



Artificial Intelligence and Machine Learning (ACCA6011)

Decision Trees and Search Algorithms Part 1 : Decision Tree

Dr. Seena Joseph

Decision Tree

Learning Problems in AI

Regression – Prediction of real value

Example : Predict the stock value for tomorrow based on the past performance;
Predict Student's Exam mark based on the internal assessment marks

Binary Classification : Prediction of Yes/ No response

Example : Predict whether John will enjoy a course or no
Predict whether a user review of the latest Apple product is positive or negative

Multiclass classification : Prediction of an instance into one of several classes

Example: Predict whether a news story is about entertainment, sports, politics or region

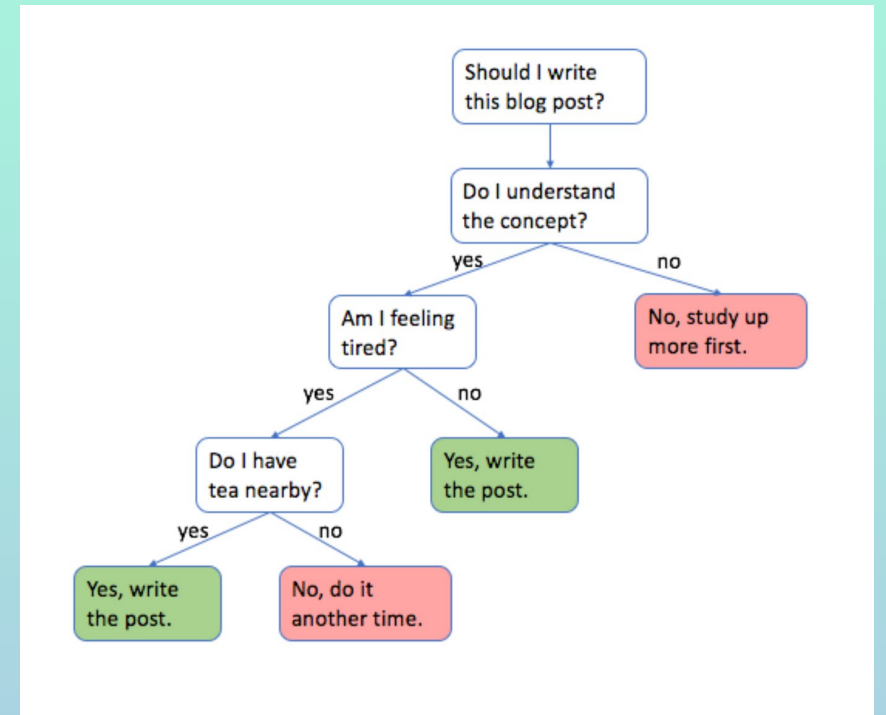
Ranking : Trying to put a set of objects in order of relevance.

