# Naive Bayes classifier

$$P(y=c|x) \propto \underline{\underline{P(y=c)}} \cdot \prod_{j=1}^{n} P(x_j | y=c)$$

| Day | Outlook | Temp | Humidity | Wind | PlayTennis ($y$) |
|---|---|---|---|---|---|
| D1 | Sunny | Hot | High | Weak | No |
| D2 | Sunny | Hot | High | Strong | No |
| D3 | Overcast | Hot | High | Weak | Yes |
| D4 | Rain | Mild | High | Weak | Yes |
| D5 | Rain | Cool | Normal | Weak | Yes |
| D6 | Rain | Cool | Normal | Strong | No |
| D7 | Overcast | Cool | Normal | Strong | Yes |
| D8 | Sunny | Mild | High | Weak | No |
| D9 | Sunny | Cool | Normal | Weak | Yes |
| D10 | Rain | Mild | Normal | Weak | Yes |
| D11 | Sunny | Mild | Normal | Strong | Yes |
| D12 | Overcast | Mild | High | Strong | Yes |
| D13 | Overcast | Hot | Normal | Weak | Yes |
| D14 | Rain | Mild | High | Strong | No |

$c \in \{No, Yes\}$

$P(y=Yes) = \frac{9}{14}$ ✓

$P(y=No) = \frac{5}{14}$ ✓

## Outlook

$P(outlook = sunny \mid y=Y)$

| | Y | N |
|---|---|---|
| Sunny | 2/9 | 3/5 |
| Overcast | 4/9 | 0 |
| Rain | 3/9 | 2/5 |

## Temp

| | Y | N |
|---|---|---|
| hot | 2/9 | 2/5 |
| mild | 4/9 | 2/5 |
| cold | 3/9 | 1/5 |

## Humidity

| | Y | N |
|---|---|---|
| high | 3/9 | 4/5 |
| Normal | 6/9 | 1/5 |

Windy

| | y | n |
|---|---|---|
| Strong | 3/9 | 3/5 |
| Weak | 6/9 | 2/5 |

$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong \rangle$

$$P(y = Yes \mid x) = P(y = Yes) \cdot P(outlook = sunny \mid y = Yes) \cdot P(temp = cold \mid y = Yes) \cdot P(Humidity = high \mid y = Yes) \cdot P(wind = strong \mid y = Yes)$$

$$= \frac{\cancel{9}}{\cancel{14}\ 7} \cdot \frac{\cancel{2}}{\cancel{9}} \cdot \frac{\cancel{3}}{\cancel{9}\ 3} \cdot \frac{\cancel{3}}{\cancel{9}\ 3} \cdot \frac{\cancel{3}}{\cancel{9}\ 3}$$

$$= \frac{1}{7} \cdot \frac{1}{3 \cdot 3 \cdot 3} = \frac{1}{189} = 0.0053$$

$$P(y = No \mid x) = P(y = No) \cdot P(outlook = sunny \mid y = No) \cdot P(temp = cold \mid y = No) \cdot P(Humidity = high \mid y = No) \cdot P(wind = strong \mid y = No)$$

$$= \frac{5}{14} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5}$$

$$= \frac{18}{875} = 0.0206$$

Belongs to class NO.

