

Welcome  
to  
ML 1.2

will start at 9:05 PM

## Agenda:

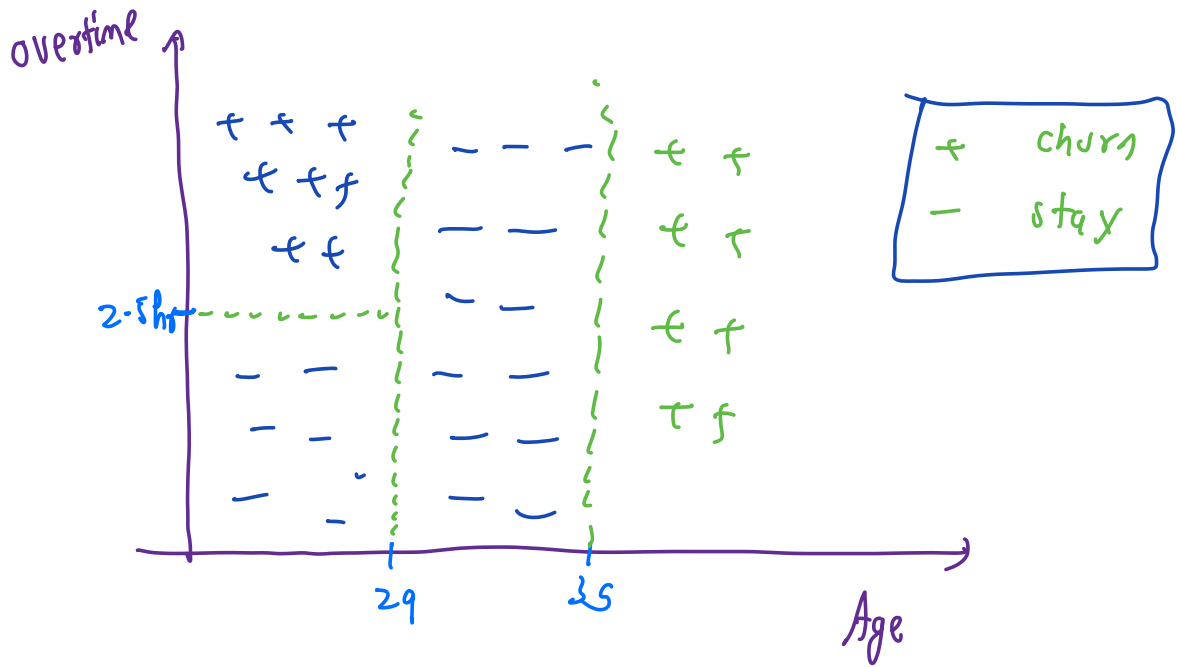
- Use Case
- Decision Tree Intuition
- Entropy
  - Weighted Entropy
  - Information Gain

