Step ①:

$|S| = 14$ } no of records in dataset

$|S_{yes}| = 9$

$|S_{no}| = 5$

$p_{yes} = 9/14$, $p_{no} = 5/14$

$E(H) = -\sum_{i=1}^{c} p(x_i)\log_2[p(x_i)]$

$= -[p_{yes}\log_2 p_{yes} + p_{no}\log_2 p_{no}]$

$= -\left[\frac{9}{14}\log_2\left(\frac{9}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right]$

$= \underline{0.94}$

Step ②: info gain based on each feature

(w) $|S| = 14$

$$\boxed{I(A) = \sum_{v \in values(A)} \frac{|S_v|}{|S|}\, entropy(S_v)}$$

$$\boxed{gain(S,A) = entropy(S) - I(A)}$$

$|S_{sunny}| = 5$, $|S_{rainy}| = 5$, $|S_{cloudy}| = 4$

$Entropy(S_{sunny}) = -\left[\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right] = 0.971$

$Entropy(S_{rainy}) = -\left[\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right] = 0.971$

$Entropy(S_{cloudy}) = -[\log_2(1)] = 0$

$I(weather) = \frac{5}{14}(0.971) + \frac{5}{14}(0.971) + 0$

$= 0.6935$

$gain(S, weather) = 0.94 - 0.6935$

$= \underline{0.2465}$

2(m) $|S_{happy}| = 4$, $|S_{bored}| = 4$, $|S_{laggy}| = 3$, $|S_{sad}| = 3$

$E(S_{hap}) = -\left[\frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right] = 0.811$

$E(S_{bored}) = -\left[\frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{3}{4}\log_2\left(\frac{3}{4}\right)\right] = 0.811$

$E(S_{lazy}) = -[\log_2(1)] = 0$

$E(S_{sad}) = -\left[\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] = 0.918$

$I(mood) = \frac{4}{14}(0.811) + \frac{4}{14}(0.811) + 0 + \frac{3}{14}(0.918)$

$= 0.6601$

$gain(S, mood) = 0.94 - 0.6601$

$= \underline{0.28}$

2(g) $|S_{com}| = 4$, $|S_{dram}| = 2$, $|S_{th}| = 4$, $|S_{rom}| = 4$

$E(S_{com}) = -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] = 1$

$E(S_{dram}) = -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] = 1$

$E(S_{th}) = -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] = 1$

$E(S_{rom}) = -[\log_2(1)] = 0$

$I(genre) = \frac{4}{14}(1) + \frac{2}{14}(1) + \frac{4}{14}(1) + 0$

$= 0.714$

$gain(S, genre) = 0.94 - 0.714$

$= \underline{0.226}$

2(a) $|S_{no}| = 6$, $|S_{yes}| = 8$

$E(S_{no}) = -\left[\frac{5}{6}\log_2\left(\frac{5}{6}\right) + \frac{1}{6}\log_2\left(\frac{1}{6}\right)\right] = 0.65$

$E(S_{yes}) = -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] = 1$

$I(alone) = \frac{6}{14}(0.65) + \frac{8}{14}$

$= 0.85$

$gain(S, alone) = 0.94 - 0.85$

$= \underline{0.09}$

① $|S| = 5$, $|S_y| = 3$, $|S_n| = 2$

$p_y = \frac{3}{5}$, $p_n = \frac{2}{5}$

$E(S) = -\left[\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right]$

$= 0.971$

②(s)

$|S_1| = 3$, $|S_2| = 2$

$E(S_1) = -\left[\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right] = 0.918$