# Decision Trees

↳ supervised Algo

↳ Classification & Regression

Classification: $y \in$ discrete set

Regression: $y \in \mathbb{R}$

x values / features

↳ discrete values

$x_1$ : outlook $\in$ {Sunny, overcast, Rainy}

$x_2$ : humidity $\in$ {High, Normal}

$x_3$ : wind $\in$ {strong, weak}

$x_4$ : temperature $\in$ {Hot, Moderate, Cold}

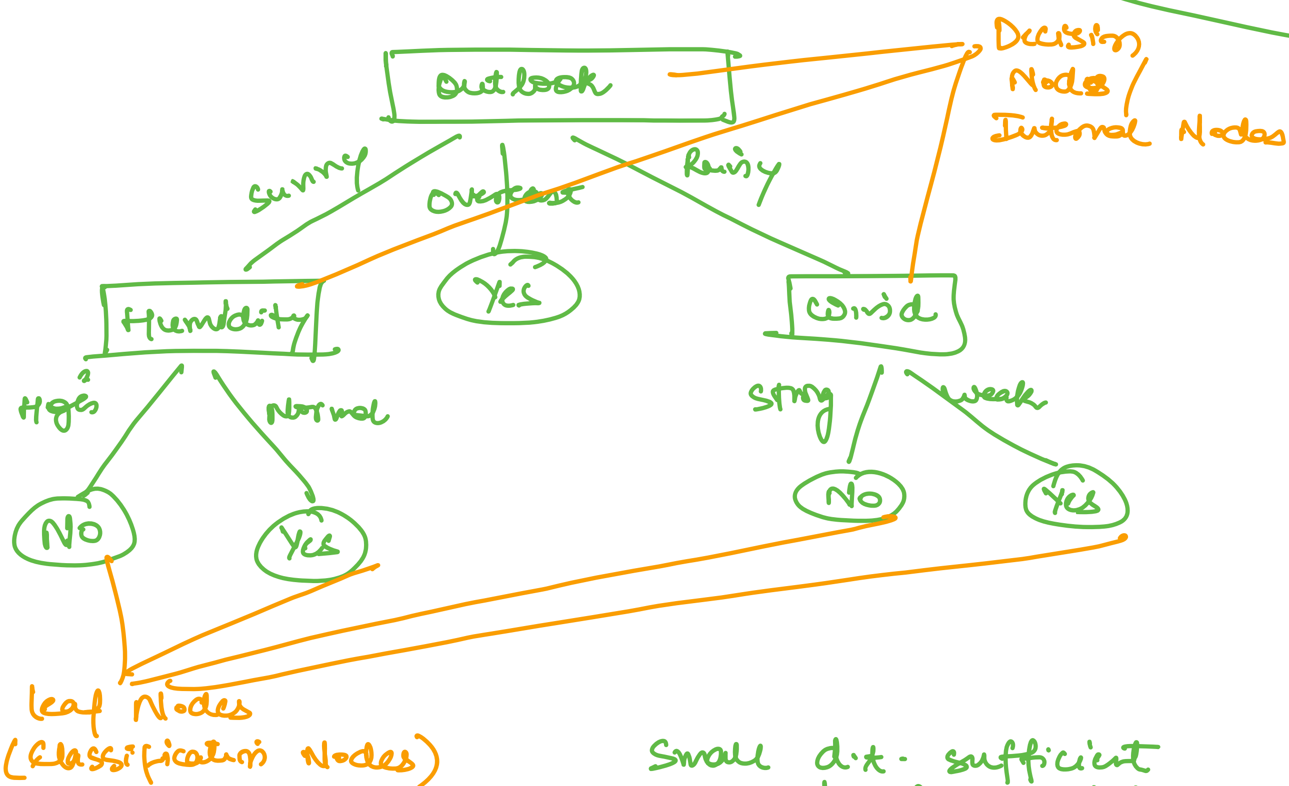
Discrete, Continuous / Numeric

0, 1, 2 . . .

