DECISION TREES

Inspire...Educate...Transform.

BY

International School of Engineering

{We Are Applied Engineering}

Disclaimer: Some of the Images and content have been taken from multiple online sources and this presentation is intended only for knowledge sharing but not for any commercial business intention

OVERVIEW

- DEFINITION OF DECISION TREE
- WHY DECISION TREE?
- DECISION TREE TERMS
- EASY EXAMPLE
- CONSTRUCTING A DECISION TREE
- CALCULATION OF ENTROPY
- ENTROPY
- TERMINATION CRITERIA
- PRUNING TREES
- APPROACHES TO PRUNE TREE
- DECISION TREE ALGORITHMS
- LIMITATIONS
- ADVANTAGES
- VIDEO OF CONSTRUCTING A DECISION TREE

DEFINITION OF ‘DECISION TREE'

- A decision tree is a natural and simple way of inducing following kind of rules.

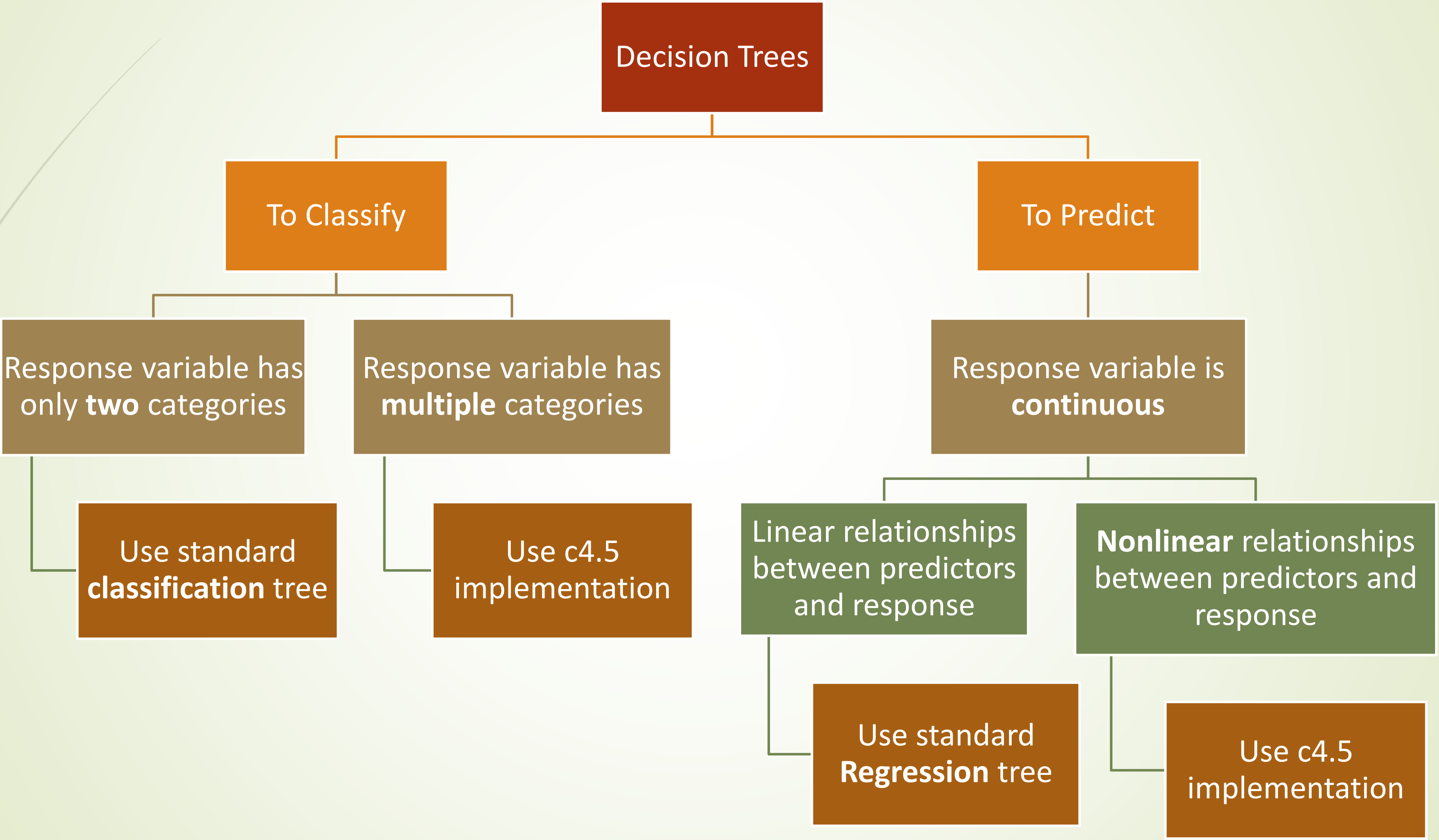
  If (Age is x) and (income is y) and (family size is z) and (credit card