Example : Predicting what order to put web pages, in response to a user query

Activity: Identify more example for each of the learning problem

Decision Trees

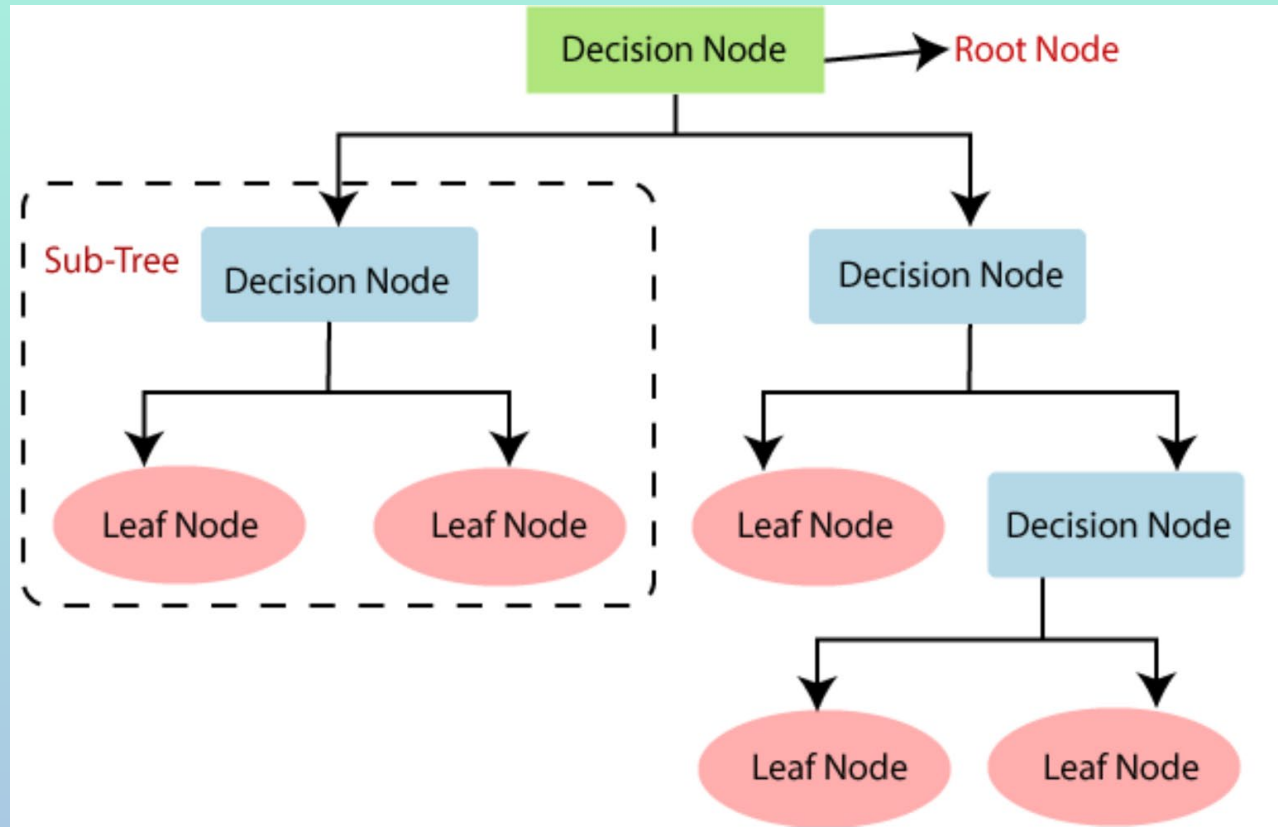
- A classic and natural model of learning
- Decision tree -> divide and conquer



- Activity : More examples to create binary tree

Decision Tree

- General structure of a Decision Tree



Source [2]

Decision Trees

- Components

- Root node: the topmost node of the tree where the decision-making process starts based
- Internal node : Represent the decision based on a feature or attribute
- Leaf nodes : Represent the final output or prediction node
- Parent and child node : base node is the parent node and the remaining nodes are child node.
- Branches : A subtree formed by splitting the tree ; outcome of a decision
- Pruning – process of removing unwanted branches from the tree
- Splitting – process of dividing the decision node/ root node into sub-nodes based on the given condition

Decision Trees

Attribute/Features – characteristics or property of dataset that is used to make decisions at each node of the tree.

Each attribute corresponds to a question that helps to split the data into subsets.

Type of features

1. Numerical features

Continuous or discrete values (age, salary, temperature)

Numerical features are used to split the tree by comparing against a threshold.. Eg is **Age > 50**

2. Categorical feature

These are discrete, non-numerical values that represent categories or classes. Example is the weather sunny, rainy or cloudy ? Split of categorical features are by checking the membership in a set of values (is the weather sunny ?)

Decision Trees

Role of feature in splitting

A feature is used to split the data-based on a specific condition

Example: is **Age** > 50 ? or Is **Salary** high ?

The purpose of feature selection is split the data into homogenous subsets, reducing impurity.

Feature selection – The decision tree algorithm evaluate the multiple features at each node and selects the one with the highest information gain or less impurity.

Decision Trees

Attribute Selection Measures (ASM)

How to select the best attribute for the root node and sub nodes.

ASM involves the collection of optimum attributes concerning the source node as well as for the sub-nodes.

The goal of **ASM** is to find the features that split the data into subsets that are homogenous/pure as possible.

Best practices for ASM are:

1. Information gain
2. Gini index

Decision Trees

Information gain- Calculates the amount of information that is provided by a feature regarding the class.

Information Gain = Entropy (S) – [(Weighted Average)*Entropy (Every Feature)]

Entropy – is the measure of impurity or disorder or uncertainty;

Entropy values : range from 0 to 1; less the entropy value better the outcome;

Entropy controls how a decision tree decides to split the data.

