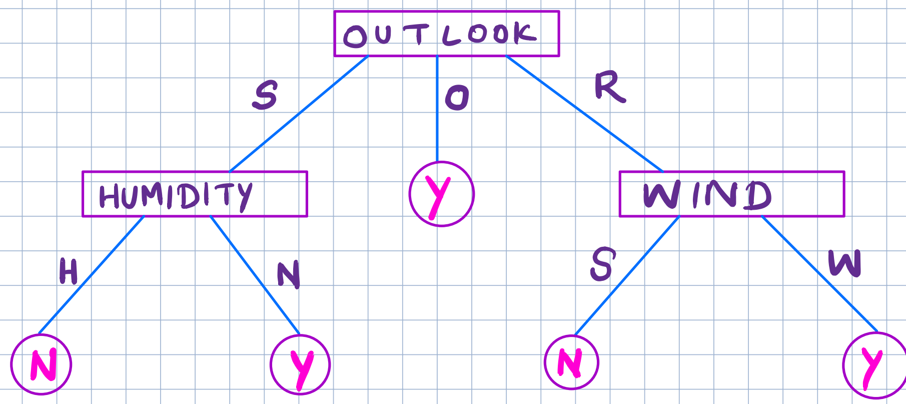
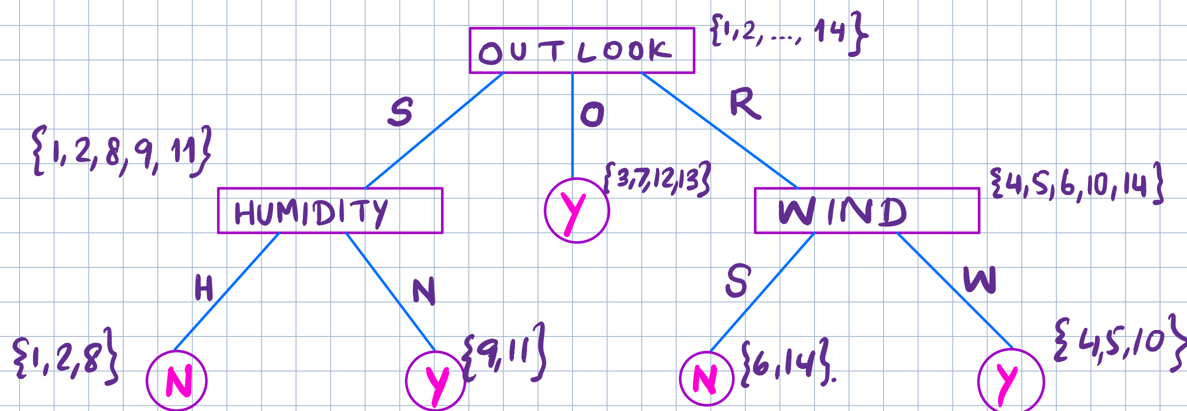

OUTLOOK
S
O
R
HUMIDITY
Y
WIND
H
N
S
W
N
Y
N
Y


OUTLOOK
{1,2,..., 14}
S
O
R
{1,2,8,9, 11}
HUMIDITY
Y
{3,7,12,13}
WIND
{4,5,6,10,14}
H
N
S
W
{1,2,8}
N
Y
{9,11}
N
{6,14}
Y
{4,5,10}

Lec-02 1/27/2020.

Decision Trees. (Ch3-Mitchell, Ch1-CIML).

Recap :-

- ERM :- design algorithms that work well on 'S' (training set), and hope it works well on 'T' (test set).

