Elective Course

Course Code: CS4103

Autumn 2025-26

Lecture #36

Artificial Intelligence for Data Science

Week-10:

MACHINE LEARNING (Part IV)

Decision Tree

Course Instructor:

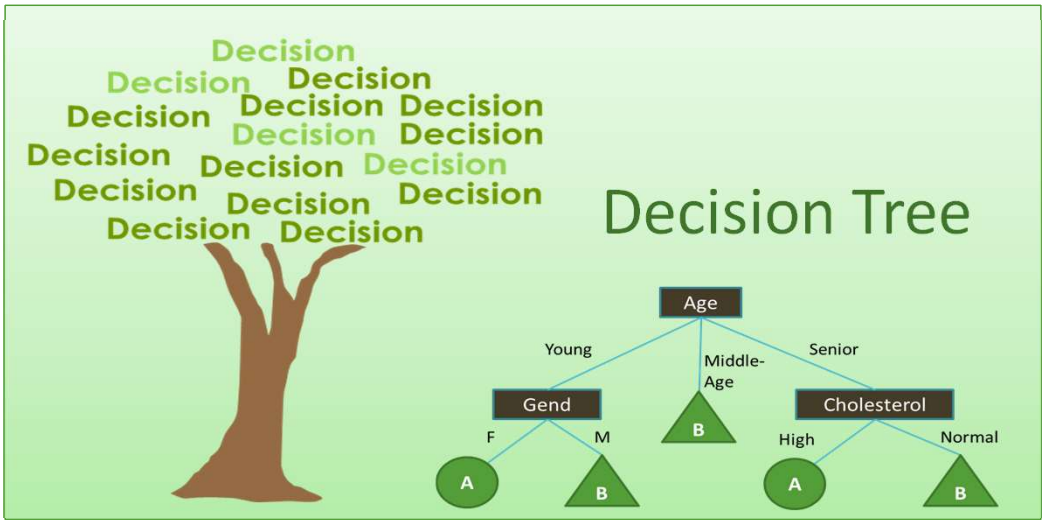
Dr. Monidipa Das

Assistant Professor

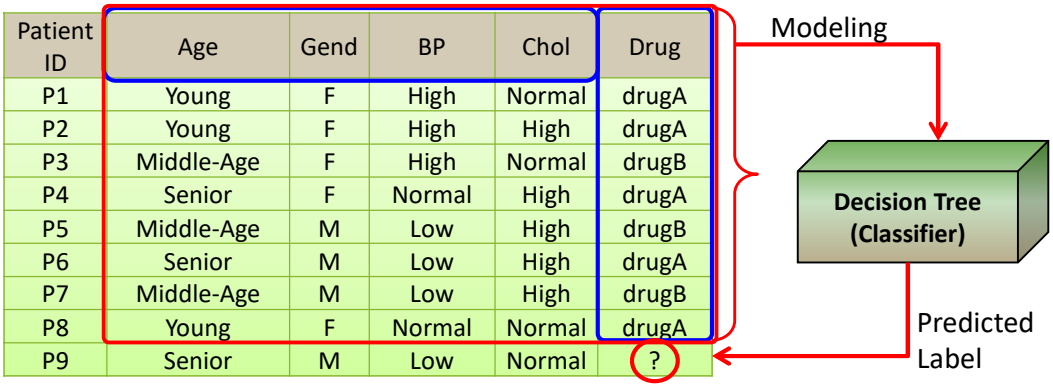
Department of Computational and Data Sciences