  spending is p) then he will accept the loan

- It is powerful and perhaps most widely used modeling technique of all

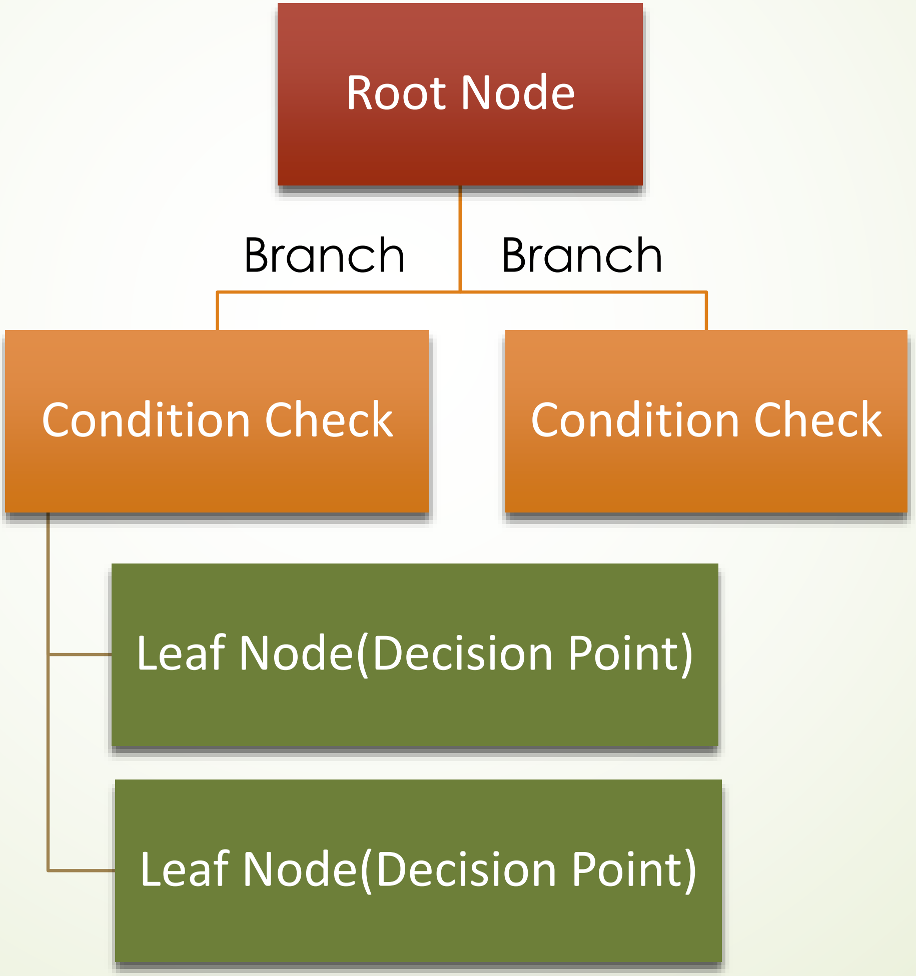
- Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance

WHY DECISION TREE?


