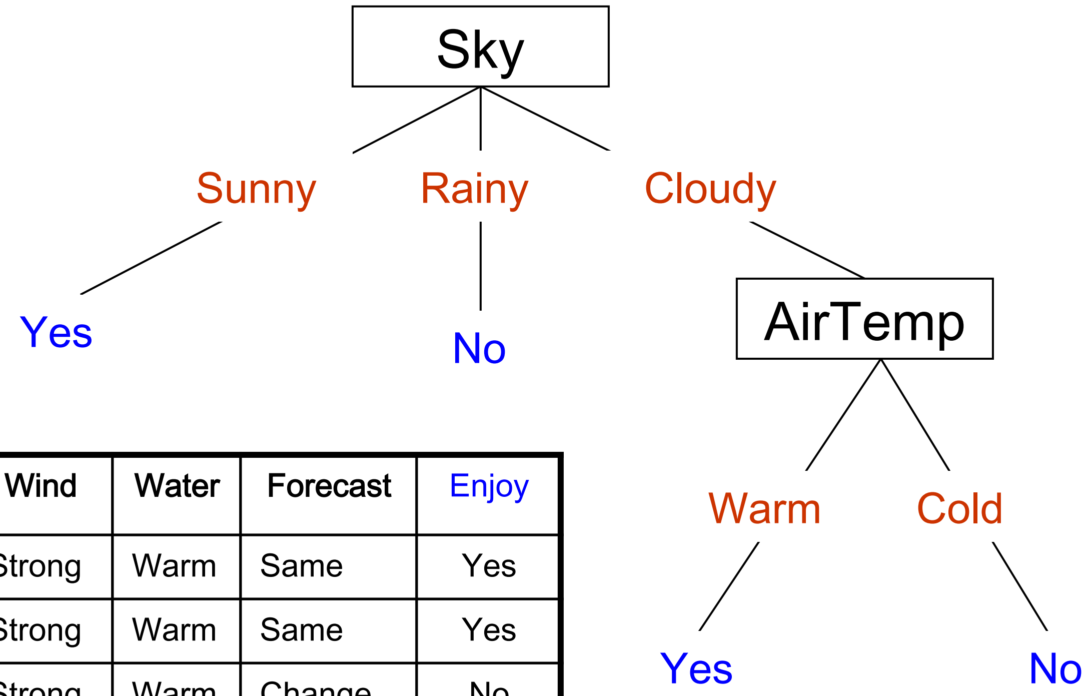
Decision Trees

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes
5	Cloudy	Warm	High	Weak	Cool	Same	Yes
6	Cloudy	Cold	High	Weak	Cool	Same	No