Indian Institute of Science Education and Research Kolkata, India 741246

2



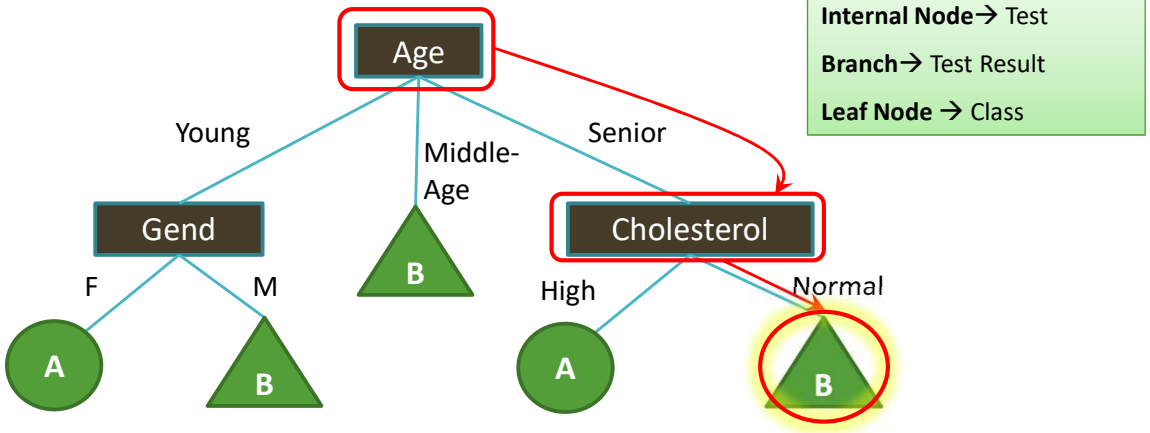
Decision Tree



Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA
P9	Senior	M	Low	Normal	?

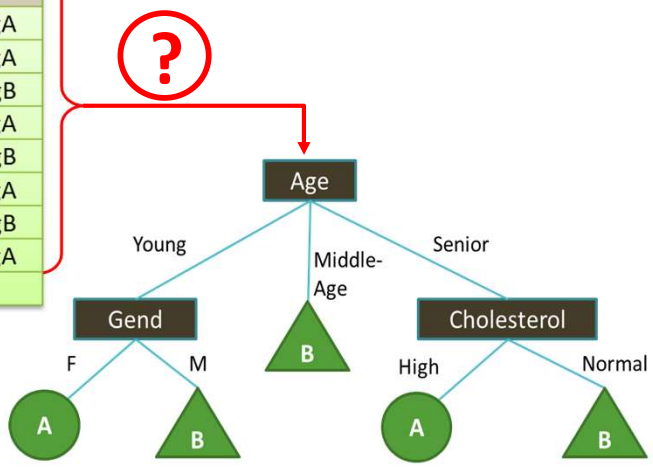
Decision Tree: Example

Patient ID	Age	Gend	BP	Chol	Drug
P9	Senior	M	Low	Normal	?



How to build a decision tree?

Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA
P9	Senior	M	Low	Normal	?

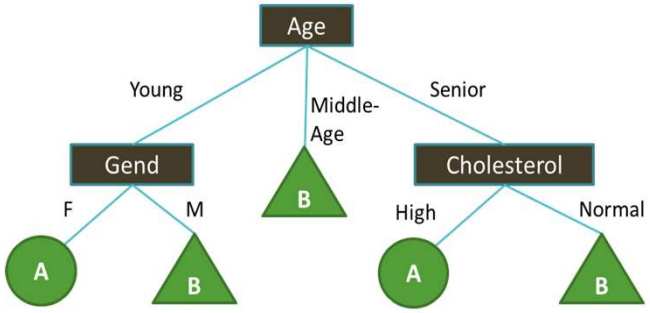


Decision Tree Learning Algorithm

1. Choose an attribute from the input feature set
2. Calculate the significance of the attribute in splitting the dataset
3. Split the dataset based on the value of the best attribute
4. Go to step 1

