

Decision Trees

Ninghao Liu

University of Georgia

February 6, 2024

Some contents adopted from “Data Mining”, Section 8.2, by Jiawei Han et al.

Decision Tree: Example

	age	income	student	credit_rating	com
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31...40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31...40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31...40	medium	no	excellent	yes
13	31...40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Figure 1: The input tabular dataset.

Decision Tree: Example

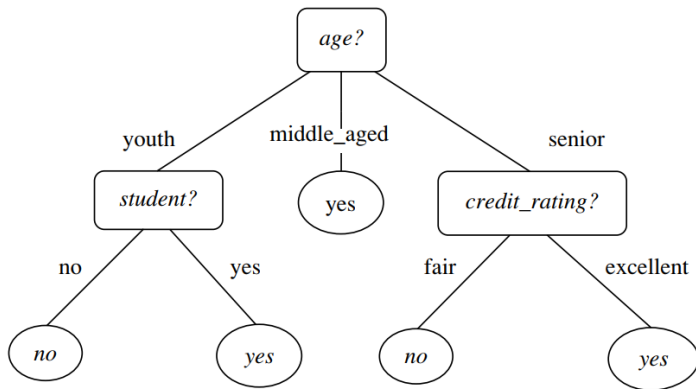
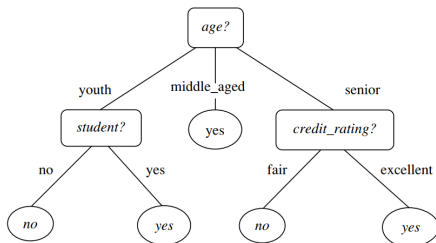


Figure 2: A decision tree for "buys_computer", indicating whether a customer is likely to purchase a computer.

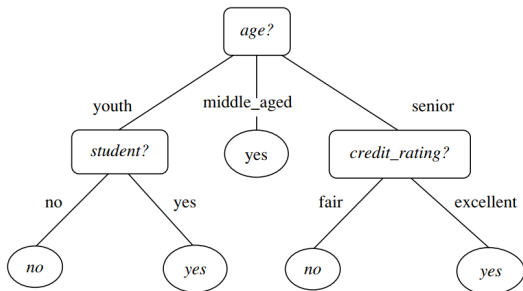
Decision Tree: Basic Concepts



A **decision tree** is a flowchart-like tree structure

- **Internal node** (nonleaf node): a test on an attribute (e.g., age).
- **Branch**: an outcome of the test.
- **Leaf node**: class label.

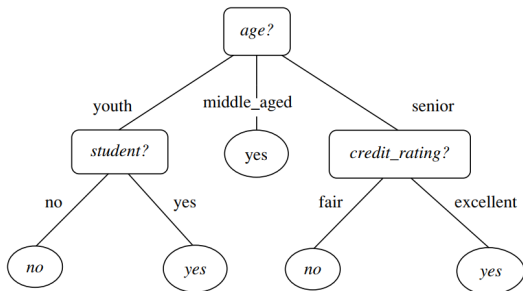
Decision Tree: Basic Concepts



Given a feature vector x whose label is unknown:

- The attribute values are tested against the decision tree.
- A **path** is traced from the root to a leaf node. The leaf node holds the classification label.
- Decision trees can easily be converted to classification **rules**.

Decision Tree: Basic Concepts



Advantages of decision trees:

- **Intuitive** and generally easy to assimilate by humans.
- **Fast** in learning and prediction.
- Have **superior accuracy** in **certain tasks**, even when compared with deep models.