- Bookmarks ✓
- Notes
  - class Notes
- Agenda

Case study: Tio Attrition

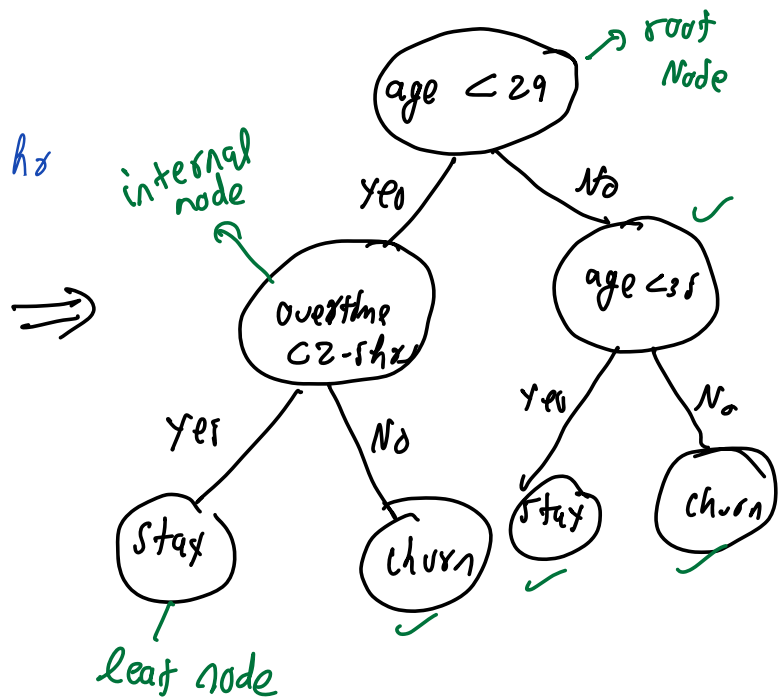