One hot encoding

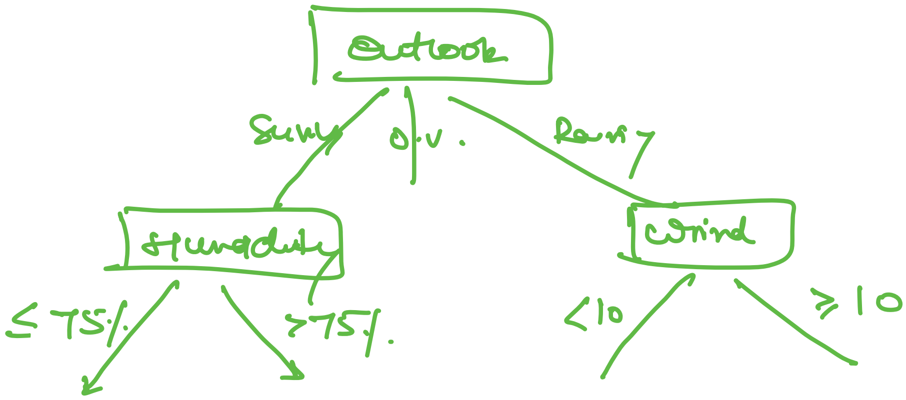
$x_j$ = {hot, moderate, cold}

1 0 0
0 1 0
0 0 1

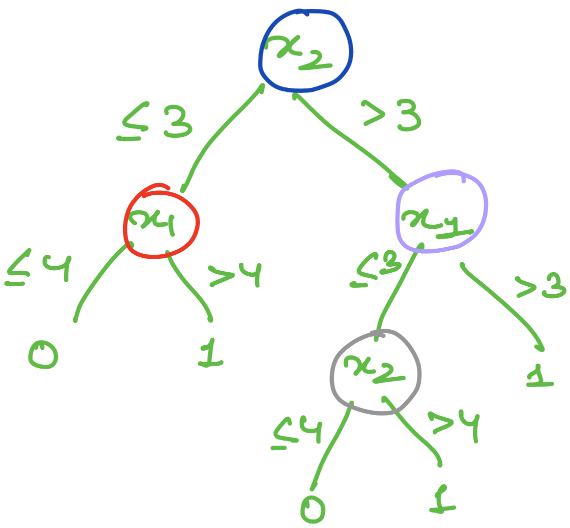


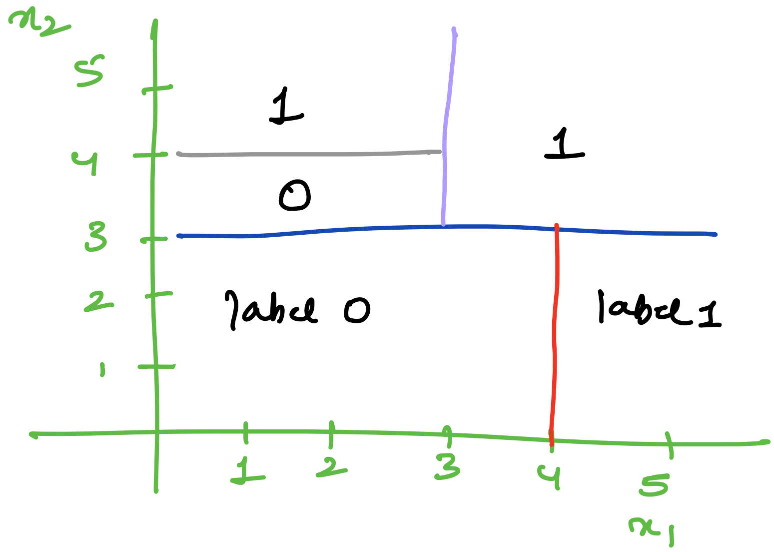

Small d.t. sufficient enough to fit the data.

* DT can handle continuous values




* In continuous values decision surface consist of axis parallel lines (Rectangles)




node

multiple comp ?

?

$2^3 = \underline{8}$

$2^{20}$

BUILD A DECISION TREE?

come at a node

↳ ① data is "pure"

↳ make it a leaf node

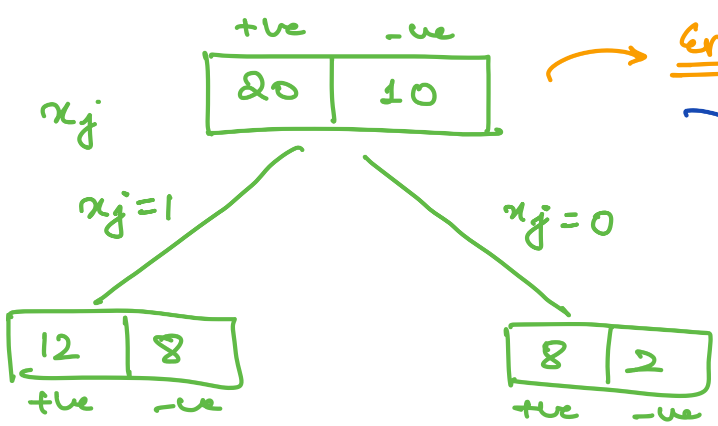
② data +ve, -ve

? Decide the attribute to split

split the data

Recurse on each branch

## Attribute to split on?




Error : leaf node classification

+ve class assign

Error: 10

Randomness: $\frac{10}{30} = 33.3\%$

label: +ve

Error: 8

Randomness: $\frac{8}{20} = \frac{2}{5} = 40\%$

label: +ve

Error: 2 Randomness: $\frac{2}{10} = 20\%$

$8+2 = 10$

$\boxed{\text{Entropy}}$: Metric to measure randomness

(Error is not ↓ but entropy of the system will go down)

$y \in \{1 \ldots r\}$ $\qquad P(y=k) = p_k$

$$\boxed{H(y) = -\sum_{k=1}^{k} p_k \log p_k}$$

Entropy of random variable $y$

$-E(\log p_k)$

Boolean case: $\quad y \in \{0, 1\}$

$p_0 \quad p_1$

$p_0 + p_1 = 1$

$$H(y) = -\left(p_0 \log p_0 + p_1 \log p_1\right)$$
$$= -\left(p_1 \log p_1 + (1-p_1)\log(1-p_1)\right)$$

