Introduction to Artificial Intelligence

Chapter 4: Learning
(1) Learning Decision Trees

Nguyễn Hải Minh, Ph.D
nhminh@fit.hcmus.edu.vn

Outline

- Form of Learning
- Learning from Decision Trees
- Summary

Learning Agents – Why learning?

1. Unknown environments
   - *i.e., a robot designed to navigate mazes must learn the layout of each new maze it encounters.*
2. Environment changes over time
   - *i.e., An agent designed to predict tomorrow's stock market prices must learn to adapt when conditions change from boom to bust.*
3. No idea how to program a solution
   - *i.e., the task to recognizing the faces of family members.*

Learning element

- Design of a learning element is affected by
  - Which *components* is to be improved
  - What *prior knowledge* the agent already has
  - What *representation* is used for the components
  - What ***feedback*** is available to learn these components

- Type of feedback:
  - Supervised learning: correct answers for each example
  - Unsupervised learning: correct answers not given
  - Reinforcement learning: occasional rewards

Supervised Learning

- Simplest form: learn a function from examples
- Problem: given a training set of N example input-output pairs

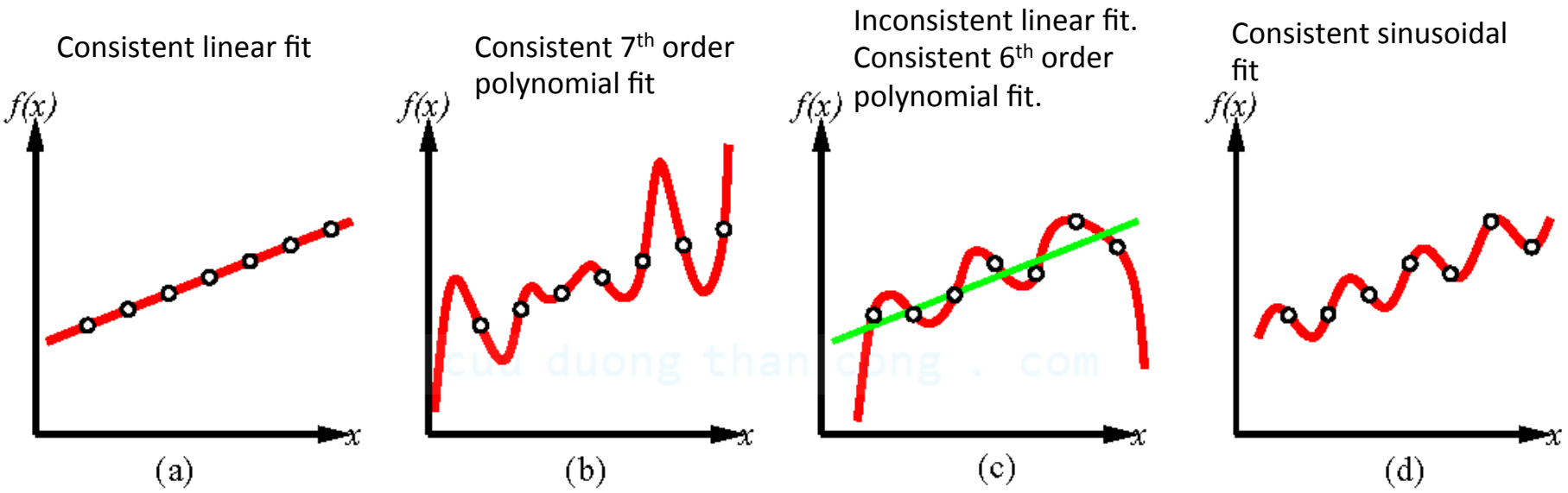
$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$$

Where each y_j was generated by an unknown function $y = f(x)$

→ Find a hypothesis h such that $h \approx f$

- To measure the accuracy of a hypothesis we give it a test set of examples that are different with the training set.

Supervised Learning

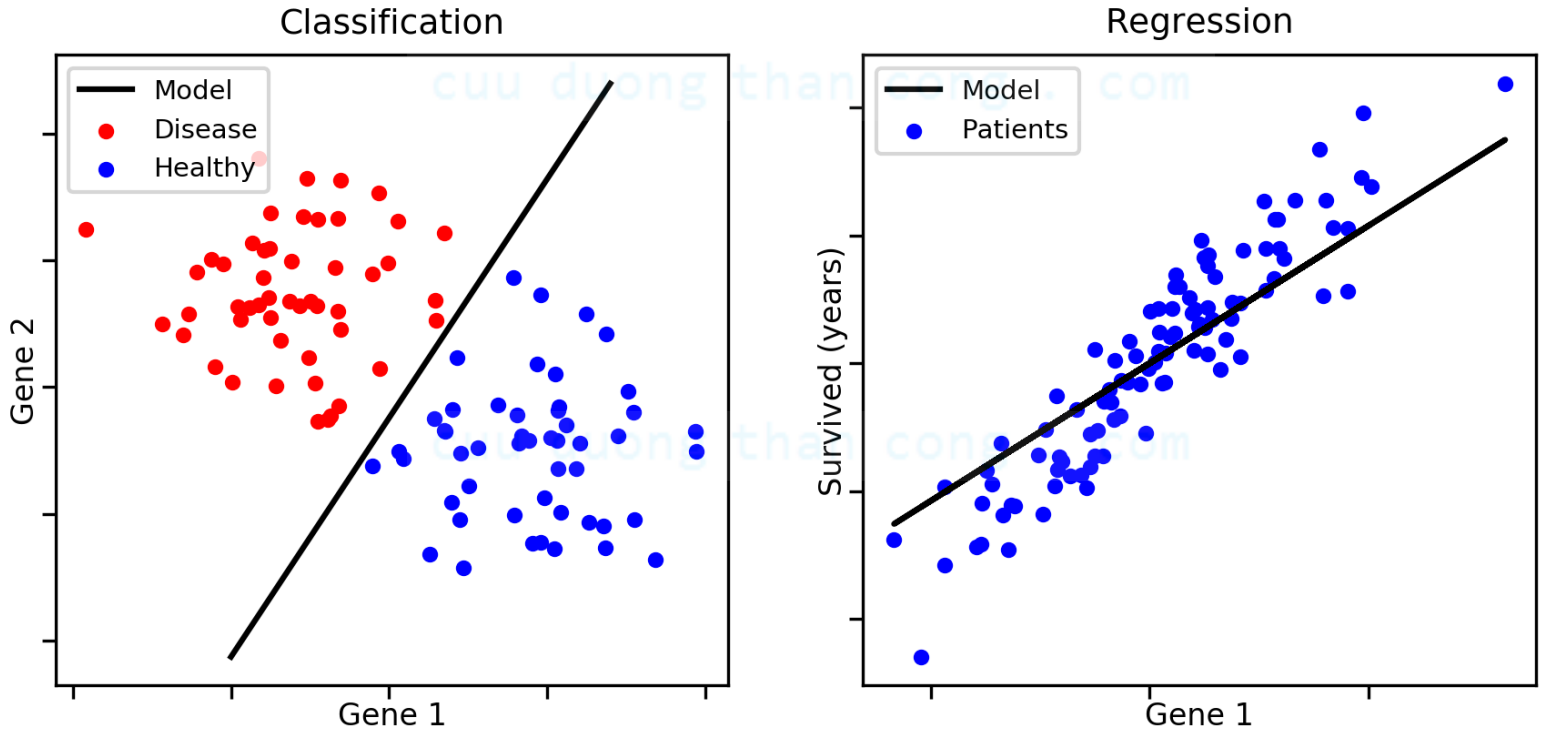


- Construct *h* so that it agrees with *f*.
- The hypothesis *h* is consistent if it agrees with *f* on all observations.
- **Ockham's razor:** Select the simplest consistent hypothesis.

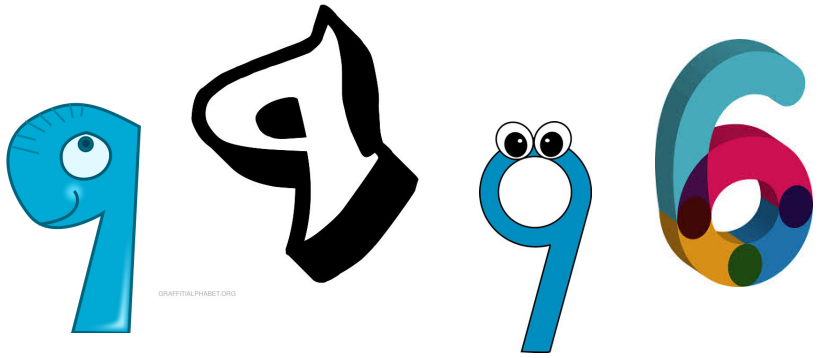
Learning problems

$h(\boldsymbol{x})$ = the predicted output value for the input $\boldsymbol{x}$.