$E(S_2) = -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] = 1$

$I(salary) = \frac{3}{5}(0.918) + \frac{1}{5}$

$= 0.9508$

$gain(S, salary) = 0.019$

(l) $|S_{mum}| = 2$, $|S_{blr}| = 2$, $|S_{hyd}| = 1$

$E(S_{mum}) = 0$

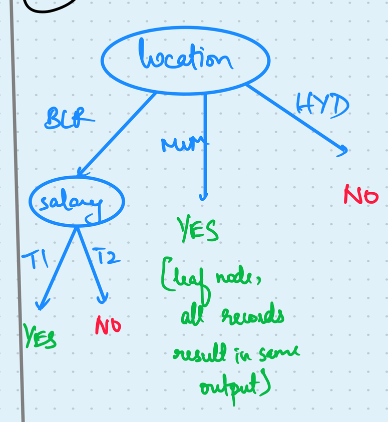
$E(S_{blr}) = 1$

$E(S_{hyd}) = 0$

$I(location) = \frac{2}{5} = 0.4$

$gain(S, loc) = 0.571$

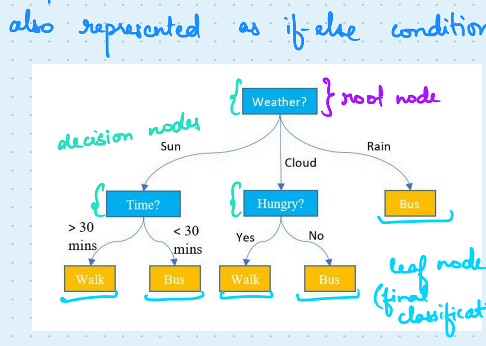
③



ID3

Decision Trees

- approximates discrete-valued target function
- also represented as if-else conditions



$\boxed{\text{represents a disjunction of conjunctions}}$

Entropy

$H(A) = -\sum p(x_i)\log_2[p(x_i)]$

$I(A) = \sum_{v \in values(A)} \frac{|S_v|}{|S|}\, entropy(S_v)$

$gain(S,A) = Entropy(S) - I(A)$

crux of the problem

ID 3

CART

decision tree

Find-S algorithm

version space

Concept Learning

model learning how to actually categorise data into categories

⇅

searching through hypothesis space for the hypothesis that best fits the given example

Algorithm:

① Initialise h to most specific hypothesis in h-space

$h = \{\phi, \phi .. \}$

② For each learning example (record in training data): → instance x

- If negative → skip
- If positive and hypothesis too specific → make more general

③ Repeat ② until all learning examples exhausted

Example

EXAMPLE	COLOR	TOUGHNESS	FUNGUS	APPEARANCE	POISONOUS
1.	GREEN	HARD	NO	WRINKELD	YES
2.	GREEN	HARD	YES	SMOOTH	NO
3.	BROWN	SOFT	NO	WRINKLED	NO
4.	ORANGE	HARD	NO	WRINKLED	YES
5.	GREEN	SOFT	YES	SMOOTH	YES
6.	GREEN	HARD	YES	WRINKLED	YES
7.	ORANGE	HARD	NO	WRINKLED	YES

ATTRIBUTES ON WHICH THE CONCEPT DEPENDS ON — CONCEPT

⓪ $h = \{\phi, \phi, \phi, \phi\}$

① $h = \{green, hard, no, wrinkled\}$

② [skip]

③ [skip]

④ $h = \{?, hard, no, wrinkled\}$

⑤ $h = \{?, ?, ?, ?, ?\}$

Limitations

- Outputs only one of many possible hypotheses
- May not perform consistency across entire dataset
- Sensitive to inconsistent datasets → sole reliance on positive examples

ALTERNATIVE: Candidate Elimination Algorithm

0: $h = \{\phi, \phi, \phi, \phi\}$

1: skip

2: senior subrub fullstack medium

3: senior ... fullstack ...

4: senior ? fullstack medium/large ?

ϕ, ϕ, ϕ, ϕ

adult, laptop, yes, high

?, , , ?

?, laptop, yes, ?

Consistent hypothesis: h consistent wrt training data if it classifies all objects of training dataset to corresponding classes

$\boxed{h(x_i) = C(x_i)\ \forall x_i \in \text{train set}}$ } h-space = concept space

Version space: Subset of hypothesis space that contains all hypotheses consistent with training data

boundaries

specific boundary (S)
set of maximally specific examples that are consistent with positive examples

general boundary (G)
set of maximally general examples that are consistent with both positive & negative examples