steps in Boosting

1. ***Initialize Model Weights***: Begin with a single weak learner and assign ***equal weights*** to all training examples.
2. ***Train Weak Learner***: Train weak learners on these datasets.
3. ***Sequential Learning***: Boosting works by training models ***sequentially*** where each model focuses on ***correcting the errors of its predecessor***.
4. ***Weight Adjustment***: Boosting assigns weights to training datapoints. ***Misclassified examples receive higher weights*** in the next iteration.

12 Comparison of Techniques

Technique	*Category*	Description
Random Forest	Bagging	Random forest constructs multiple decision trees on ***bootstrapped subsets*** of the data and aggregates their predictions for final output, ***reducing overfitting and variance***.
Random Subspace Method	Bagging	Trains models on ***random subsets of input features*** to enhance diversity and improve generalization while reducing overfitting.
Gradient Boosting Machines (GBM)	Bagging	Gradient Boosting Machines ***sequentially builds decision trees***, with each tree correcting errors of the previous ones, ***enhancing predictive accuracy iteratively***.
Extreme Gradient Boosting (XGBoost)	Bagging	XGBoost do ***optimizations like tree pruning, regularization***, and ***parallel processing*** for robust and efficient predictive models.
AdaBoost (Adaptive Boosting)	Bagging	AdaBoost ***focuses on challenging examples*** by assigning weights to data points. Combines weak classifiers with ***weighted voting*** for final predictions.
CatBoost	Bagging	CatBoost ***specialize in handling categorical features natively*** without extensive preprocessing with high predictive accuracy and automatic overfitting handling.

13 Algorithm Applications Table

Algorithm	Used For	*Typical Application*
Random Forest	Classification, Regression	***Fraud detection, churn prediction***
Random Subspace Method	Feature diversity, ensemble models	***Image/text recognition***
Gradient Boosting (GBM)	Sequential boosting	***Risk scoring, price prediction***
XGBoost	Fast, regularized boosting	***Financial forecasting, competitions***
AdaBoost	Weighted weak learners	***Face detection, spam filtering***
CatBoost	Categorical data	***E-commerce, recommendation systems***