- Entropy :- measure of randomness of distributions.

$$H(p) := \sum_x p(x) \log \frac{1}{p(x)}, \quad p(x) \text{ - probability of } x.$$

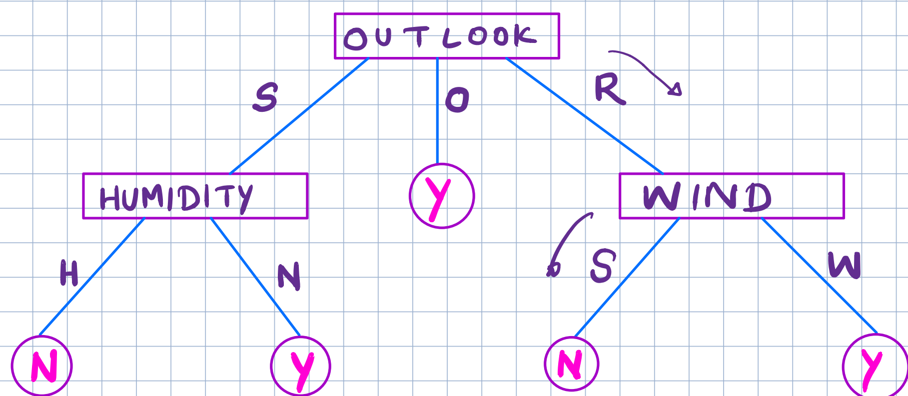
Higher entropy $\Leftrightarrow$ more randomness
$\Leftrightarrow$ more uncertainty about outcomes.

Features :- (Outlook, Temp, Humidity, Wind)
4 attributes.

Labels :- Y/N

- What are decision trees?
- How to infer/predict using decision trees?
- How to design a "good" decision tree?

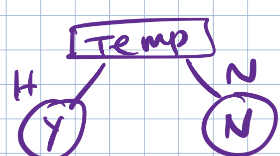
A **decision tree** is a tree, each of whose "non-leaf" node represents an "attribute", edges represent values of the attributes of the parent node, and leaf nodes correspond to labels.



- How to infer using decision trees?

  (Rain, High, High, Strong).

  Traverse the tree from root node, output the leaf node you end at.



- Q:- How to design a "good" decision tree from training data?

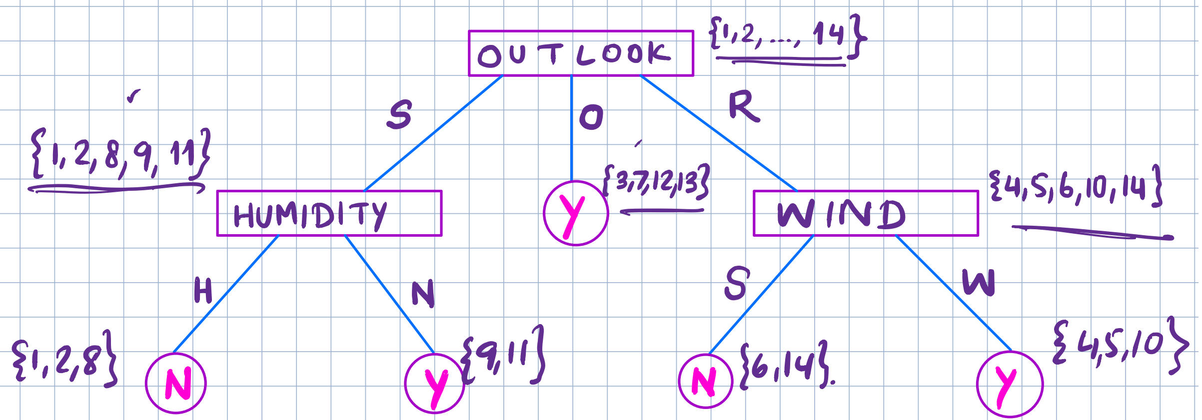
  - Our goal is to map features into labels.

    "Some attributes" are more informative about the labels than others.