DT intuition

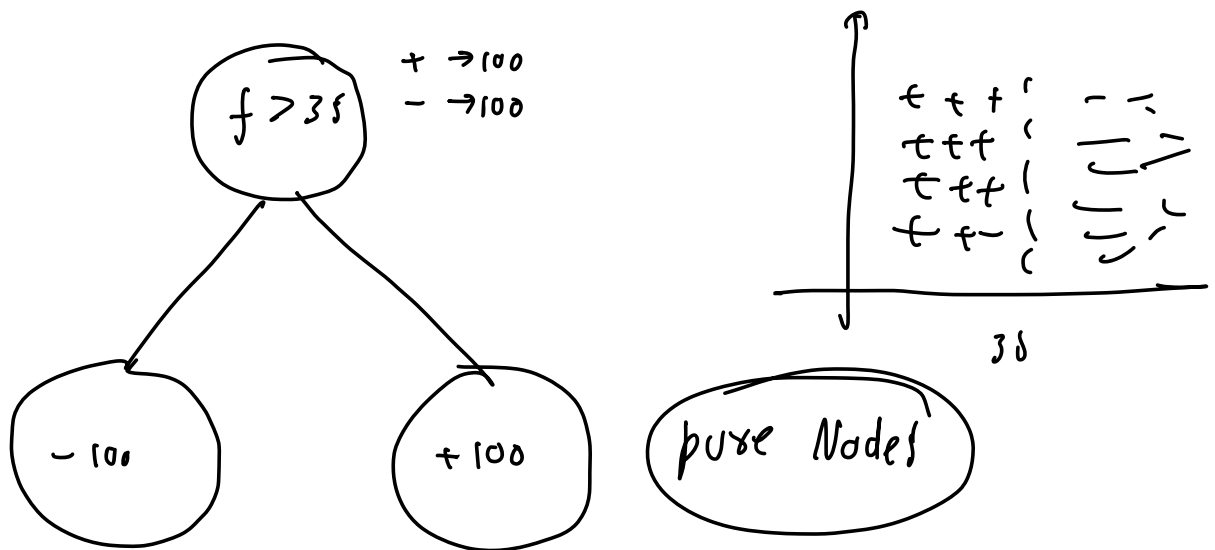
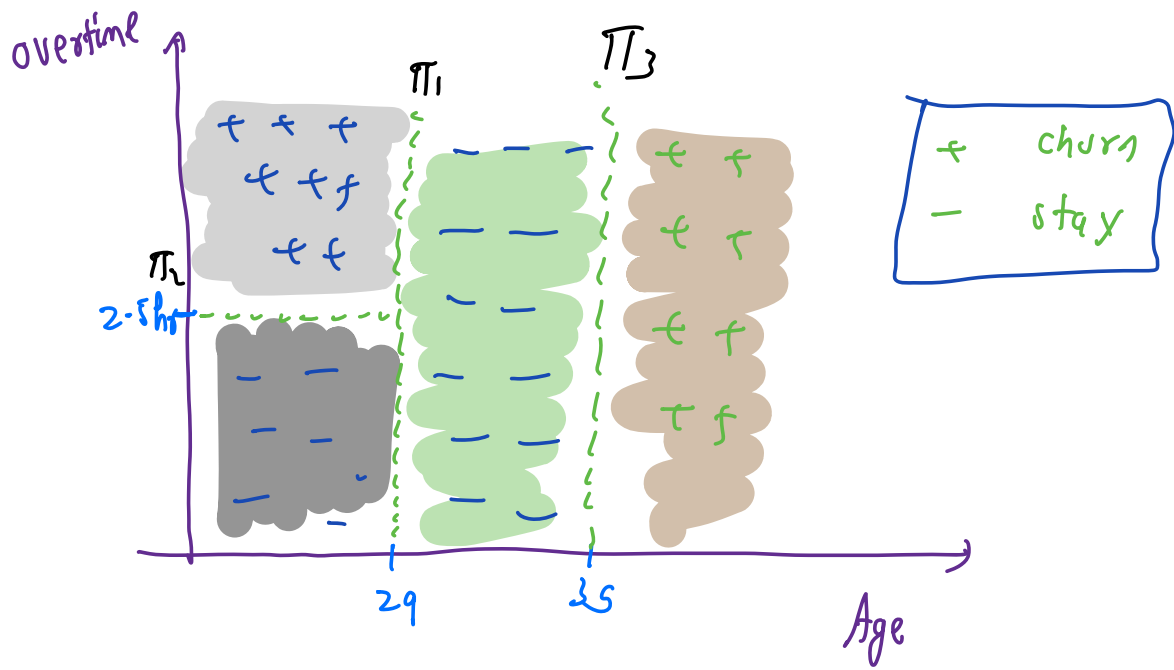
Age  
Overtime ] → Attrition