- Discrete valued function ⇒ classification
- Continuous valued function ⇒ regression



Classification



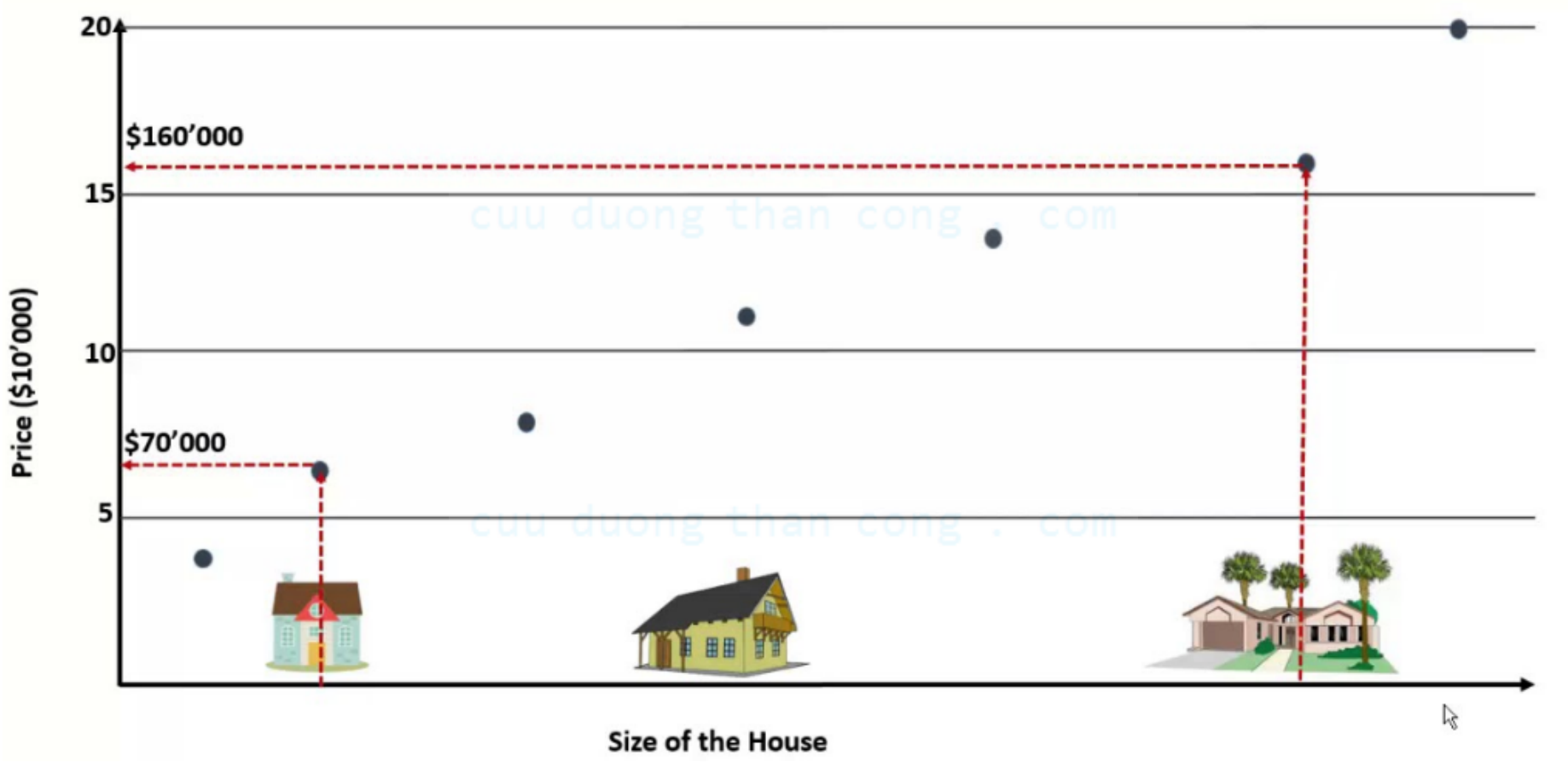
- Is this number 9?
  - 2 classes: Yes/No

- Will you pass or fail the exam?
  - 2 classes: Fail/Pass

- Is this an apple, an orange or a tomato?
  - 3 classes: Apple/Orange/ Tomato

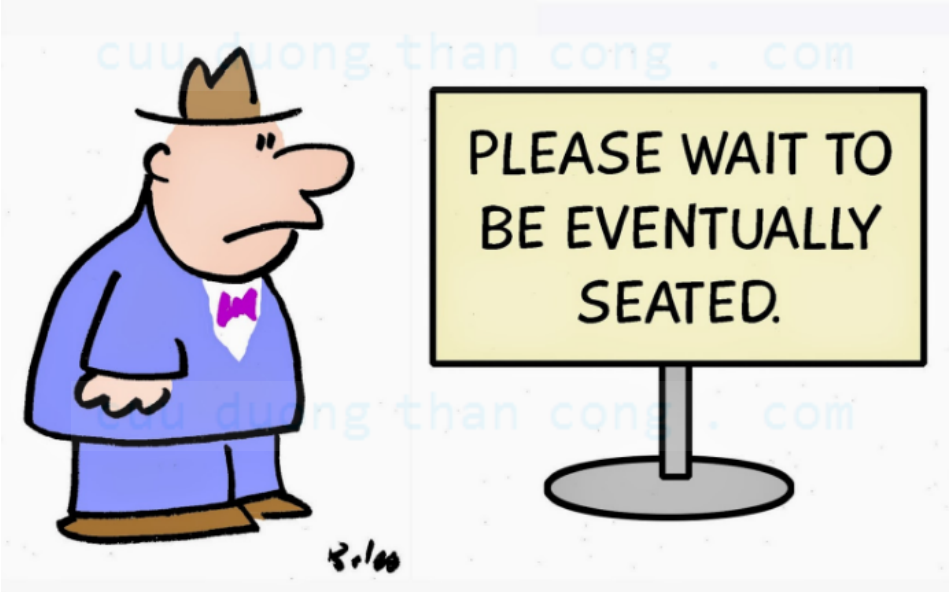
Regression

- Estimating the price of a house



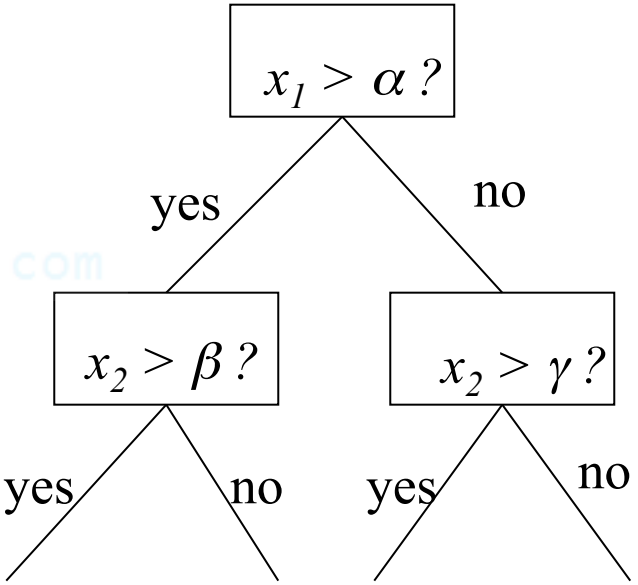
A classification problem example

Predicting whether a certain person will wait to have a seat in a restaurant.