- Steps to design a decision tree (from data):-

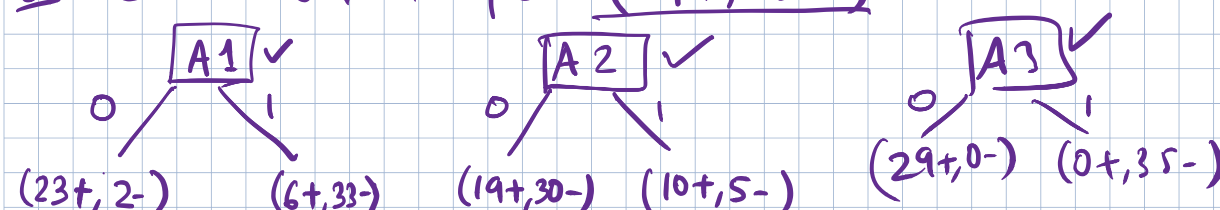
Start with a root node:- (put all examples there.)

LOOP:-
1. Select a "good" attribute for the node
2. For each possible value of attribute, create a child node, (with corresponding) edges, put examples in corresponding nodes.
3. If a node has all examples with same label, $\boxed{\text{STOP}}$



Hierarchical partitioning

Ex:- S ⟶ 64 examples (29+, 35−)   A3>A1>A2



- Select an attribute that divides the examples of that node into subsets that are as pure as possible.
  - shorter trees.

    (OCCAM'S RAZOR) ⟶ simplest explanation is best.

helps prevent $\boxed{\text{over fitting}}$.

Information gain (entropy of labels) (Qiunlan '75).

(*) reduction in randomness of labels in the examples present at a node

'S' $\longrightarrow$ set of examples $(n = |S|)$

labels $\to \{1, \dots, \ell\}$

'A' $\longrightarrow$ attribute (columns of 'S').

$IG(\underline{\underline{S}}; A) \longrightarrow$

'S' $\longrightarrow$ 'n',

n_i = # examples in S with label 'i'

$n_1 + n_2 + \dots + n_\ell = \underline{\underline{n}}$

empirical distribution of labels of 'S' : $\boxed{\left(\frac{n_1}{n}, \frac{n_2}{n}, \dots, \frac{n_\ell}{n}\right)}$

$$\boxed{H(S) := \sum_{i=1}^{\ell} \left(\frac{n_i}{n}\right) \cdot \log\left(\frac{n}{n_i}\right).}$$

'A' $\longrightarrow$ OUTLOOK $\longrightarrow \{S, O, R\}$ $\to$ $a = 3$,

Humidity $\longrightarrow \{H, N\}$ $\to$ $a = 2$

'A' $\longrightarrow \{v_1, v_2, \dots, v_a\}$.

$$\boxed{IG(S; A) = H(S) - \overbrace{\sum_{j=1}^{a} \frac{|S_{v_j}|}{|S|} \cdot H(S_{v_j})}^{\text{cond. entropy.}}}$$

'S' $\longrightarrow S_{v_1}$, S_{v_2}, $\vdots$, S_{v_a}

$$\boxed{|S_{v_1}| + \dots + |S_{v_a}| = |S|}$$

(2+, 88-)

(99999+, 2-)

Tennis:- $(5N, 9Y)$ $n_1 = 5$, $n_2 = 9$, 14

$$H(S) = \frac{5}{14}\cdot\log\frac{14}{5} + \frac{9}{14}\log\frac{14}{9} = 0.94\ldots$$

Humidity — 'A', $\{N, H\}$ $\{v_1, v_2\}$

$\underbrace{\{6+, 1-\}}_{S_{v_1}}$ $\underbrace{\{3+, 4-\}}_{S_{v_2}}$

$$H(S_{v_1}) = \frac{6}{7}\log\frac{7}{6} + \frac{1}{7}\log\frac{7}{1} \approx$$

$$H(S_{v_2}) = \frac{3}{7}\log\frac{7}{3} + \frac{4}{7}\log\frac{7}{4}.$$

$$IG(S; \text{Humidity}) = 0.94 - \frac{7}{14}(\quad) - \frac{7}{14}\cdot(\quad)$$

$$= 0.151. \checkmark$$

$$IG(S; \text{wind}) = 0.048,\ldots \checkmark$$

Ex:- compute $IG(S; \text{OUTLOOK})$, $IG(S; \text{Temp})$

verify $(\checkmark, \checkmark)$

Complete steps to get a decision tree.