$p \to 0 \qquad \underset{0}{p} \log \underset{-\infty}{p} \to \dfrac{\log p}{1/p} = \dfrac{\frac{\partial}{\partial p}(\log p)}{\frac{\partial}{\partial p}(1/p)} = \dfrac{1/p}{-1/p^2}$

$= -p \to 0$

$p \to 0 \quad p \log p \to 0$

$p \to 1 \quad p \log p \to 0$

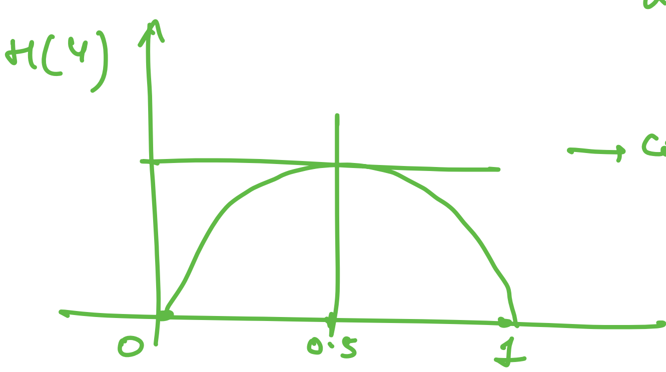
maxima, minima?

$$\frac{d}{dp_1}\left(p_1 \log p_1 + (1-p_1)\log(1-p_1)\right)$$

$$\log p_1 + \frac{p_1}{p_1} - \frac{(1-p_1)}{(1-p_1)} - \log(1-p_1)$$

$$\log p_1 + \cancel{1} - \cancel{1} - \log(1-p_1) = 0$$

$$\log p_1 = \log(1-p_1)$$
$$p_1 = 1 - p_1$$
$$2p_1 = 1 \quad \Rightarrow p_1 = 0.5 = 1/2$$




$\rightarrow$ concave $f^n$

$y \in \{1 \dots r\}$

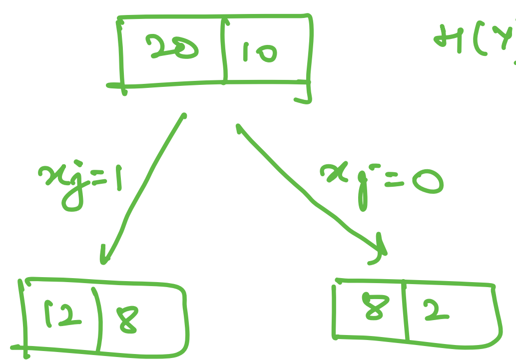
$H(Y) = -\sum_{k=1}^{r} p_k \log p_k$ is maximum at $p_k = 1/r \ \forall k$

$H(Y) \rightarrow$ Entropy is maximum when the distribution is uniform

<u>Goal:</u> reduce Entropy of class labels.

Before split

$$H(Y) = -\left(\frac{20}{30}\log\frac{20}{30} + \frac{10}{30}\log\frac{10}{30}\right)$$




$$H(Y|x_j=0) = -\left(\frac{8}{10}\log\frac{8}{10} + \frac{2}{10}\log\frac{2}{10}\right)$$

$$H(Y|x_j=1) = -\left(\frac{12}{20}\log\frac{12}{20} + \frac{8}{20}\log\frac{8}{20}\right)$$

<u>Combine:</u>

1 0

$$\boxed{H(Y|x_j) = \sum_g P(x_j=g)\, H(Y|x_j=g)}$$

$$= P(x_j=1)\cdot H(Y|x_j=1) + P(x_j=0)\cdot P(Y|x_j=0)$$

$$= -\left(\frac{20}{30}\cdot\left(\frac{12}{20}\log\frac{12}{20} + \frac{8}{20}\log\frac{8}{20}\right) + \frac{10}{30}\left(\frac{8}{10}\log\frac{8}{10} + \frac{2}{10}\log\frac{2}{10}\right)\right)$$

After split

Mutual Information $= H(Y) - H(Y|x_j)$

$MI(Y, x_j) =$

less entropy → Entropy ↓ → Pure Info

MI maximize

$\boxed{\text{Gini Index:}}$

$$H(Y) = -\sum_{k=1}^{r} p_k \log p_k \qquad p_k = P(y=k)$$

$$Gini(Y) = 1 - \sum_{k=1}^{r} p_k^2$$

$p_k \cdot p_k$

Earlier there was $\log p_k$ now we have $p_k$ instead

→ DT Q's
→ Metrics

} 2hrs