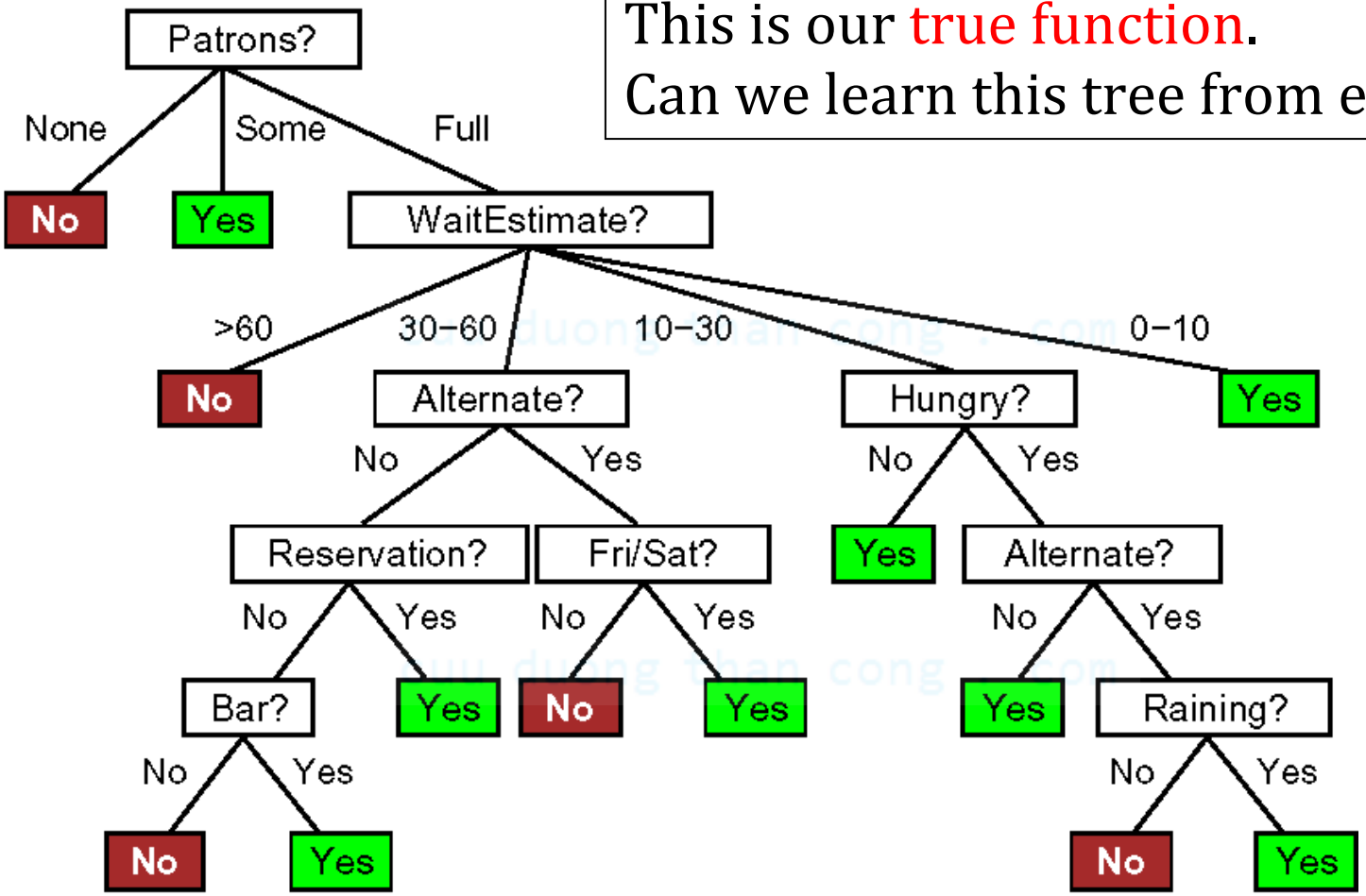
Learning Decision trees

- "Divide and conquer": Split data into smaller and smaller subsets
- Splits usually on a single variable



The wait@restaurant decision tree

This is our true function.
Can we learn this tree from examples?



Inductive learning of decision tree

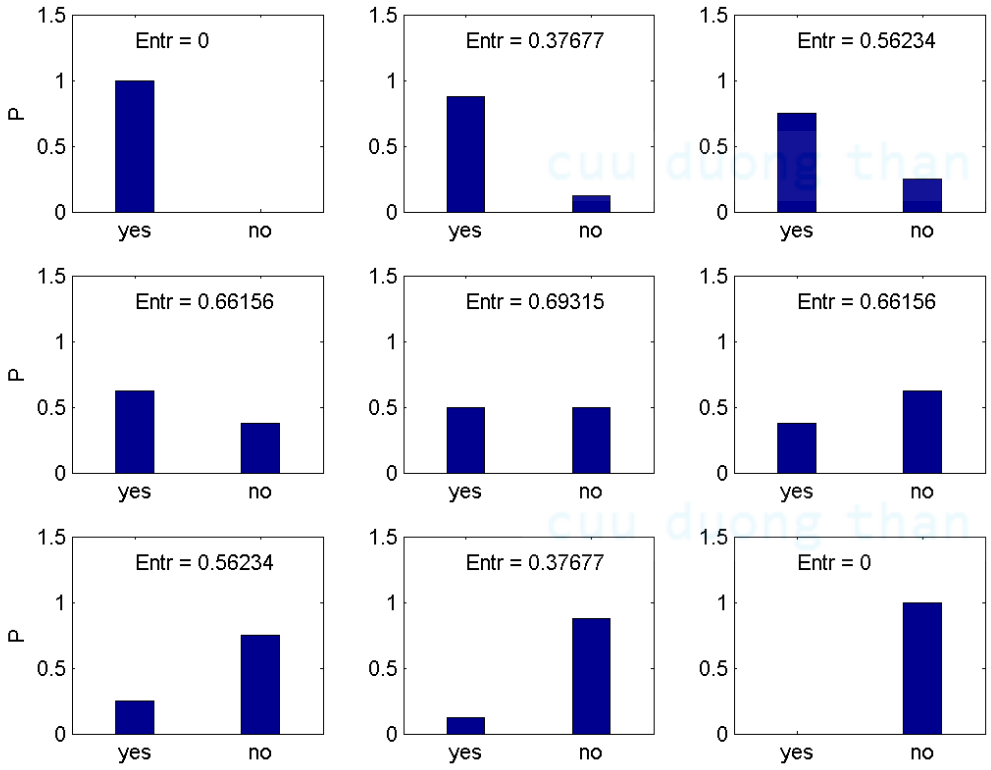
- **Simplest:** Construct a decision tree with one leaf for every example = memory based learning.
  Not very good generalization.
- **Advanced:** Split on each variable so that the purity of each split increases (i.e. either only yes or only no)
- Purity measured,e.g, with entropy
  - Entropy is a measure of the uncertainty of a random variable V with one value v_k

Entropy: $$H(V) = \sum_k P(v_k) \log_2 \frac{1}{P(v_k)} = -\sum_k P(v_k) \log_2 P(v_k)$$

  - v_k: 1 class in V (yes/no in binary classification)
  - $P(v_k)$: the proportion of the number of elements in class v_k to the number of elements in V

Entropy

- Entropy is a measure of the uncertainty of a random variable with only one value



The entropy is maximal when all possibilities are equally likely.

The goal of the decision tree is to decrease the entropy in each node.

Entropy is zero in a pure "yes" node (or pure "no" node).

Decision tree learning example

Problem: decide whether to wait for a table at a restaurant, based on the following attributes:

1. **Alternate:** is there an alternative restaurant nearby?
2. **Bar:** is there a comfortable bar area to wait in?
3. **Fri/Sat:** is today Friday or Saturday?
4. **Hungry:** are we hungry?
5. **Patrons:** number of people in the restaurant (None, Some, Full)
6. **Price:** price range ($, $$, $$$)
7. **Raining:** is it raining outside?
8. **Reservation:** have we made a reservation?
9. **Type:** kind of restaurant (French, Italian, Thai, Burger)
10. **WaitEstimate:** estimated waiting time (0-10, 10-30, 30-60, >60)

Decision tree learning example

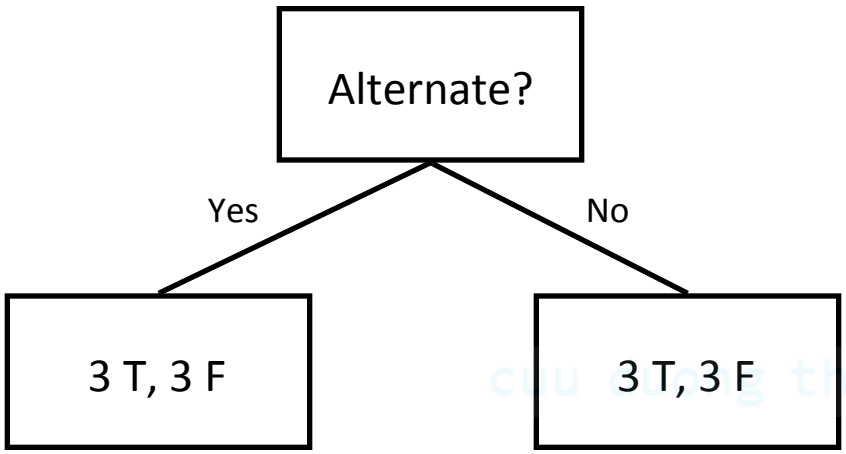
T = True, F = False

Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	T	F	F	T	Some	$$$	F	T	French	0–10	T
X_2	T	F	F	T	Full	$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	$$$	F	T	French	>60	F
X_6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	$	F	F	Burger	30–60	T

$$H(S) = -\left(6/12\right)\log_2\left(6/12\right) - \left(6/12\right)\log_2\left(6/12\right) = 1$$

6 True,
6 False

Decision tree learning example



Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	T	F	F	T	Some	$$$	F	T	French	0–10	T
X_2	T	F	F	T	Full	$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	$$$	F	T	French	>60	F
X_6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	$	F	F	Burger	30–60	T

- Calculate Average Entropy of attribute Alternate:

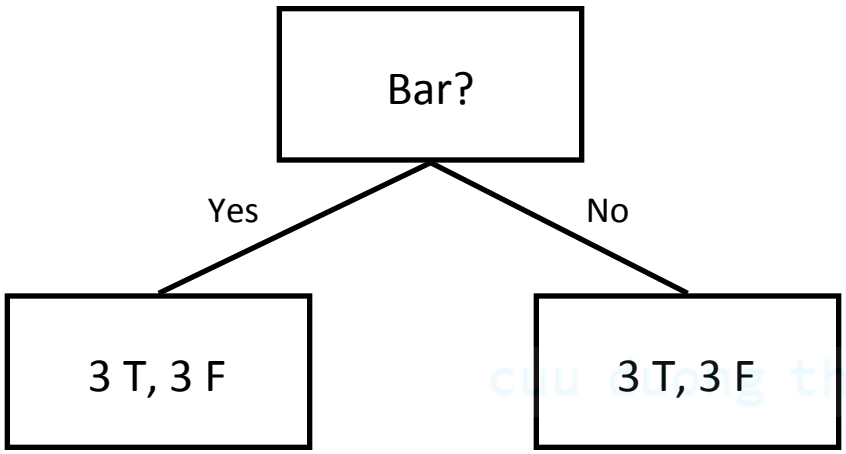
$$\mathrm{AE}_{\mathrm{Alternate}} = P(\mathrm{Alt}=T) \times \mathrm{H}(\mathrm{Alt}=T) + P(\mathrm{Alt}=F) \times H(\mathrm{Alt}=F)$$

$$\mathrm{AE}_{\mathrm{Alternate}} = \frac{6}{12}\left[-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)\right]+\frac{6}{12}\left[-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)\right]=1$$

- Information Gained (difference in entropy from before to after the set S is split on attribute Alternate)

$$IG(\mathrm{Alternate}, S) = H(S) - \mathrm{AE}_{\mathrm{Alternate}} = 1 - 1 = 0$$

Decision tree learning example

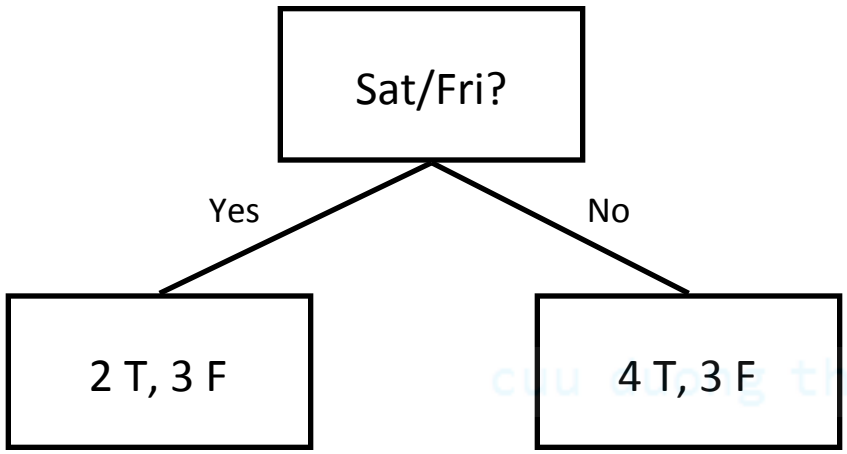


Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	*T*	*F*	*F*	*T*	*Some*	*$$$*	*F*	*T*	*French*	*0–10*	*T*
X_2	*T*	*F*	*F*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*30–60*	*F*
X_3	*F*	*T*	*F*	*F*	*Some*	*$*	*F*	*F*	*Burger*	*0–10*	*T*
X_4	*T*	*F*	*T*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*10–30*	*T*
X_5	*T*	*F*	*T*	*F*	*Full*	*$$$*	*F*	*T*	*French*	*>60*	*F*
X_6	*F*	*T*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Italian*	*0–10*	*T*
X_7	*F*	*T*	*F*	*F*	*None*	*$*	*T*	*F*	*Burger*	*0–10*	*F*
X_8	*F*	*F*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Thai*	*0–10*	*T*
X_9	*F*	*T*	*T*	*F*	*Full*	*$*	*T*	*F*	*Burger*	*>60*	*F*
X_{10}	*T*	*T*	*T*	*T*	*Full*	*$$$*	*F*	*T*	*Italian*	*10–30*	*F*
X_{11}	*F*	*F*	*F*	*F*	*None*	*$*	*F*	*F*	*Thai*	*0–10*	*F*
X_{12}	*T*	*T*	*T*	*T*	*Full*	*$*	*F*	*F*	*Burger*	*30–60*	*T*

$$\mathrm{AE}_{\mathrm{Bar}} = \frac{6}{12}\left[-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)\right]+\frac{6}{12}\left[-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)\right]=1$$

$$IG(\mathrm{Bar}, S) = H(S) - \mathrm{AE}_{\mathrm{Bar}} = 1 - 1 = 0$$

Decision tree learning example



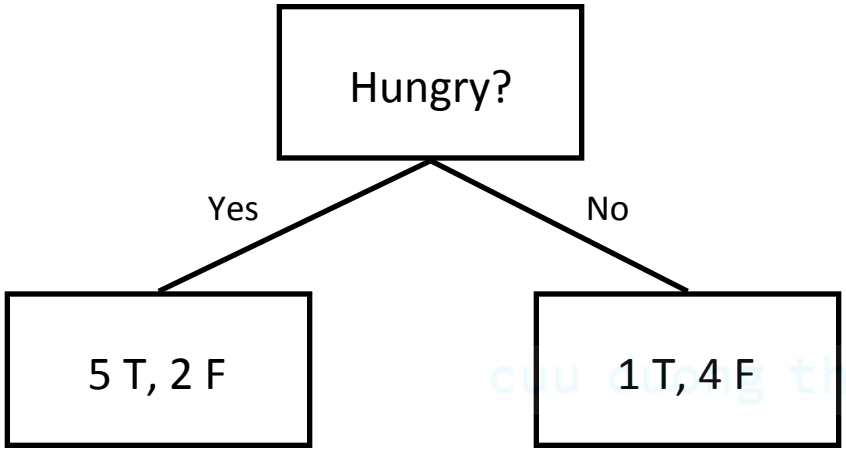
Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	*T*	*F*	*F*	*T*	*Some*	*$$$*	*F*	*T*	*French*	*0–10*	*T*
X_2	*T*	*F*	*F*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*30–60*	*F*
X_3	*F*	*T*	*F*	*F*	*Some*	*$*	*F*	*F*	*Burger*	*0–10*	*T*
X_4	*T*	*F*	*T*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*10–30*	*T*
X_5	*T*	*F*	*T*	*F*	*Full*	*$$$*	*F*	*T*	*French*	*>60*	*F*
X_6	*F*	*T*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Italian*	*0–10*	*T*
X_7	*F*	*T*	*F*	*F*	*None*	*$*	*T*	*F*	*Burger*	*0–10*	*F*
X_8	*F*	*F*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Thai*	*0–10*	*T*
X_9	*F*	*T*	*T*	*F*	*Full*	*$*	*T*	*F*	*Burger*	*>60*	*F*
X_{10}	*T*	*T*	*T*	*T*	*Full*	*$$$*	*F*	*T*	*Italian*	*10–30*	*F*
X_{11}	*F*	*F*	*F*	*F*	*None*	*$*	*F*	*F*	*Thai*	*0–10*	*F*
X_{12}	*T*	*T*	*T*	*T*	*Full*	*$*	*F*	*F*	*Burger*	*30–60*	*T*

$$\mathrm{AE}_{\mathrm{Sat/Fri?}} = \frac{5}{12}\left[-\left(\tfrac{2}{5}\right)\log_2\left(\tfrac{2}{5}\right)-\left(\tfrac{3}{5}\right)\log_2\left(\tfrac{3}{5}\right)\right]+$$

$$\frac{7}{12}\left[-\left(\tfrac{4}{7}\right)\log_2\left(\tfrac{4}{7}\right)-\left(\tfrac{3}{7}\right)\log_2\left(\tfrac{3}{7}\right)\right]=0.979$$

$$IG(\mathrm{Sat/Fri}, S) = H(S) - \mathrm{AE}_{\mathrm{Sat/Fri}} = 1 - 0.979 = 0.021$$