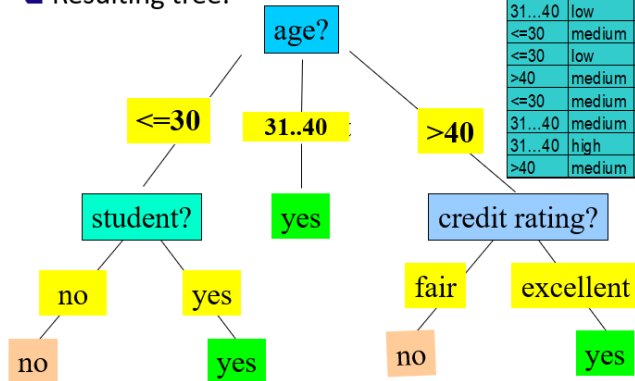
Decision Trees Induction

Brief history of decision tree learning:

- **ID3**: Iterative Dichotomiser 3.
- **C4.5**: C4.5 can be used for classification tasks and has been extended to handle regression in later versions (e.g., C5.0).
- **CART**: Classification and Regression Trees.

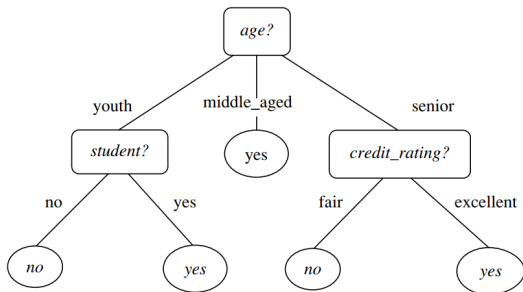
Decision Trees Induction

- ❑ Training data set: Buys_computer
- ❑ The data set follows an example of Quinlan's ID3 (Playing Tennis)
- ❑ Resulting tree:



age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

Decision Trees Induction



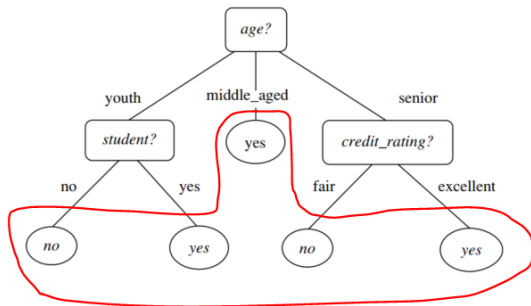
Given a training set $\mathcal{D} = \{(\mathbf{x}, y)\}$, the general idea of decision tree induction is:

- **Greedy:** Non-backtracking.
- **Top-down:** Starting from root to intermediate nodes to leafs.
- **Divide-and-Conquer:** The **training data** is recursively **partitioned** into smaller subsets as the tree is being built.

Data Partition?

age	income	student	credit rating	com
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

Data Partition?



- Training samples are partitioned after following paths to the leafs.
- The samples within each leaf node are expected to have the same label.

Decision Trees Induction

Basic algorithm:

- At start, all the training examples are at the root.
- Attributes are **categorical** (if continuous-valued, they are discretized in advance).
- Examples are **partitioned recursively** based on selected attributes.
- Test attributes are **selected** on the basis of a heuristic or statistical measure (e.g., information gain).

Conditions for stopping:

- All samples for a given node belong to the same class.
- There are no remaining attributes for further partitioning – majority voting is employed for classifying the leaf.
- There are no samples left.

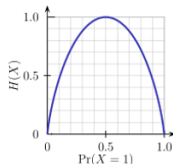
How to Select Attributes?

- Entropy (Information Theory)
 - A measure of uncertainty associated with a random variable
 - Calculation: For a discrete random variable Y taking m distinct values $\{y_1, \dots, y_m\}$,
 - $H(Y) = -\sum_{i=1}^m p_i \log(p_i)$, where $p_i = P(Y = y_i)$

How to Select Attributes?

- Entropy (Information Theory)

- A measure of uncertainty associated with a random variable
- Calculation: For a discrete random variable Y taking m distinct values $\{y_1, \dots, y_m\}$,
 - $H(Y) = -\sum_{i=1}^m p_i \log(p_i)$, where $p_i = P(Y = y_i)$
- Interpretation:
 - Higher entropy \Rightarrow higher uncertainty
 - Lower entropy \Rightarrow lower uncertainty



m = 2

How to Select Attributes?

To be continued ...