Decision Trees

How to calculate entropy ?

$$\text{Entropy}(s) = -P(\text{Yes})\log_2 P(\text{Yes}) - P(\text{No})\log_2 P(\text{No})$$

where $P(\text{Yes}) \rightarrow$ probability/Chance of getting yes

$P(\text{No}) \rightarrow$ probability/Chance of getting No

Decision Trees

Gini Index – Measures the impurity or purity used during the creation of a decision tree algorithm.

$$Gini(D) = 1 - \sum_{i=1}^k p_i^2$$

Where p_i is the probability of choosing a sample of class i .
K is the number of classes

A lower Gini Index represents the better split

Decision Trees

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Decision Trees

Example

Dataset: Weather and Decision to Play

Weather	Windy	Play?
Sunny	No	No
Sunny	Yes	No
Overcast	No	Yes
Rainy	No	Yes
Rainy	Yes	No
Rainy	No	Yes
Overcast	Yes	Yes
Sunny	No	Yes

- **Feature 1** : Weather (Sunny, Overcast, Rainy)
- **Feature 2**: Windy (Yes, No)
- **Target**: Play? (Yes, No)

Decision Trees

Example for calculating information gain

Number of instances for Yes or No for the Target variable, Play.

Yes = 5

No = 3

Step 1:

Entropy for the whole dataset, S

$$\begin{aligned}\text{Entropy}(S) &= -P(\text{Yes})\log_2 P(\text{Yes}) - P(\text{No})\log_2 P(\text{No}) \\ &= -5/8\log_2(5/8) - (3/8)\log_2(3/8) \\ &= -(0.625\log_2 0.625 + 0.375\log_2 0.375) \\ &= -(0.625 \times -0.678 + 0.375 \times -1.415) = 0.954\end{aligned}$$

Decision Trees

Step 2:

Calculate the entropy for Weather

Sunny; Overcast; Rainy

Sunny = 3 instances with 2 No and 1 Yes (entropy =0.918)

Overcast = 2 instances ; 2 Yes (entropy =0)

Rainy = 3 instances with 2 No and 1 Yes ; (entropy =0.918)

Weighted entropy : $\frac{3}{8} \times 0.918 + \frac{2}{8} \times 0 + \frac{3}{8} \times 0.918 = 0.688$

Step 3: entropy (s) – entropy (weather)

= 0.954 - 0.688 = **0.266**

Decision Trees

Example Gini Index

Number of instances for Yes or No for the Target variable, Play.

Yes = 5

No = 3

Step 1:

$$\begin{aligned} \textit{Gini}(S) &= 1 - \left(\left(\frac{5^2}{8} \right) + \left(\frac{3^2}{8} \right) \right) \\ &= 1 - (0.3906 + 0.1406) = 1 - 0.5312 = 0.4688 \end{aligned}$$

Decision Trees

Example Gini Index

Step 2 : Calculate Gini Index for Splitting on "Weather"

Sunny : 3 instances, 2 "No", 1 "Yes" (Gini = 0.444).

Overcast: 2 instances, 2 "Yes" (Gini = 0).

Rainy: 3 instances, 2 "Yes", 1 "No" (Gini = 0.444).

Calculate the weighted Gini Index

$$= 3/8 \times 0.444 + 2/8 \times 0 + 3/8 \times 0.444 = 0.333$$

Decision Trees

Example interpretation

Information Gain for splitting on "Weather" is **0.266**. This shows that splitting on this feature reduces uncertainty (entropy), but the reduction is moderate.

Gini Index after the split is **0.333**, indicating that the groups formed by this split are somewhat pure, though not entirely.

Decision Trees

Practical Demo

Decision Tree

- **Pruning** is a process of deleting the unnecessary nodes from a tree to get the optimal decision tree.
- It actually minimizes the tree size without hampering its accuracy. It is of two types, **cost complexity**, and **error reduction** pruning.

Reference

[1] M. Bansal, A. Goyal and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," Decision Analytics Journal, vol. 3, 2022.DOI: 10.1016/j.dajour.2022.100071.

[\[2\] Decision Tree Algorithm in Machine Learning – Javatpoint](#)