Decision Tree :

Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

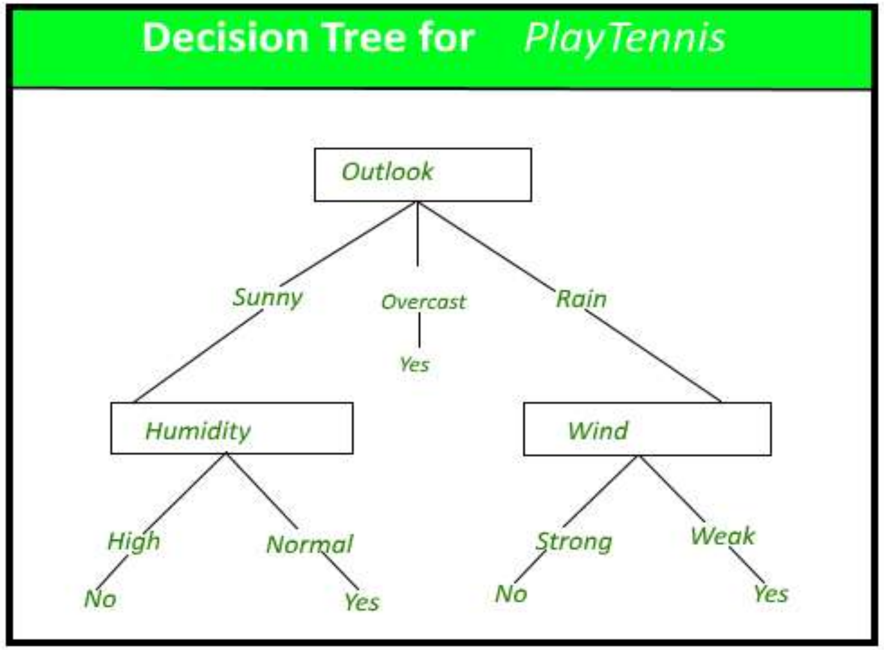


Fig. 1: A decision tree for the concept Play Tennis.

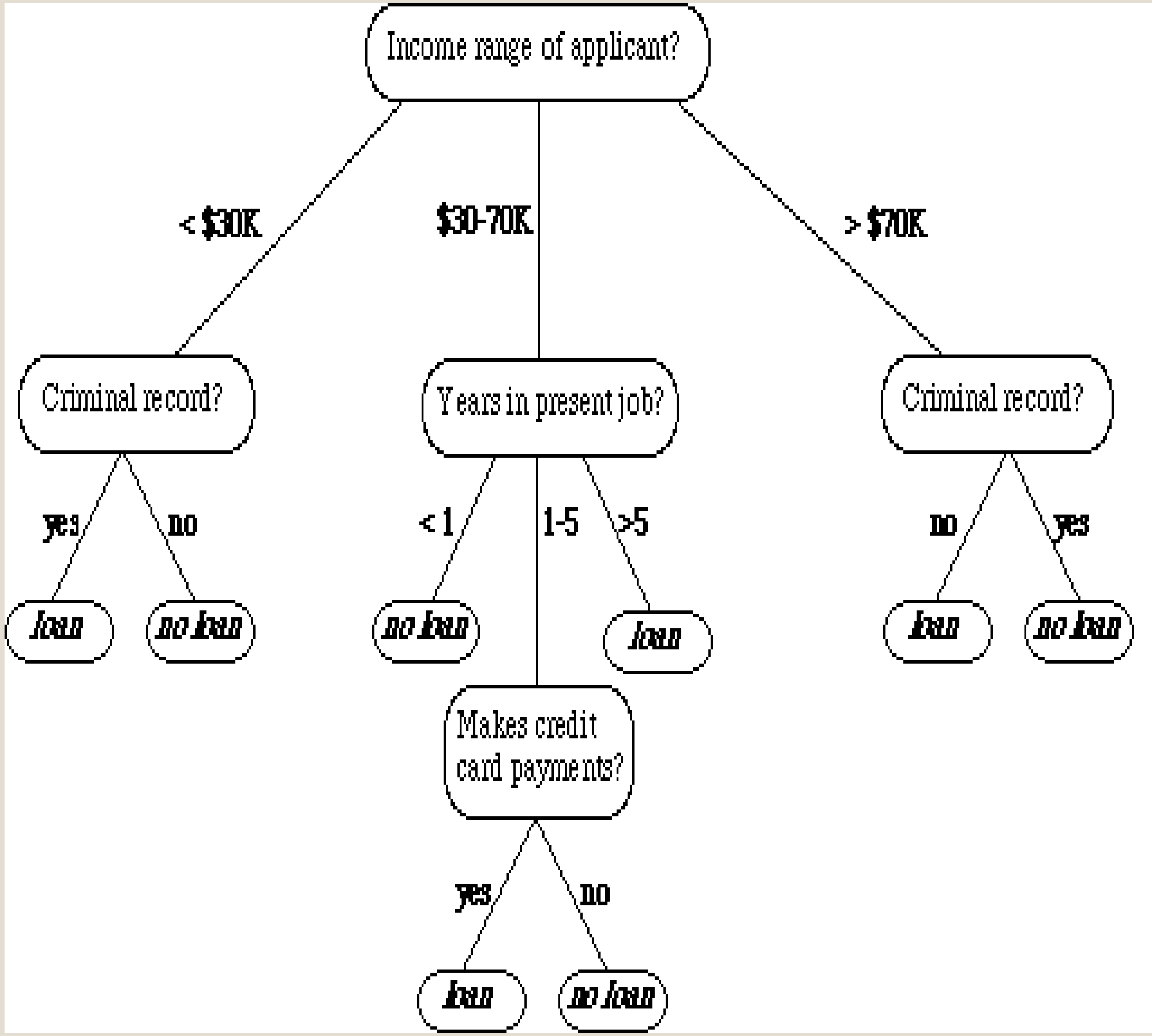
The strengths of decision tree methods are:

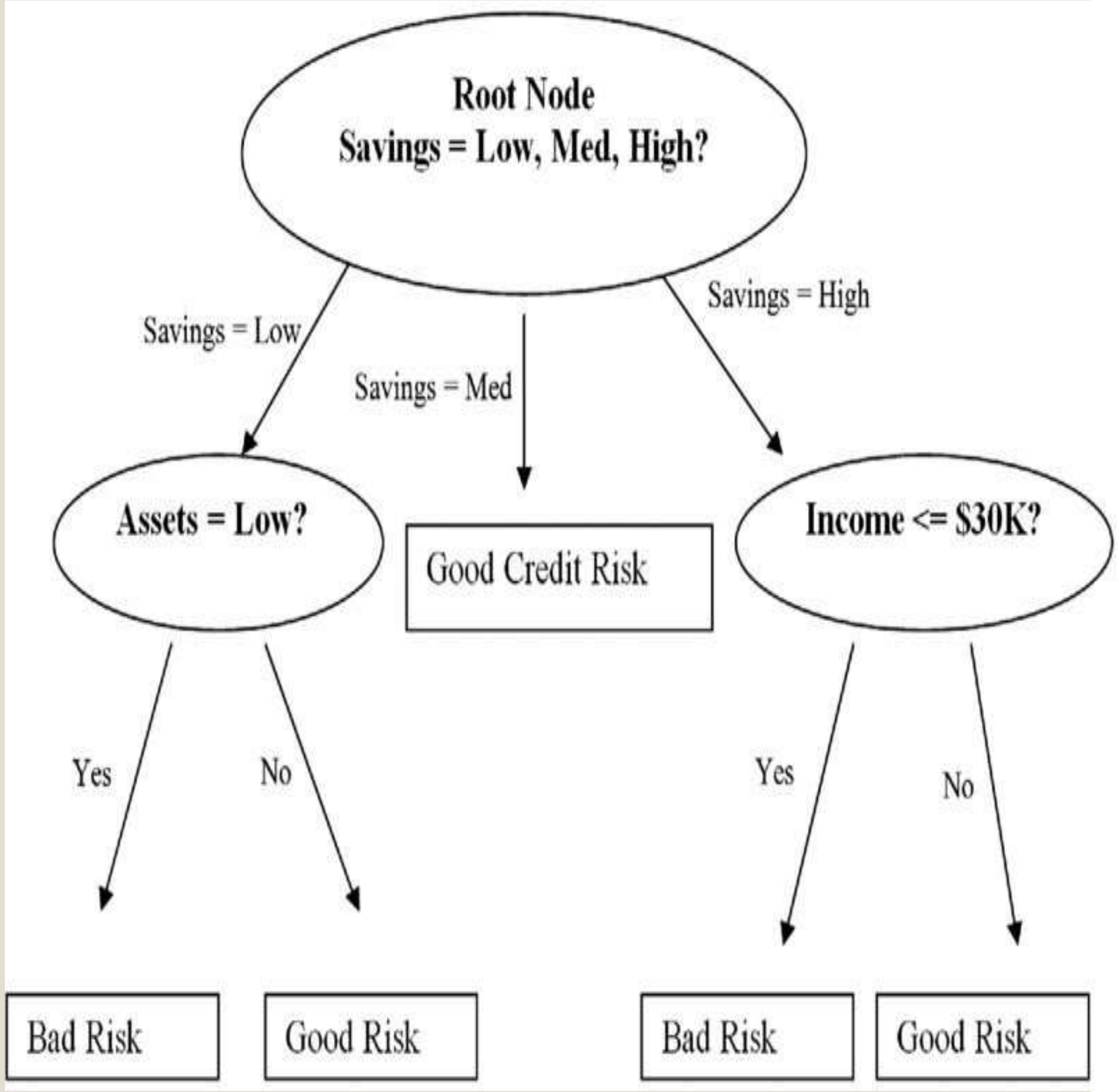
- Decision trees are able to generate understandable rules.
- Decision trees perform classification without requiring much computation.
- Decision trees are able to handle both continuous and categorical variables.
- Decision trees provide a clear indication of which fields are most important for prediction or classification.