Decision Trees
To Classify
To Predict
Response variable has only **two** categories
Response variable has **multiple** categories
Response variable is **continuous**
Use standard **classification** tree
Use c4.5 implementation
Linear relationships between predictors and response
Nonlinear relationships between predictors and response
Use standard **Regression** tree
Use c4.5 implementation

Source: http://www.simafore.com/blog/bid/62482/2-main-differences-between-classification-and-regression-trees

DECISION TREE TERMS



Root Node

Branch

Branch

Condition Check

Condition Check

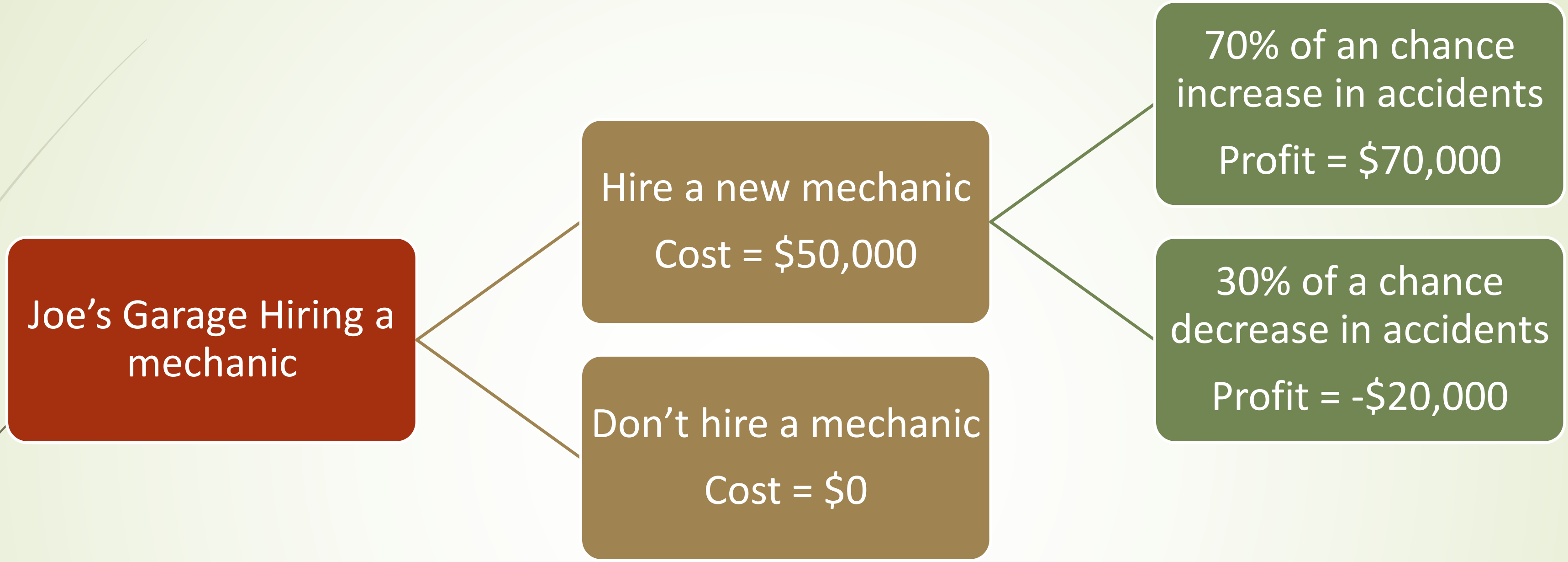
Leaf Node(Decision Point)

Leaf Node(Decision Point)

EASY EXAMPLE

- Joe’s garage is considering hiring another mechanic.
- The mechanic would cost them an additional $50,000 / year in salary and benefits.
- If there are a lot of accidents in Iowa City this year, they anticipate making an additional $75,000 in net revenue.
- If there are not a lot of accidents, they could lose $20,000 off of last year’s total net revenues.
- Because of all the ice on the roads, Joe thinks that there will be a 70% chance of “a lot of accidents” and a 30% chance of “fewer accidents”.
- Assume if he doesn’t expand he will have the same revenue as last year.

continued



Joe’s Garage Hiring a mechanic
Hire a new mechanic
Cost = $50,000
Don’t hire a mechanic
Cost = $0
70% of an chance increase in accidents
Profit = $70,000
30% of a chance decrease in accidents
Profit = -$20,000

- Estimated value of “Hire Mechanic” = NPV =.7(70,000) + .3(- $20,000) - $50,000 = - $7,000
- Therefore you should not hire the mechanic