Decision tree learning example

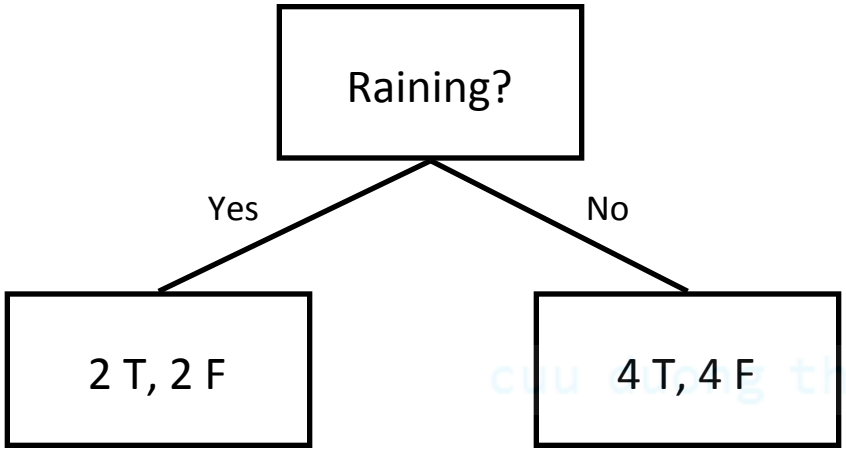


Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	*T*	*F*	*F*	*T*	*Some*	*$$$*	*F*	*T*	*French*	*0–10*	*T*
X_2	*T*	*F*	*F*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*30–60*	*F*
X_3	*F*	*T*	*F*	*F*	*Some*	*$*	*F*	*F*	*Burger*	*0–10*	*T*
X_4	*T*	*F*	*T*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*10–30*	*T*
X_5	*T*	*F*	*T*	*F*	*Full*	*$$$*	*F*	*T*	*French*	*>60*	*F*
X_6	*F*	*T*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Italian*	*0–10*	*T*
X_7	*F*	*T*	*F*	*F*	*None*	*$*	*T*	*F*	*Burger*	*0–10*	*F*
X_8	*F*	*F*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Thai*	*0–10*	*T*
X_9	*F*	*T*	*T*	*F*	*Full*	*$*	*T*	*F*	*Burger*	*>60*	*F*
X_{10}	*T*	*T*	*T*	*T*	*Full*	*$$$*	*F*	*T*	*Italian*	*10–30*	*F*
X_{11}	*F*	*F*	*F*	*F*	*None*	*$*	*F*	*F*	*Thai*	*0–10*	*F*
X_{12}	*T*	*T*	*T*	*T*	*Full*	*$*	*F*	*F*	*Burger*	*30–60*	*T*

$$\mathrm{AE}_{\mathrm{Hungry}} = \frac{7}{12}\left[-\left(\tfrac{5}{7}\right)\log_2\left(\tfrac{5}{7}\right)-\left(\tfrac{2}{7}\right)\log_2\left(\tfrac{2}{7}\right)\right]+\frac{5}{12}\left[-\left(\tfrac{1}{5}\right)\log_2\left(\tfrac{1}{5}\right)-\left(\tfrac{4}{5}\right)\log_2\left(\tfrac{4}{5}\right)\right]=0.804$$

$$IG(\mathrm{Hungry}, S) = H(S) - \mathrm{AE}_{\mathrm{Hungry}} = 1 - 0.804 = 0.196$$

Decision tree learning example

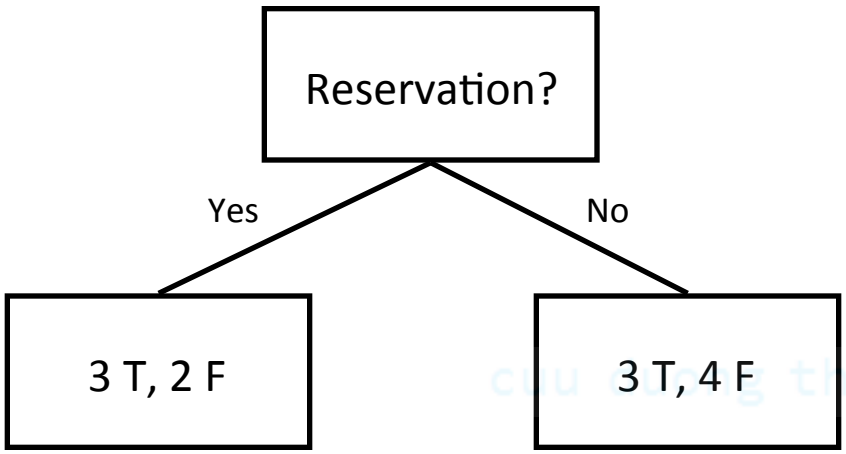


Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	*T*	*F*	*F*	*T*	*Some*	*$$$*	*F*	*T*	*French*	*0–10*	*T*
X_2	*T*	*F*	*F*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*30–60*	*F*
X_3	*F*	*T*	*F*	*F*	*Some*	*$*	*F*	*F*	*Burger*	*0–10*	*T*
X_4	*T*	*F*	*T*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*10–30*	*T*
X_5	*T*	*F*	*T*	*F*	*Full*	*$$$*	*F*	*T*	*French*	*>60*	*F*
X_6	*F*	*T*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Italian*	*0–10*	*T*
X_7	*F*	*T*	*F*	*F*	*None*	*$*	*T*	*F*	*Burger*	*0–10*	*F*
X_8	*F*	*F*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Thai*	*0–10*	*T*
X_9	*F*	*T*	*T*	*F*	*Full*	*$*	*T*	*F*	*Burger*	*>60*	*F*
X_{10}	*T*	*T*	*T*	*T*	*Full*	*$$$*	*F*	*T*	*Italian*	*10–30*	*F*
X_{11}	*F*	*F*	*F*	*F*	*None*	*$*	*F*	*F*	*Thai*	*0–10*	*F*
X_{12}	*T*	*T*	*T*	*T*	*Full*	*$*	*F*	*F*	*Burger*	*30–60*	*T*

$$\mathrm{AE}_{\mathrm{Raining}} = \frac{4}{12}\left[-\left(\tfrac{2}{4}\right)\log_2\left(\tfrac{2}{4}\right)-\left(\tfrac{2}{4}\right)\log_2\left(\tfrac{2}{4}\right)\right]+\frac{8}{12}\left[-\left(\tfrac{4}{8}\right)\log_2\left(\tfrac{4}{8}\right)-\left(\tfrac{4}{8}\right)\log_2\left(\tfrac{4}{8}\right)\right]= 1$$

$$IG(\mathrm{Raining}, S) = H(S) - \mathrm{AE}_{\mathrm{Raining}} = 1 - 1 = 0$$

Decision tree learning example

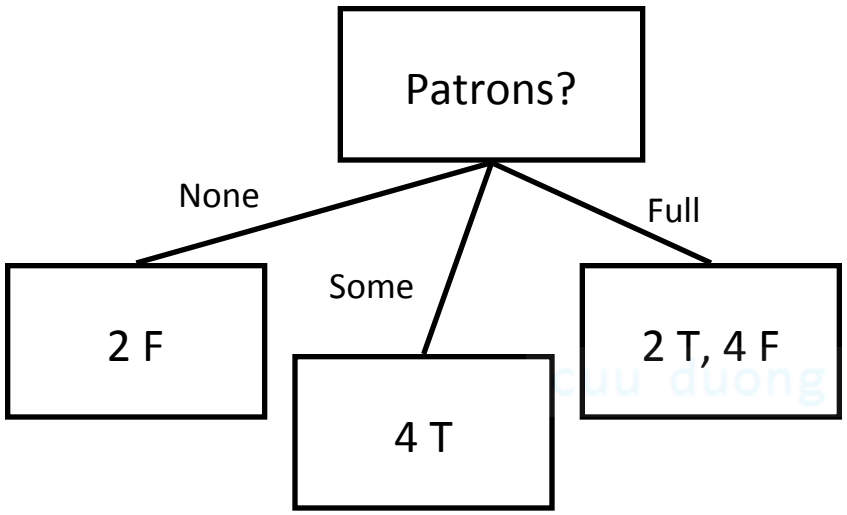


Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	*T*	*F*	*F*	*T*	*Some*	*$$$*	*F*	*T*	*French*	*0–10*	*T*
X_2	*T*	*F*	*F*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*30–60*	*F*
X_3	*F*	*T*	*F*	*F*	*Some*	*$*	*F*	*F*	*Burger*	*0–10*	*T*
X_4	*T*	*F*	*T*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*10–30*	*T*
X_5	*T*	*F*	*T*	*F*	*Full*	*$$$*	*F*	*T*	*French*	*>60*	*F*
X_6	*F*	*T*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Italian*	*0–10*	*T*
X_7	*F*	*T*	*F*	*F*	*None*	*$*	*T*	*F*	*Burger*	*0–10*	*F*
X_8	*F*	*F*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Thai*	*0–10*	*T*
X_9	*F*	*T*	*T*	*F*	*Full*	*$*	*T*	*F*	*Burger*	*>60*	*F*
X_{10}	*T*	*T*	*T*	*T*	*Full*	*$$$*	*F*	*T*	*Italian*	*10–30*	*F*
X_{11}	*F*	*F*	*F*	*F*	*None*	*$*	*F*	*F*	*Thai*	*0–10*	*F*
X_{12}	*T*	*T*	*T*	*T*	*Full*	*$*	*F*	*F*	*Burger*	*30–60*	*T*

$$\mathrm{AE}_{\mathrm{Reservation}} = \frac{5}{12}\left[-\left(\tfrac{3}{5}\right)\log_2\left(\tfrac{3}{5}\right)-\left(\tfrac{2}{5}\right)\log_2\left(\tfrac{2}{5}\right)\right]+\frac{7}{12}\left[-\left(\tfrac{3}{7}\right)\log_2\left(\tfrac{3}{7}\right)-\left(\tfrac{4}{7}\right)\log_2\left(\tfrac{4}{7}\right)\right]=0.979$$

$$IG(\text{Reservation}, S) = H(S) - \mathrm{AE}_{\mathrm{Reservation}} = 1 - 0.979 = 0.021$$