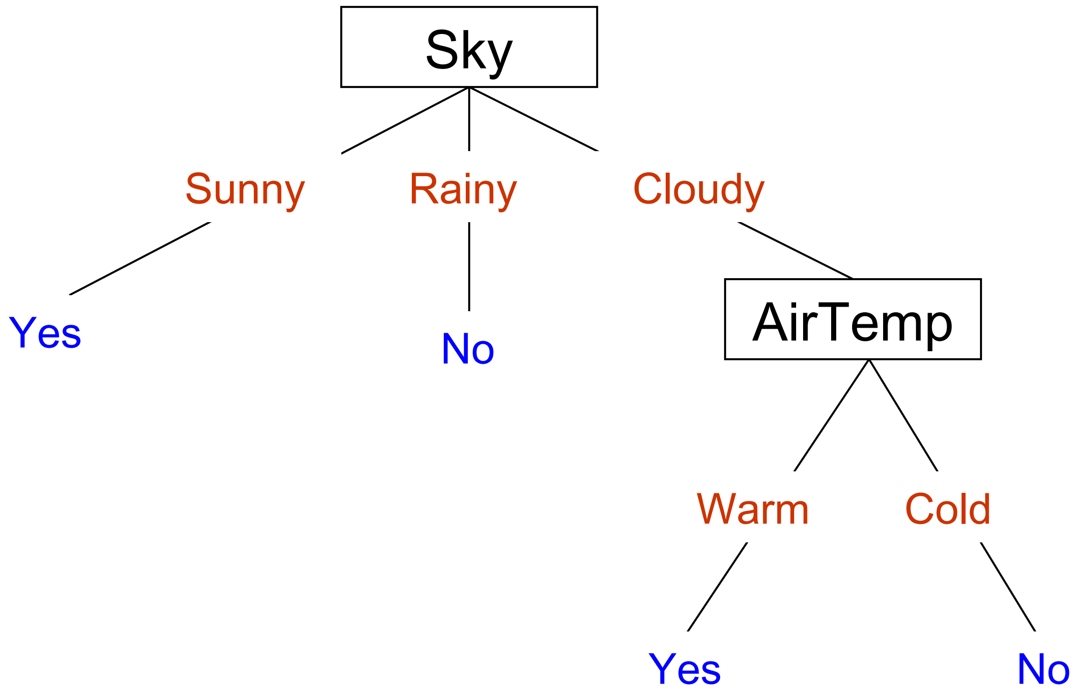
Decision Trees



No.	Sky	AirTemp	Humidity	Wind	Water	Forecast	Enjoy
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes
5	Cloudy	Warm	High	Weak	Cool	Same	Yes
6	Cloudy	Cold	High	Weak	Cool	Same	No

Cao Hoang Tru
CSE Faculty - HCMUT

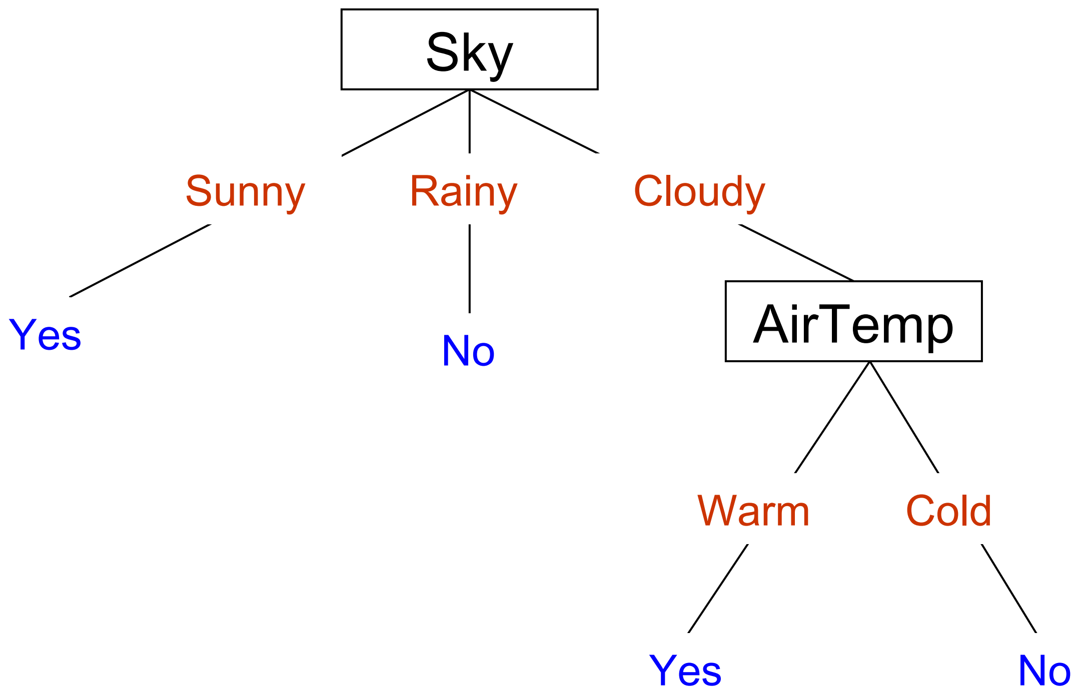
Decision Trees



(Sky = Sunny) ∨ (Sky = Cloudy ∧ AirTemp = Warm)