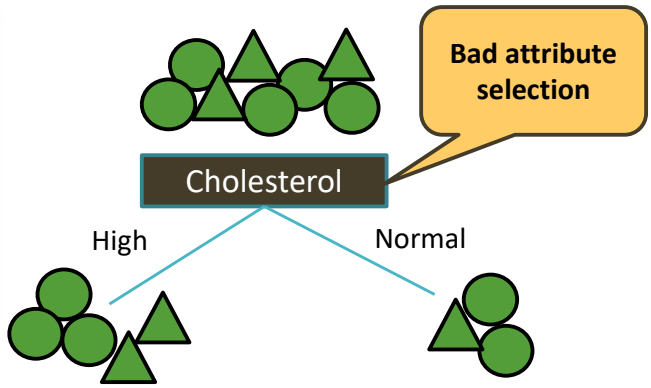
If training examples are perfectly classified along a branch, then STOP iterating there

Grow tree just deep enough for perfect classification if possible (or can approximate at chosen depth)



Finding the best attribute

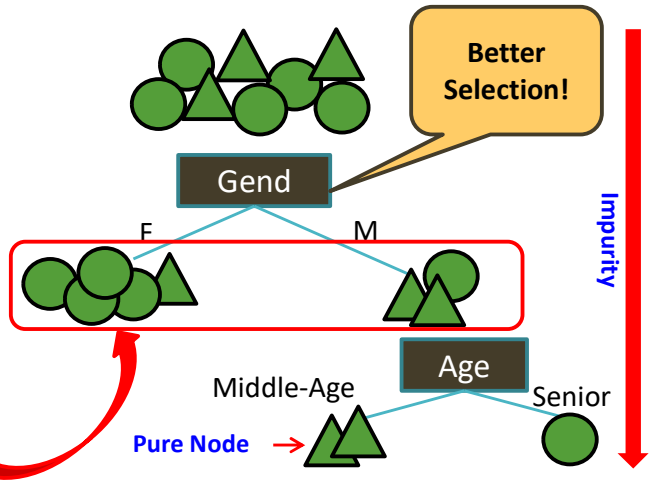
Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA



Drug-A
Drug-B

Finding the best attribute

Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA



Drug-A
Drug-B

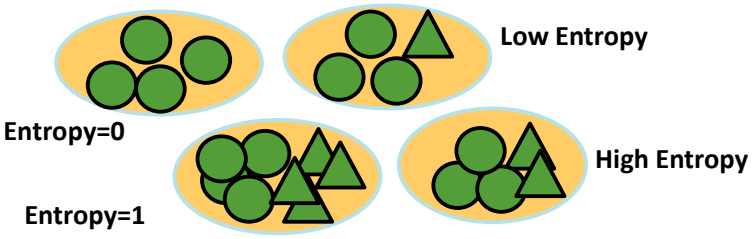
More **Predictiveness**
Lesser **Impurity**

Entropy

9

- Measure of uncertainty or randomness
- The lower the entropy the more homogeneous or purer the node

$$Entropy = -P(A)logP(A) - P(B)\log P(B)$$



$$Entropy = -\sum_{i=1}^{c} p_i \log_2 p_i$$

Decision Tree Building: Example

10

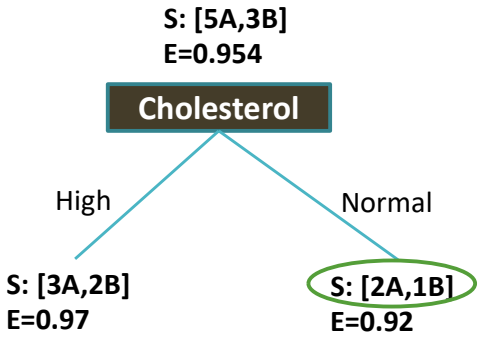
Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA

S: [5A,3B]

E=-P(A)log(P(A))-P(B)log(P(B))
E=-(5/8)log(5/8)-(3/8)log(3/8)=**0.954**

Decision Tree Building: Example

Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA

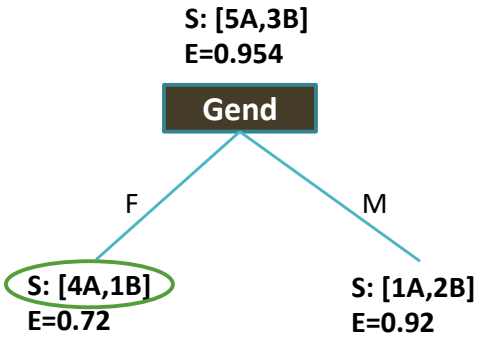


S: [5A,3B]