Decision tree learning example



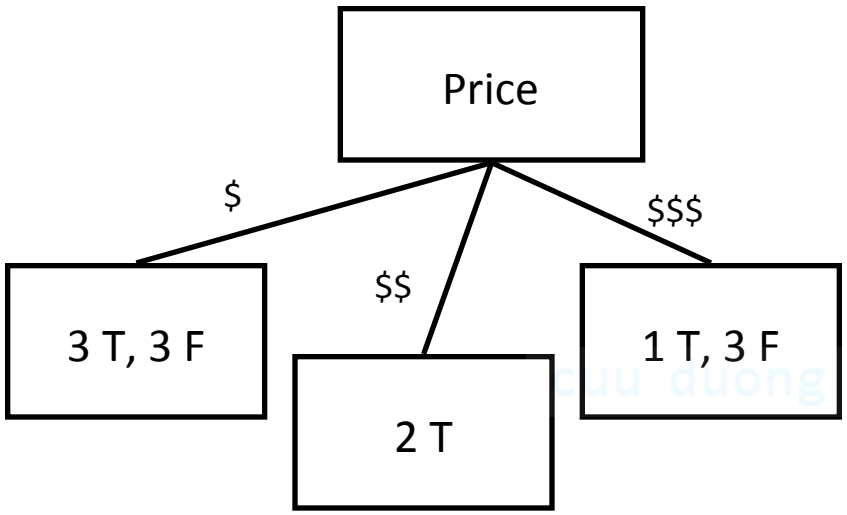
Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	*T*	*F*	*F*	*T*	*Some*	*$$$*	*F*	*T*	*French*	*0–10*	*T*
X_2	*T*	*F*	*F*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*30–60*	*F*
X_3	*F*	*T*	*F*	*F*	*Some*	*$*	*F*	*F*	*Burger*	*0–10*	*T*
X_4	*T*	*F*	*T*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*10–30*	*T*
X_5	*T*	*F*	*T*	*F*	*Full*	*$$$*	*F*	*T*	*French*	*>60*	*F*
X_6	*F*	*T*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Italian*	*0–10*	*T*
X_7	*F*	*T*	*F*	*F*	*None*	*$*	*T*	*F*	*Burger*	*0–10*	*F*
X_8	*F*	*F*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Thai*	*0–10*	*T*
X_9	*F*	*T*	*T*	*F*	*Full*	*$*	*T*	*F*	*Burger*	*>60*	*F*
X_{10}	*T*	*T*	*T*	*T*	*Full*	*$$$*	*F*	*T*	*Italian*	*10–30*	*F*
X_{11}	*F*	*F*	*F*	*F*	*None*	*$*	*F*	*F*	*Thai*	*0–10*	*F*
X_{12}	*T*	*T*	*T*	*T*	*Full*	*$*	*F*	*F*	*Burger*	*30–60*	*T*

$$\mathrm{AE}_{\mathrm{Patrons}} = \frac{2}{12}\left[-\left(\tfrac{0}{2}\right)\log_2\left(\tfrac{0}{2}\right)-\left(\tfrac{2}{2}\right)\log_2\left(\tfrac{2}{2}\right)\right]+\frac{4}{12}\left[-\left(\tfrac{4}{4}\right)\log_2\left(\tfrac{4}{4}\right)-\left(\tfrac{0}{4}\right)\log_2\left(\tfrac{0}{4}\right)\right]$$

$$+\frac{6}{12}\left[-\left(\tfrac{2}{6}\right)\log_2\left(\tfrac{2}{6}\right)-\left(\tfrac{4}{6}\right)\log_2\left(\tfrac{4}{6}\right)\right]=0.541$$

$$IG(\mathrm{Patrons}, S) = H(S) - \mathrm{AE}_{\mathrm{Patrons}} = 1 - 0.541 = 0.459$$

Decision tree learning example



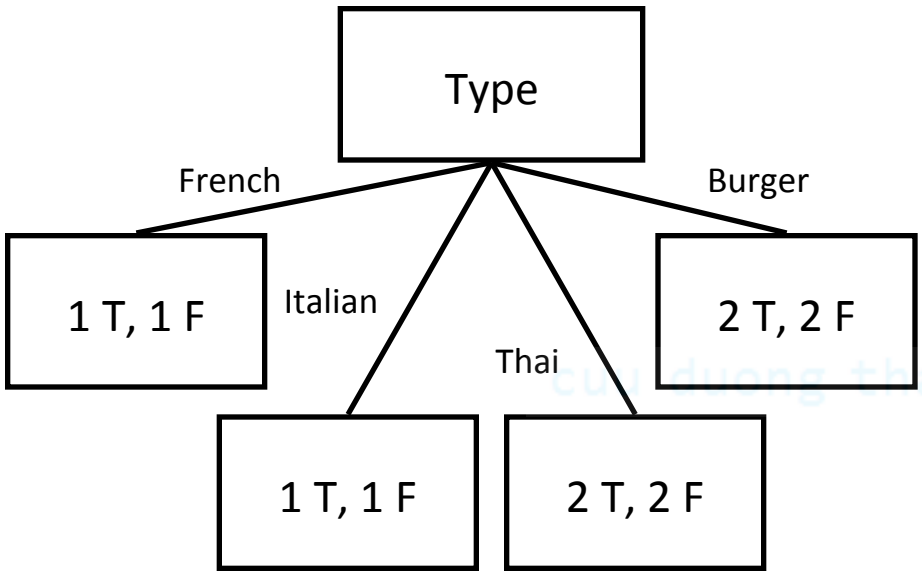
Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	*T*	*F*	*F*	*T*	*Some*	*$$$*	*F*	*T*	*French*	*0–10*	*T*
X_2	*T*	*F*	*F*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*30–60*	*F*
X_3	*F*	*T*	*F*	*F*	*Some*	*$*	*F*	*F*	*Burger*	*0–10*	*T*
X_4	*T*	*F*	*T*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*10–30*	*T*
X_5	*T*	*F*	*T*	*F*	*Full*	*$$$*	*F*	*T*	*French*	*>60*	*F*
X_6	*F*	*T*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Italian*	*0–10*	*T*
X_7	*F*	*T*	*F*	*F*	*None*	*$*	*T*	*F*	*Burger*	*0–10*	*F*
X_8	*F*	*F*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Thai*	*0–10*	*T*
X_9	*F*	*T*	*T*	*F*	*Full*	*$*	*T*	*F*	*Burger*	*>60*	*F*
X_{10}	*T*	*T*	*T*	*T*	*Full*	*$$$*	*F*	*T*	*Italian*	*10–30*	*F*
X_{11}	*F*	*F*	*F*	*F*	*None*	*$*	*F*	*F*	*Thai*	*0–10*	*F*
X_{12}	*T*	*T*	*T*	*T*	*Full*	*$*	*F*	*F*	*Burger*	*30–60*	*T*

$$\text{AE}_{\text{Price}} = \frac{6}{12}\left[-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)-\left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)\right]+\frac{2}{12}\left[-\left(\tfrac{2}{2}\right)\log_2\left(\tfrac{2}{2}\right)-\left(\tfrac{0}{2}\right)\log_2\left(\tfrac{0}{2}\right)\right]$$

$$+\frac{4}{12}\left[-\left(\tfrac{1}{4}\right)\log_2\left(\tfrac{1}{4}\right)-\left(\tfrac{3}{4}\right)\log_2\left(\tfrac{3}{4}\right)\right]=0.770$$

$$IG(\text{Price}, S) = H(S) - \text{AE}_{\text{Price}} = 1 - 0.770 = 0.23$$

Decision tree learning example



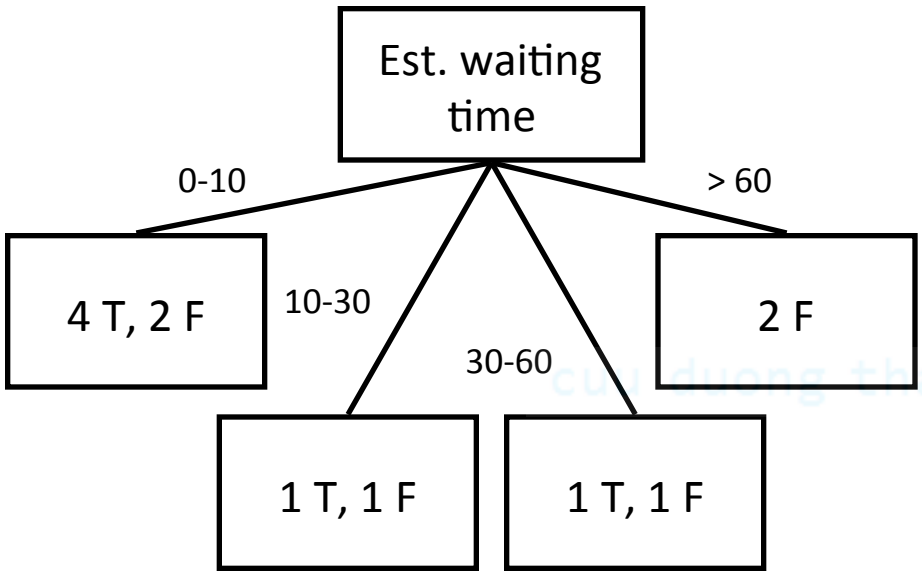
Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	T	F	F	T	Some	$$$	F	T	French	0–10	T
X_2	T	F	F	T	Full	$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	$$$	F	T	French	>60	F
X_6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	$	F	F	Burger	30–60	T