The weaknesses of decision tree methods :

- Decision trees are less appropriate for estimation tasks where the goal is to predict the value of a continuous attribute.
- Decision trees are prone to errors in classification problems with many class and relatively small number of training examples.
- Decision tree can be computationally expensive to train.

Decision Tree Introduction with example:





In Decision Tree the major challenge is to identification of the attribute for the root node in each level. This process is known as *attribute selection*. We have two popular attribute selection measures:

- Information Gain
- Gini Index

1. Information Gain

When we use a node in a decision tree to partition the training instances into smaller subsets the entropy changes. Information gain is a measure of this change in entropy.

Entropy:

Entropy is the measure of uncertainty of a random variable, it characterizes the impurity of an arbitrary collection of examples. The higher the entropy more the information content.

Example:

```
For the set X = {a,a,a,b,b,b,b,b}

Total intances: 8

Instances of b: 5

Instances of a: 3
```

$$Entropy H(X) = -\left[\left(\frac{3}{8}\right)log_2\frac{3}{8} + \left(\frac{5}{8}\right)log_2\frac{5}{8}\right]$$

```
= -[0.375 * (-1.415) + 0.625 * (-0.678)]

=-(-0.53-0.424)

= 0.954
```

GINI Index:

Gini index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen. But what is actually meant by ‘impurity’? If all the elements belong to a single class, then it can be called pure. The degree of Gini index varies between 0 and 1, where 0 denotes that all elements belong to a certain class or if there exists only one class, and 1 denotes that the elements are randomly distributed across various classes. A Gini Index of 0.5 denotes equally distributed elements into some classes.

Formula for Gini Index:

$$\text{Gini} = 1 - \sum_{i=1}^{n} (p_i)^2$$

where p_i is the probability of an object being classified to a particular class.

Example of GINI Index:

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Negative	High	Low	Down
Positive	Low	High	Up
Positive	High	High	Up
Negative	Low	High	Down
Positive	Low	Low	Down
Negative	High	High	Down
Negative	Low	High	Down
Positive	Low	Low	Down
Positive	High	High	Up

Let’s start by calculating the Gini Index for ‘Past Trend’.

P(Past Trend=Positive): 6/10

P(Past Trend=Negative): 4/10

If (Past Trend = Positive & Return = Up), probability = 4/6

If (Past Trend = Positive & Return = Down), probability = 2/6

Gini index = 1 - ((4/6)^2 + (2/6)^2) = 0.45

If (Past Trend = Negative & Return = Up), probability = 0

If (Past Trend = Negative & Return = Down), probability = 4/4

Gini index = 1 - ((0)^2 + (4/4)^2) = 0

Weighted sum of the Gini Indices can be calculated as follows:

Gini Index for Past Trend = (6/10)*0.45 + (4/10)*0 = 0.27

**By this way we can calculate GINI Index for ‘Open Interest’, ‘Trading Volume ’.

Gini Index is preferred over Information gain:

Gini Index, unlike information gain, isn’t computationally intensive as it doesn’t involve the logarithm function used to calculate entropy in information gain, which is why Gini Index is preferred over Information gain.

Decision Tree Regression:

Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.

Discrete output example: A weather prediction model that predicts whether or not there’ll be rain in a particular day.

Continuous output example: A profit prediction model that states the probable profit that can be generated from the sale of a product.

Criteria	Logistic Regression	Decision Tree Classification
Interpretability	Less interpretable	More interpretable
Decision Boundaries	Linear and single decision boundary	Bisects the space into smaller spaces
Ease of Decision Making	A decision threshold has to be set	Automatically handles decision making
Robustness to noise	Robust to noise	Majorly affected by noise
Scalability	Requires a large enough training set	Can be trained on a small training set

Thanks to the audience .