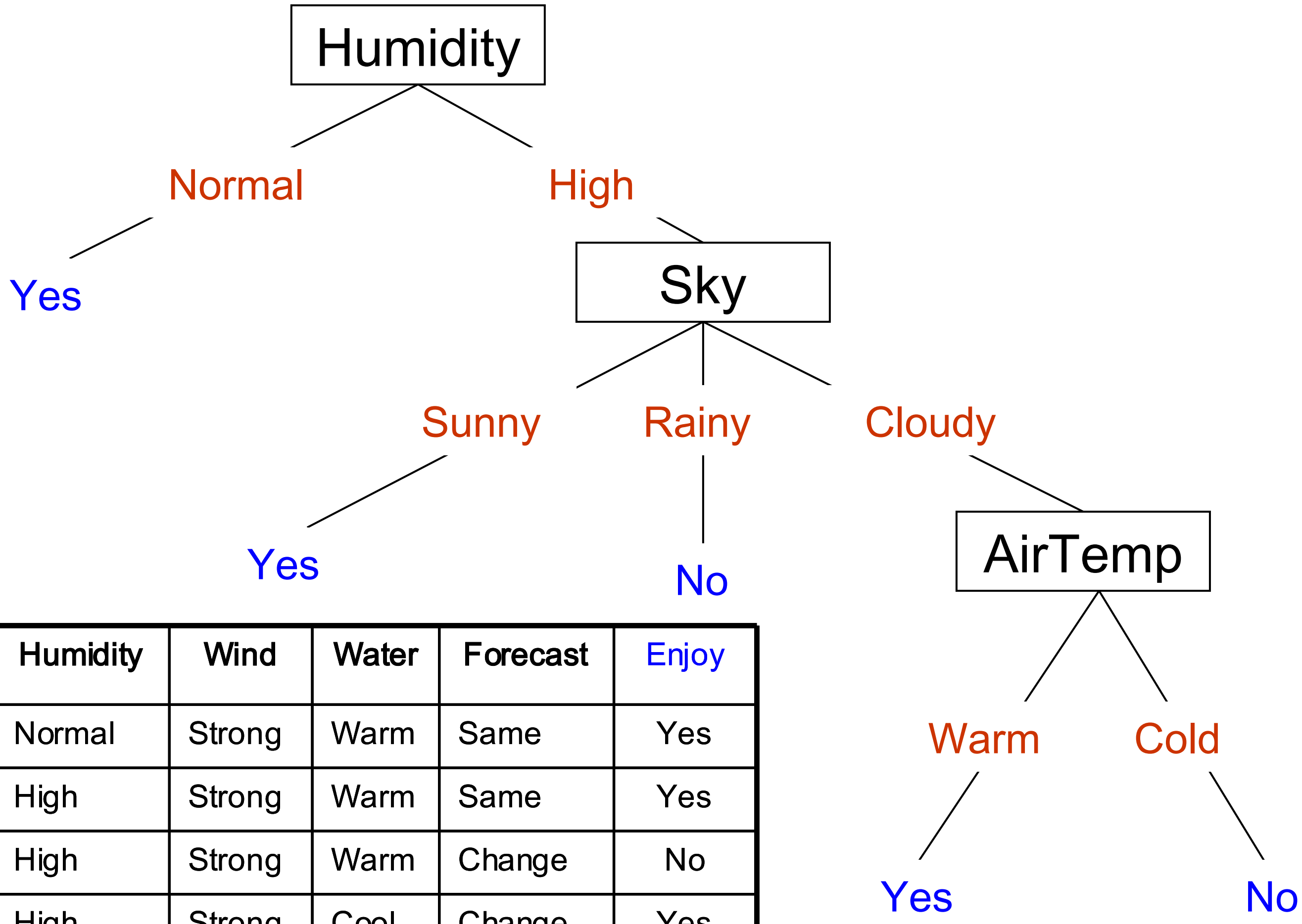
Cao Hoang Tru
CSE Faculty - HCMUT

Decision Trees



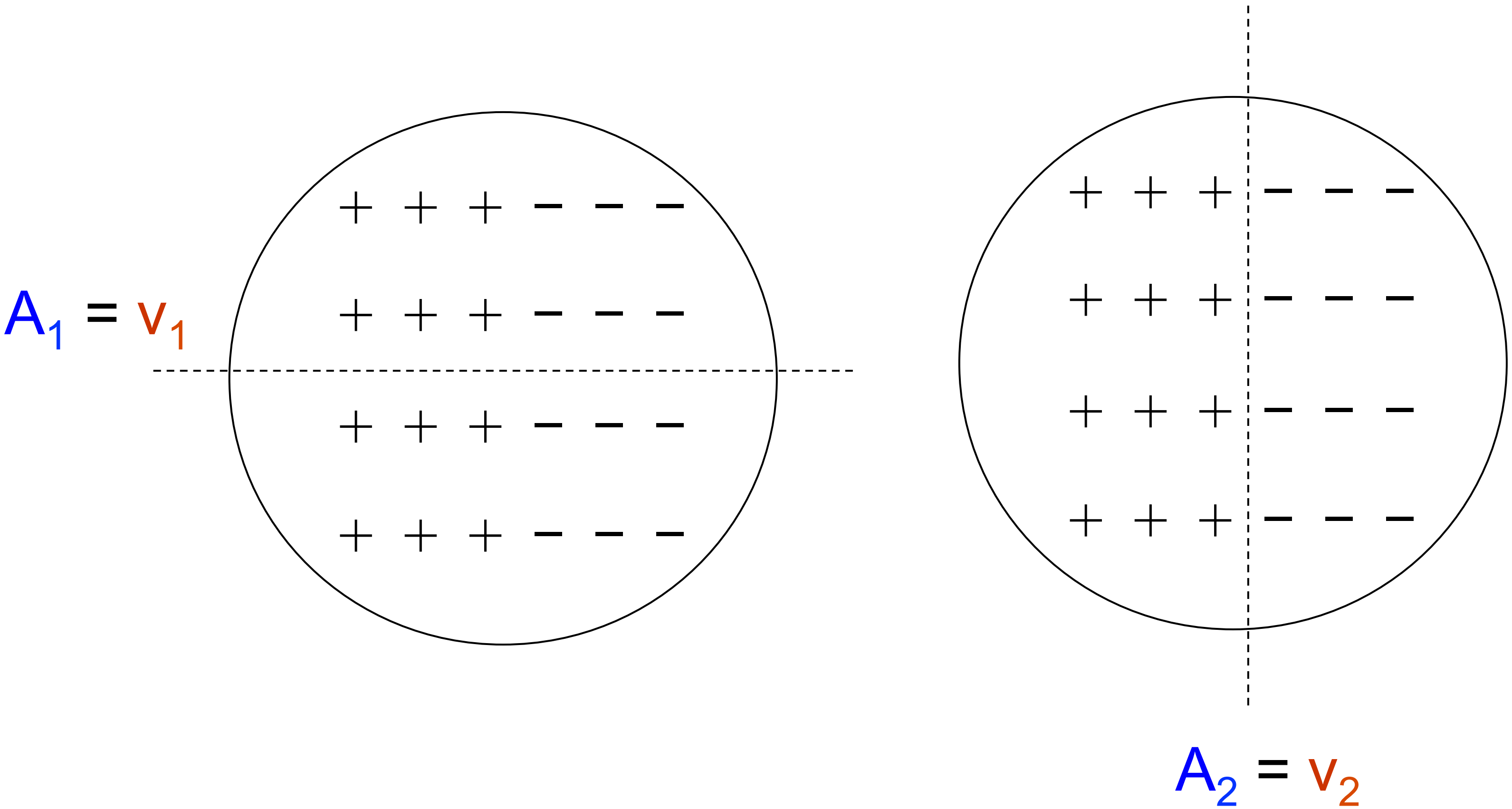
7	Rainy	Warm	Normal	Weak	Cool	Same	?
8	Cloudy	Warm	High	Strong	Cool	Change	?

Decision Trees



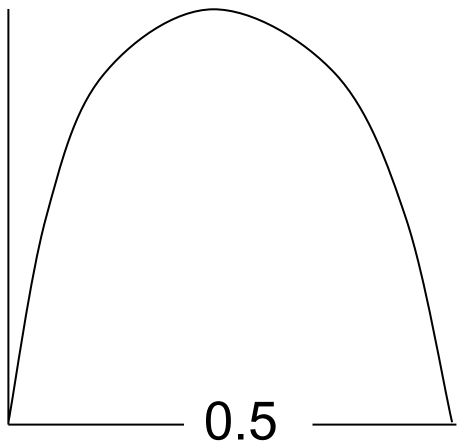
No.	Sky	AirTemp	Humidity	Wind	Water	Forecast	Enjoy
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes
5	Cloudy	Warm	High	Weak	Cool	Same	Yes
6	Cloudy	Cold	High	Weak	Cool	Same	No

Decision Trees



Homogenity of Examples

- Entropy(S) = $-\ p_+\log_2 p_+ - p_-\log_2 p_-$

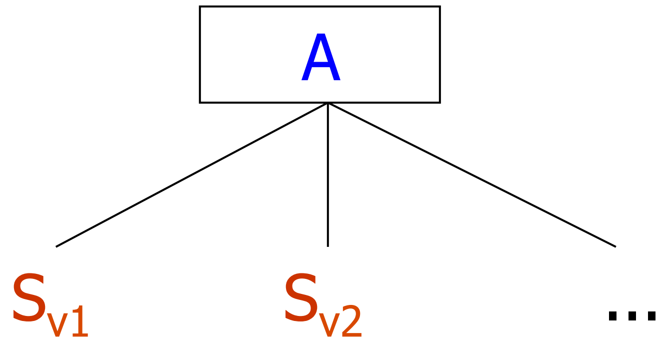


Homogenity of Examples

- Entropy(S) = $\sum_{i=1,c} - p_i \log_2 p_i$ impurity measure

Information Gain

- $Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} (|S_v|/|S|).Entropy(S_v)$