CONSTRUCTING A DECISION TREE

Two Aspects

- Which attribute to choose?
  - Information Gain
    - ENTROPY
- Where to stop?
  - Termination criteria

CALCULATION OF ENTROPY

- Entropy is a measure of uncertainty in the data

  $\text{Entropy}(S) = \sum_{(i=1 \text{ to } l)} -|S_i|/|S| * \log_2(|S_i|/|S|)$

  - S = set of examples
  - S_i = subset of S with value v_i under the target attribute
  - l = size of the range of the target attribute

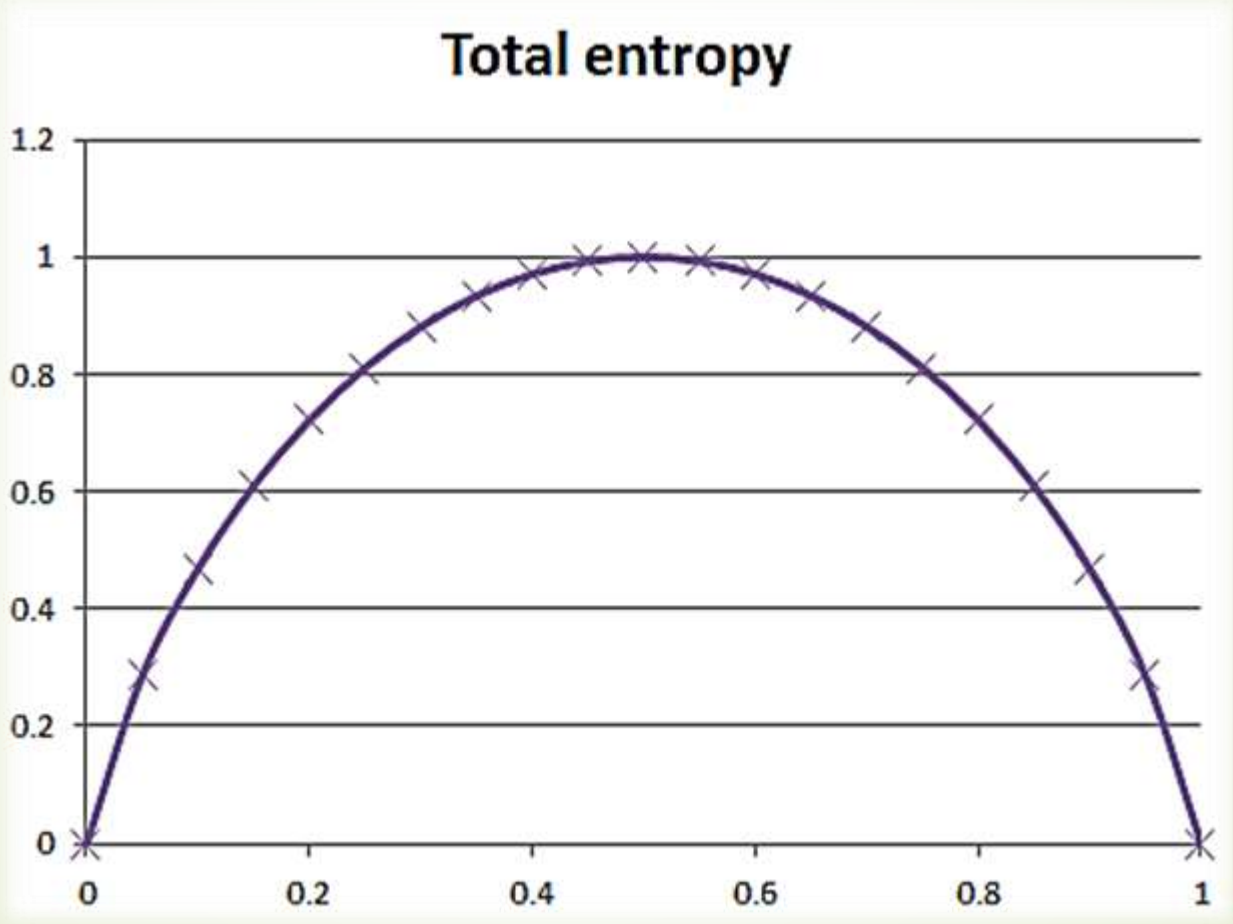
ENTROPY

- Let us say, I am considering an action like a coin toss. Say, I have five coins with probabilities for heads 0, 0.25, 0.5, 0.75 and 1. When I toss them which one has highest uncertainty and which one has the least?