$$\text{AE}_{\text{Type}} = \frac{2}{12}\left[-\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right)-\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right)\right]+\frac{2}{12}\left[-\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right)-\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right)\right]$$

$$+\frac{4}{12}\left[-\left(\tfrac{2}{4}\right)\log_2\left(\tfrac{2}{4}\right)-\left(\tfrac{2}{4}\right)\log_2\left(\tfrac{2}{4}\right)\right]+\frac{4}{12}\left[-\left(\tfrac{2}{4}\right)\log_2\left(\tfrac{2}{4}\right)-\left(\tfrac{2}{4}\right)\log_2\left(\tfrac{2}{4}\right)\right]=1$$

$$IG(\text{Type}, S) = H(S) - \text{AE}_{\text{Alternate}} = 1 - 1 = 0$$

Decision tree learning example



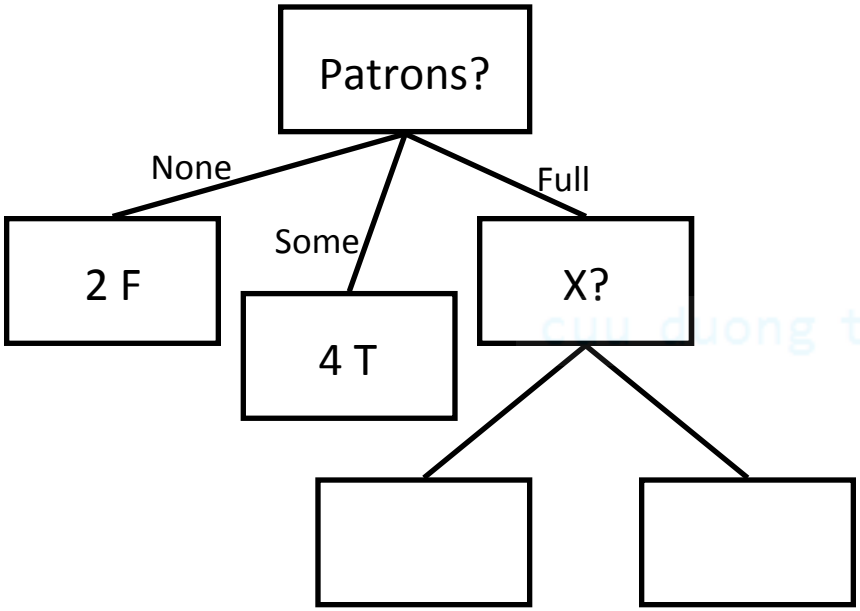
Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	*T*	*F*	*F*	*T*	*Some*	*$$$*	*F*	*T*	*French*	*0–10*	*T*
X_2	*T*	*F*	*F*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*30–60*	*F*
X_3	*F*	*T*	*F*	*F*	*Some*	*$*	*F*	*F*	*Burger*	*0–10*	*T*
X_4	*T*	*F*	*T*	*T*	*Full*	*$*	*F*	*F*	*Thai*	*10–30*	*T*
X_5	*T*	*F*	*T*	*F*	*Full*	*$$$*	*F*	*T*	*French*	*>60*	*F*
X_6	*F*	*T*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Italian*	*0–10*	*T*
X_7	*F*	*T*	*F*	*F*	*None*	*$*	*T*	*F*	*Burger*	*0–10*	*F*
X_8	*F*	*F*	*F*	*T*	*Some*	*$$*	*T*	*T*	*Thai*	*0–10*	*T*
X_9	*F*	*T*	*T*	*F*	*Full*	*$*	*T*	*F*	*Burger*	*>60*	*F*
X_{10}	*T*	*T*	*T*	*T*	*Full*	*$$$*	*F*	*T*	*Italian*	*10–30*	*F*
X_{11}	*F*	*F*	*F*	*F*	*None*	*$*	*F*	*F*	*Thai*	*0–10*	*F*
X_{12}	*T*	*T*	*T*	*T*	*Full*	*$*	*F*	*F*	*Burger*	*30–60*	*T*

$$\text{AE}_{\text{Est.waiting time}} = \frac{6}{12}\left[-\left(\tfrac{4}{6}\right)\log_2\left(\tfrac{4}{6}\right)-\left(\tfrac{2}{6}\right)\log_2\left(\tfrac{2}{6}\right)\right]+\frac{2}{12}\left[-\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right)-\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right)\right]$$

$$+\frac{2}{12}\left[-\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right)-\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right)\right]+\frac{2}{12}\left[-\left(\tfrac{0}{2}\right)\log_2\left(\tfrac{0}{2}\right)-\left(\tfrac{2}{2}\right)\log_2\left(\tfrac{2}{2}\right)\right]=0.792$$

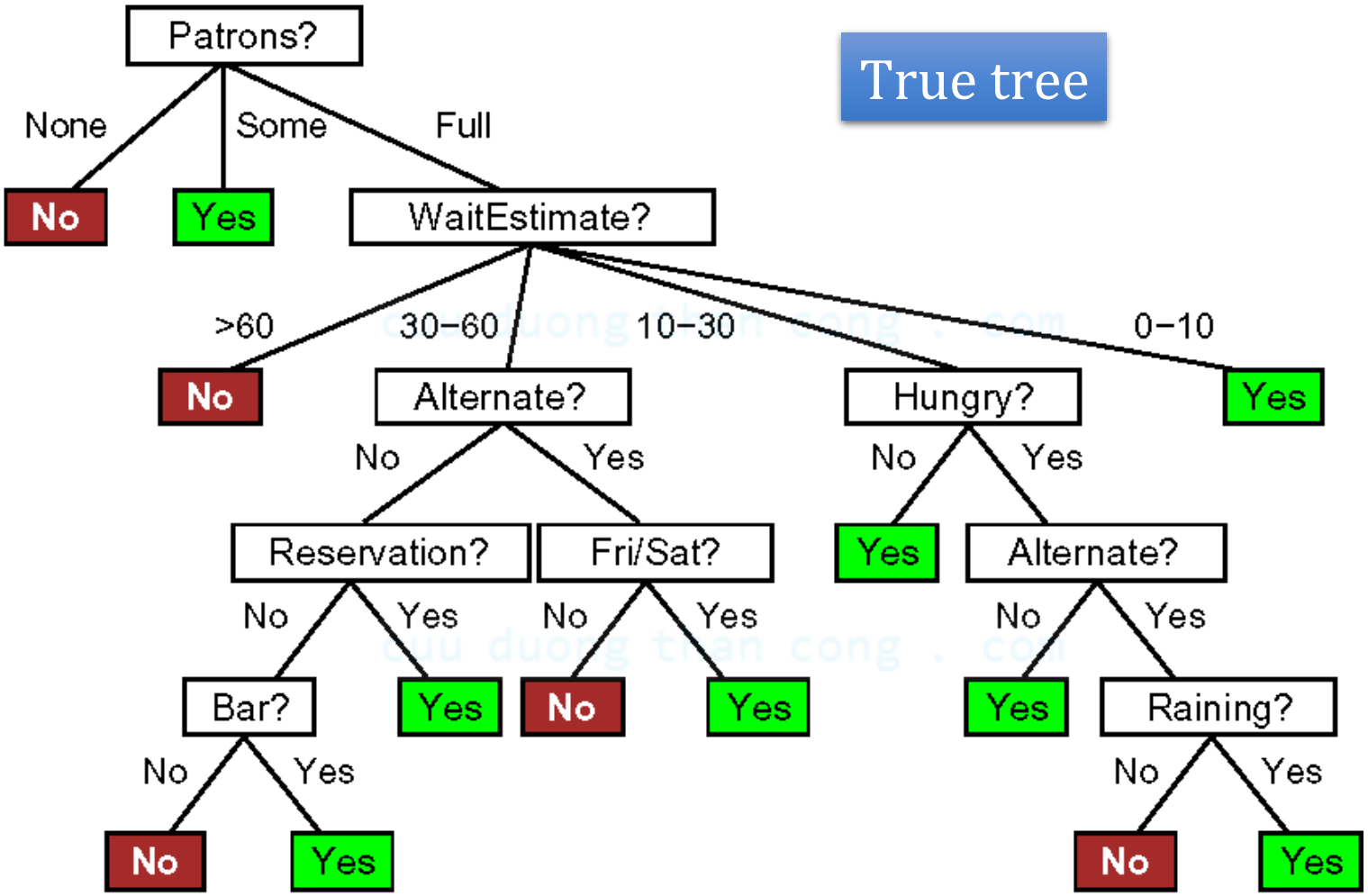
IG(Est.Waiting time, S) = H(S) – AE Est.Waiting time = 1 – 0.792 = 0.208