Cao Hoang Tru
CSE Faculty - HCMUT

Example

- Entropy(S) $= - p_+\log_2 p_+ - p_-\log_2 p_- = - (4/6)\log_2(4/6) - (2/6)\log_2(2/6)$

  $= 0.389 + 0.528 = 0.917$

- Gain(S, Sky)

  $= \text{Entropy}(S) - \sum_{v \in \{\text{Sunny, Rainy, Cloudy}\}} (|S_v|/|S|)\text{Entropy}(S_v)$

  $= \text{Entropy}(S) - [(3/6).\text{Entropy}(S_{\text{Sunny}}) + (1/6).\text{Entropy}(S_{\text{Rainy}}) +$

  $(2/6).\text{Entropy}(S_{\text{Cloudy}})]$

  $= \text{Entropy}(S) - (2/6).\text{Entropy}(S_{\text{Cloudy}})$

  $= \text{Entropy}(S) - (2/6)[- (1/2)\log_2(1/2) - (1/2)\log_2(1/2)]$

  $= 0.917 - 0.333 = 0.584$

Example

- $\text{Entropy}(S) = -\,p_+\log_2 p_+ - p_-\log_2 p_- = -\,(4/6)\log_2(4/6) - (2/6)\log_2(2/6)$

  $= 0.389 + 0.528 = 0.917$

- Gain(S, Water)

  $= \text{Entropy}(S) - \sum_{v\in\{\text{Warm, Cool}\}}(|S_v|/|S|)\text{Entropy}(S_v)$

  $= \text{Entropy}(S) - [(3/6).\text{Entropy}(S_{\text{Warm}}) + (3/6).\text{Entropy}(S_{\text{Cool}})]$

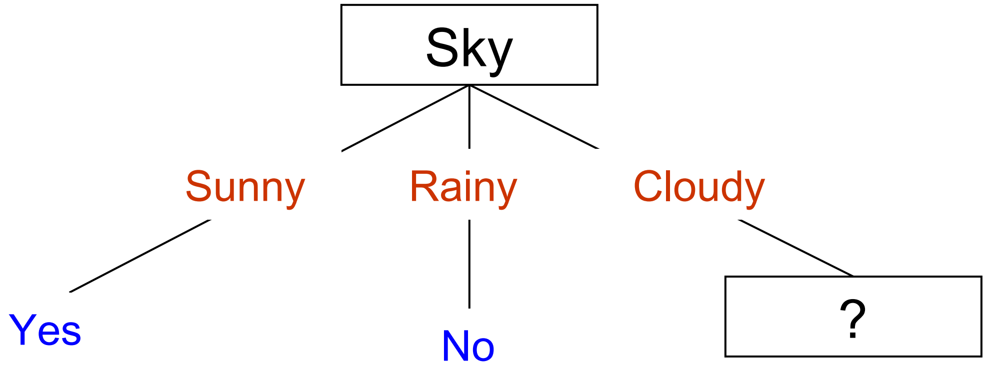
  $= \text{Entropy}(S) - (3/6).2.[-\,(2/3)\log_2(2/3) - (1/3)\log_2(1/3)]$

  $= \text{Entropy}(S) - 0.389 - 0.528$

  $= 0$

Cao Hoang Tru
CSE Faculty - HCMUT

Example



- Gain(S_{Cloudy}, AirTemp)
  $= \text{Entropy}(S_{Cloudy}) - \sum_{v \in \{Warm,\ Cold\}} (|S_v|/|S|)\text{Entropy}(S_v)$
  $= 1$
- Gain(S_{Cloudy}, Humidity)
  $= \text{Entropy}(S_{Cloudy}) - \sum_{v \in \{Normal,\ High\}} (|S_v|/|S|)\text{Entropy}(S_v)$
  $= 0$

Inductive Bias

- Hypothesis space: complete!

Cao Hoang Tru
CSE Faculty - HCMUT

Inductive Bias

- Hypothesis space: complete!
- Shorter trees are preferred over larger trees
- Prefer the simplest hypothesis that fits the data

Inductive Bias

- Decision Tree algorithm: searches incompletely thru a complete hypothesis space.

  ⇒ Preference bias

- Cadidate-Elimination searches completely thru an incomplete hypothesis space.

  ⇒ Restriction bias

Cao Hoang Tru
CSE Faculty - HCMUT