$$H = - ipi \log 2\ pi$$

- Information gain = Entropy of the system before split – Entropy of the system after split

ENTROPY: MEASURE OF RANDOMNESS



TERMINATION CRITERIA

- All the records at the node belong to one class
- A significant majority fraction of records belong to a single class
- The segment contains only one or very small number of records
- The improvement is not substantial enough to warrant making the split

PRUNING TREES

- The decision trees can be grown deeply enough to perfectly classify the training examples which leads to overfitting when there is noise in the data
- When the number of training examples is too small to produce a representative sample of the true target function.
- Practically, pruning is not important for classification

APPROACHES TO PRUNE TREE

- Three approaches
  - –Stop growing the tree earlier, before it reaches the point where it perfectly classifies the training data,
  - –Allow the tree to over fit the data, and then post-prune the tree.
  - –Allow the tree to over fit the data, transform the tree to rules and then post-prune the rules.

Pessimistic pruning

Take the upper bound error at the node and sub-trees

$$e= [f+\frac{z^2}{2N}+z\sqrt{\frac{f}{N}} - \frac{f^2}{N}+\frac{z^2}{4N^2}]/[1+\frac{z^2}{N}]$$

Cost complexity pruning

J(Tree, S) = ErrorRate(Tree, S) + a |Tree|

Play with several values a starting from 0

Do a K-fold validation on all of them and find the best pruning α

TWO MOST POPULAR DECISION TREE ALGORITHMS

- **Cart**
  - –Binary split
  - –Gini index
  - –Cost complexity pruning
- **C5.0**
  - –Multi split
  - –Info gain
  - –pessimistic pruning

LIMITATIONS

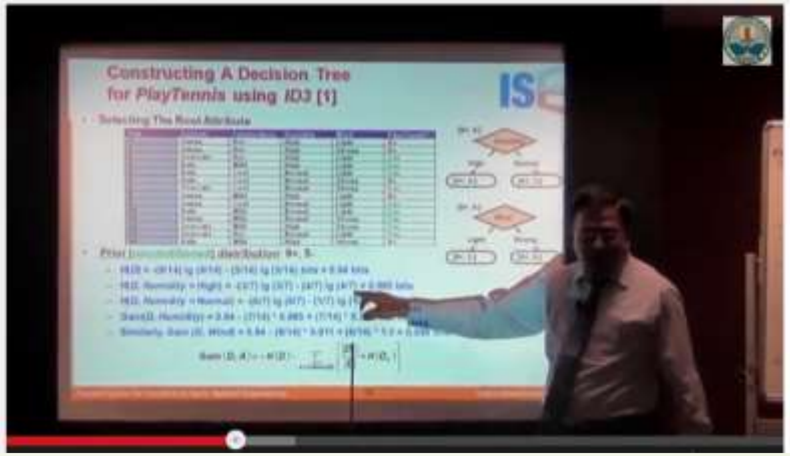
- Class imbalance
- When there are more records and very less number of attributes/features

ADVANTAGES

- They are fast
- Robust
- Requires very little experimentation
- You may also build some intuitions about your customer base. E.g. “Are customers with different family sizes truly different?

For Detailed Description on CONSTRUCTING A DECISION TREE with example

Check out our video



Constructing A Decision Tree for *PlayTennis* using *ID3* [1]

- Selecting The Root Attribute
- Prior (unconditional) distribution: 9+, 5-

International School of Engineering

INSOFE

Inspire...Educate...Transform.

Plot no 63/A, 1st Floor, Road No 13, Film Nagar, Jubilee
Hills, Hyderabad-500033

For Individuals (+91) 9502334561/62
For Corporates (+91) 9618 483 483

Facebook: www.facebook.com/insofe

Slide share: www.slideshare.net/INSOFE