E=-P(A)log(P(A))-P(B)log(P(B))
E=-(5/8)log(5/8)-(3/8)log(3/8)=**0.954**

Decision Tree Building: Example

Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA



S: [5A,3B]

E=-P(A)log(P(A))-P(B)log(P(B))
E=-(5/8)log(5/8)-(3/8)log(3/8)=**0.954**

Decision Tree Building: Example

Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA

Select the tree with **higher information gain** after splitting

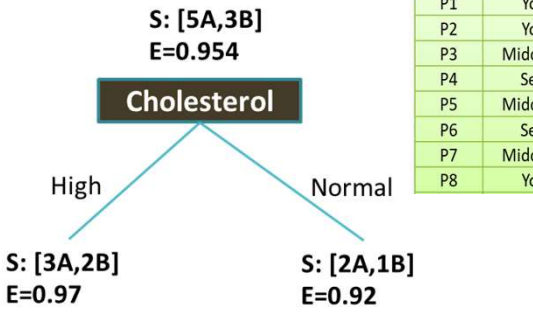
What is information gain?

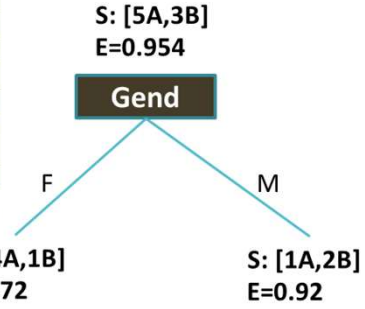
- Measure of increased level of certainty after splitting
- $Information\ Gain = (Entropy\ before\ split) - (WeightedEntropy\ after\ split)$

$WeightedEntropy\ after\ split$ ↓ $Information\ Gain$ ↑

Decision Tree Building: Example

Patient ID	Age	Gend	BP	Chol	Drug
P1	Young	F	High	Normal	drugA
P2	Young	F	High	High	drugA
P3	Middle-Age	F	High	Normal	drugB
P4	Senior	F	Normal	High	drugA
P5	Middle-Age	M	Low	High	drugB
P6	Senior	M	Low	High	drugA
P7	Middle-Age	M	Low	High	drugB
P8	Young	F	Normal	Normal	drugA



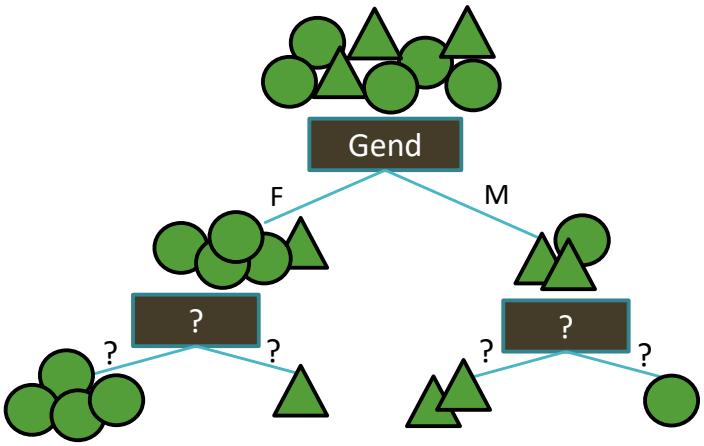


?

$$InfoGain(S, Cholesterol) = 0.954 - \left[\left(\frac{5}{8}\right) \times 0.97 + \left(\frac{3}{8}\right) \times 0.92\right] = 0.003$$

$$InfoGain(S, Gend) = 0.954 - \left[\left(\frac{5}{8}\right) \times 0.72 + \left(\frac{3}{8}\right) \times 0.92\right] = 0.159$$