Decision tree learning example

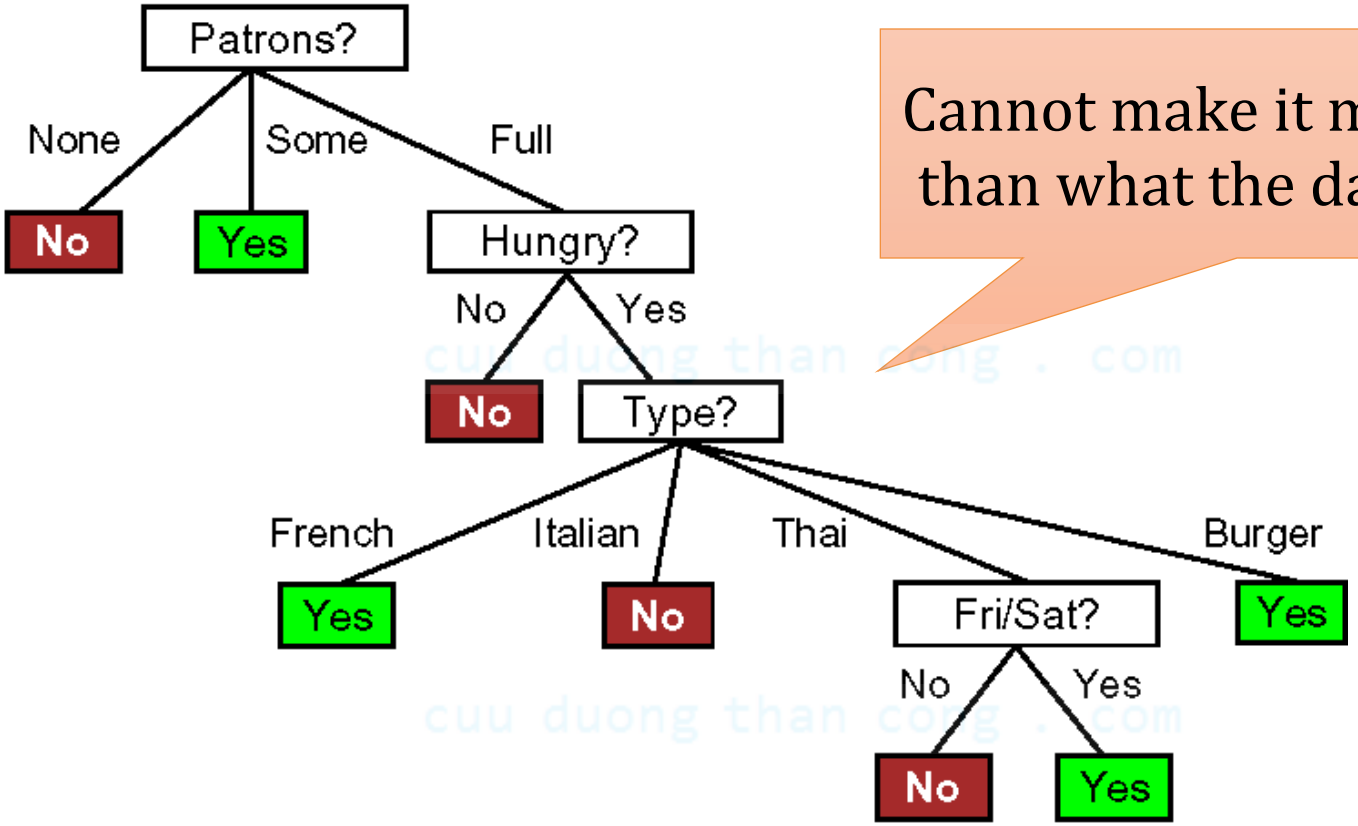


- Largest Information Gain (0.459) / Smallest Entropy (0.541) achieved by splitting on Patrons.
- Continue like this, making new splits, always purifying nodes.

Decision tree learning example



Decision tree learning example



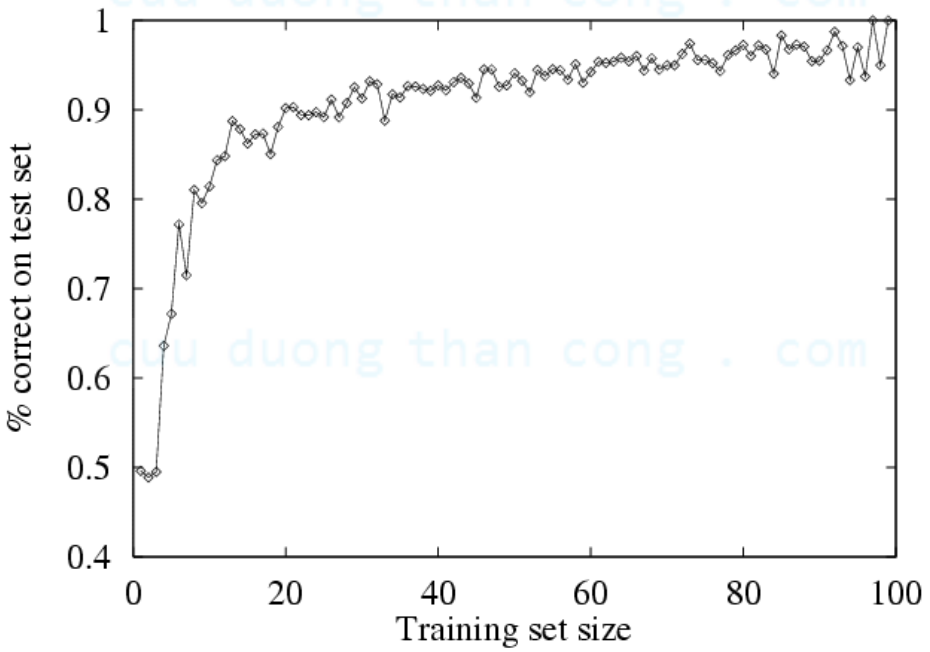
Cannot make it more complex than what the data supports.

Induced tree (from examples)

Performance measurement

- How do we know that $h \approx f$?
  1. Use theorems of computational/statistical learning theory
  2. Try h on a new test set of examples
     (use same distribution over example space as training set)

Learning curve = % correct on test set as a function of training set size



Summary

- Learning needed for unknown environments
- For supervised learning, the aim is to find a simple hypothesis approximately consistent with training examples
- Decision tree learning using information gain
- Learning performance = prediction accuracy measured on test set

Next week

- Individual Assignment 5
- Chapter 4: Learning (cont)
  - Learning Probabilistic Model
  - Artificial Neural Network
- Final Review

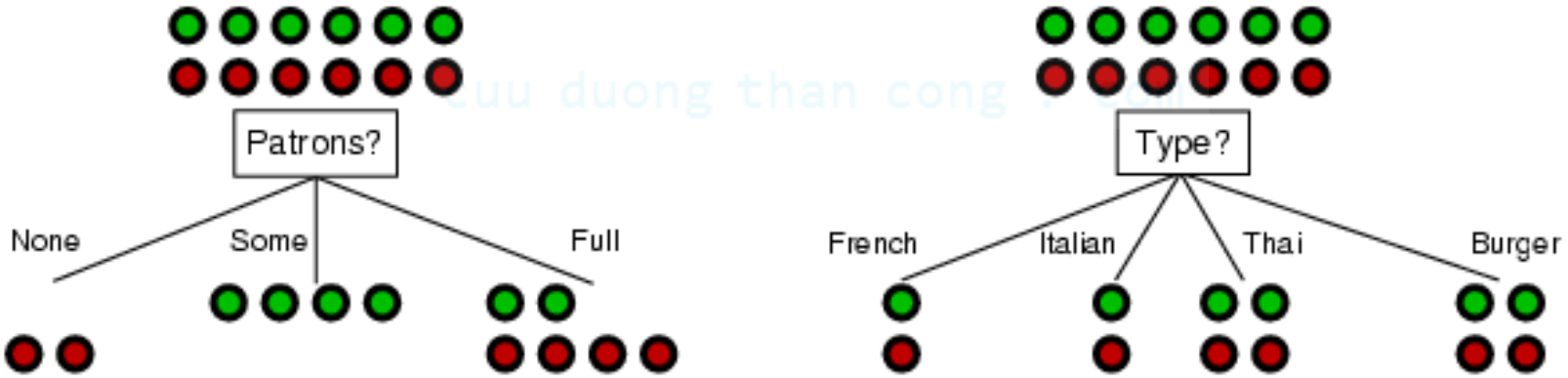
Individual Assignment 4

- Given KB as follows. Prove that there is no pit in square 1,2 (i.e., $\neg P_{1,2}$) using Resolution algorithm (clearly show each pair of sentences to be resolved)

$$KB = (B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})) \wedge \neg B_{1,1}$$

Choosing an attribute

- Idea: a good attribute splits the examples into subsets that are (ideally) "all positive" or "all negative"



- *Patrons?* is a better choice