Decision Tree Building: Example



Various Decision Tree Algorithms

	Splitting Criteria
ID3	Information Gain
CART	Gini impurity
C4.5	Gain Ratio

Decision Tree with Real-Valued Inputs

17

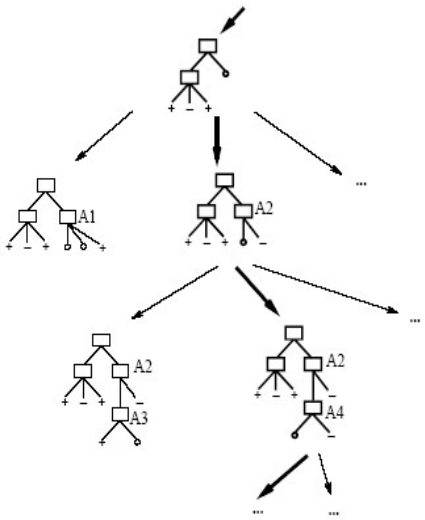
- Infinite number of possible split values!!!
- Discretized into ranges
- Or use Threshold splits
  - Binary tree: split on attribute X at value t
    - One branch: $X < t$
    - Other branch: $X \geq t$
- Need to find optimal split points within the range of values for that feature
  - How to find the split with the highest gain?
    - For each continuous feature A:
    - Sort examples according to the value of A
    - For each ordered pair (a1,a2) with different labels
    - Check the mid-point as a possible threshold, i.e.

Example:
- Length (L): 10 15 21 28 32 40 50
- labels: - + + - + + -
- Check thresholds: L < 12.5; L < 24.5; L < 45

Hypothesis Space Search in Decision Trees

18

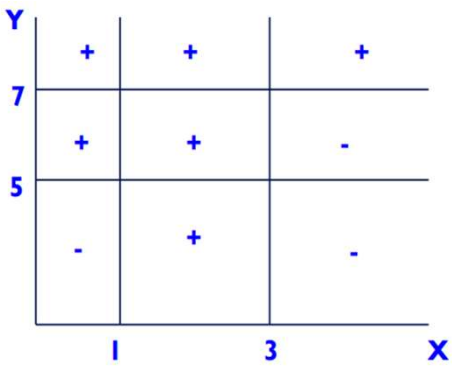
- Hypothesis space (all possible trees) is complete
  - Target function is included in there
- Conduct a search of the space of decision trees which can represent all possible discrete functions.
- Goal: to find the best decision tree

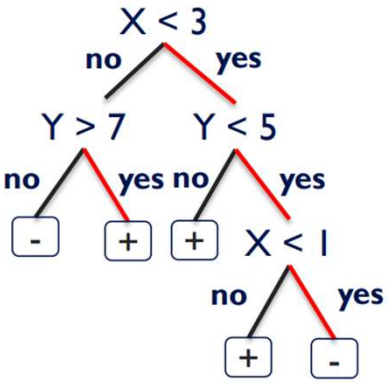


Decision Boundaries

19

- Divides the features space into axis-parallel rectangles, each labeled with one of the labels





Decision Tree: Advantages and Disadvantages

20

- Advantages
  - Can handle very large datasets with many attributes
  - Flexible: several attribute types, classification and regression tasks, missing values…
  - Interpretability: provide rules and attribute importance
- Disadvantages
  - High variance
  - Not always competitive with other algorithms in terms of accuracy

Decision Tree: Summary

- Decision trees are one of the most popular ML tools
  - Easy to understand, implement, and use
  - Computationally cheap
- Information gain, Gini impurity, etc. to select attributes
- Can be used for *regression* and *density estimation* as well
- Decision trees overfit
  - Must use tricks to find "simple trees"
  - Pruning
  - Use ensembles of different trees (random forests)

Questions?