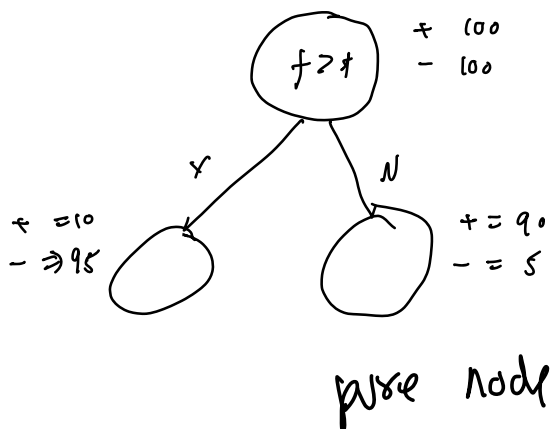
```

if age < 29:
    if overtime < 2.5 hr
        stay
    else :
        churn
else :
    if age < 35:
        stay
    else :
        churn
  
```

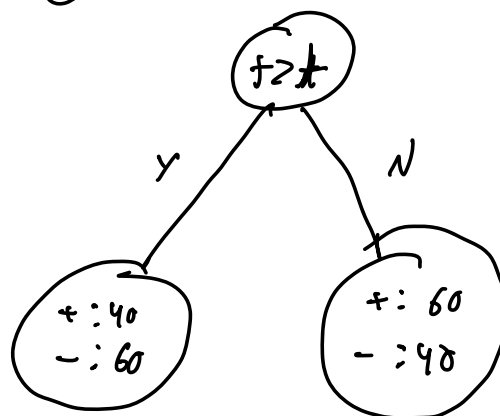




①



②



Entropy : Impurity

$\Downarrow$

measure impurity

$$H(y) = - \sum_{i=1}^K p(y_i) \cdot \log_2(p(y_i))$$

$\downarrow$   
2

2 classes : 0 & 1

$$H(y) = - P(0) \log(P(0)) - P(y=1) \cdot \log(y=1)$$

$$P(0) = p$$

$$P(1) = 1-p$$

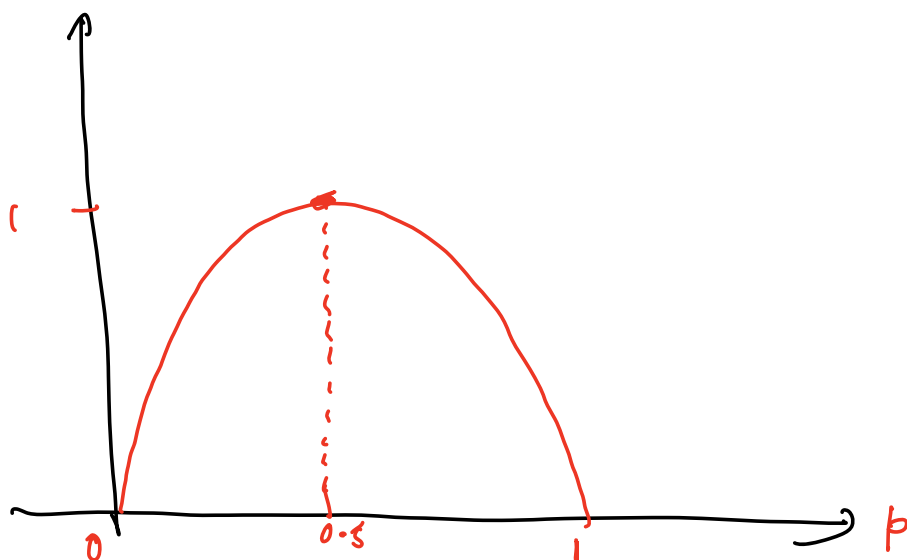
$$H(y) = -p \log p - (1-p) \log (1-p)$$

$$P(0) = \frac{1}{6}$$

$$P(1) = \frac{2}{6}$$

y
0
1
1
0
0
0

$$H(y) = -p \log p - (1-p) \log (1-p)$$



$$H(y) = -p \log p - (1-p) \log (1-p)$$

Case 1:  $p = 1$

import math

$$-(1 * \text{math} \cdot \log(1)) \rightarrow (0.000001) * \text{math} \cdot \log(0.000001) \approx 0$$

Case 2:

$$p = 0 \approx 0$$

Case 3

$$p = 0.5 \\ H \Rightarrow 1$$

$$(1-p) = 0.5$$

Case 4

$$\begin{aligned} &\rightarrow 99\% \rightarrow H(y) = 0.0801 \\ &- \rightarrow 1\% \end{aligned}$$

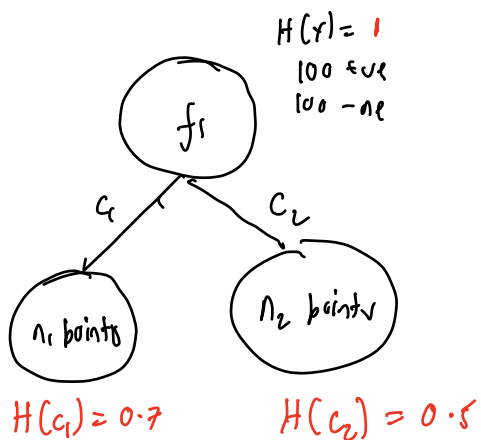
Break: 10:25

$$f_1 = c_1 / c_2$$

$$f_2 = a_1 / a_2 / a_3$$

$$f_3 = b_1 / b_2$$

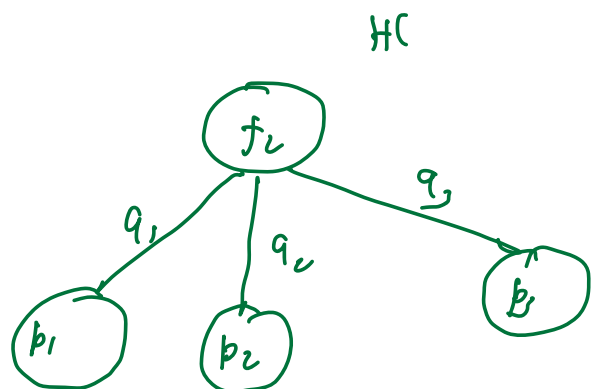
$$f_4 = d_1 / d_2$$



$$n = n_1 + n_2$$

$$\Rightarrow \frac{n_1}{n} * H(c_1) + \frac{n_2}{n} * H(c_2)$$

weighted entropy



$$H(a_1) = 0.5$$

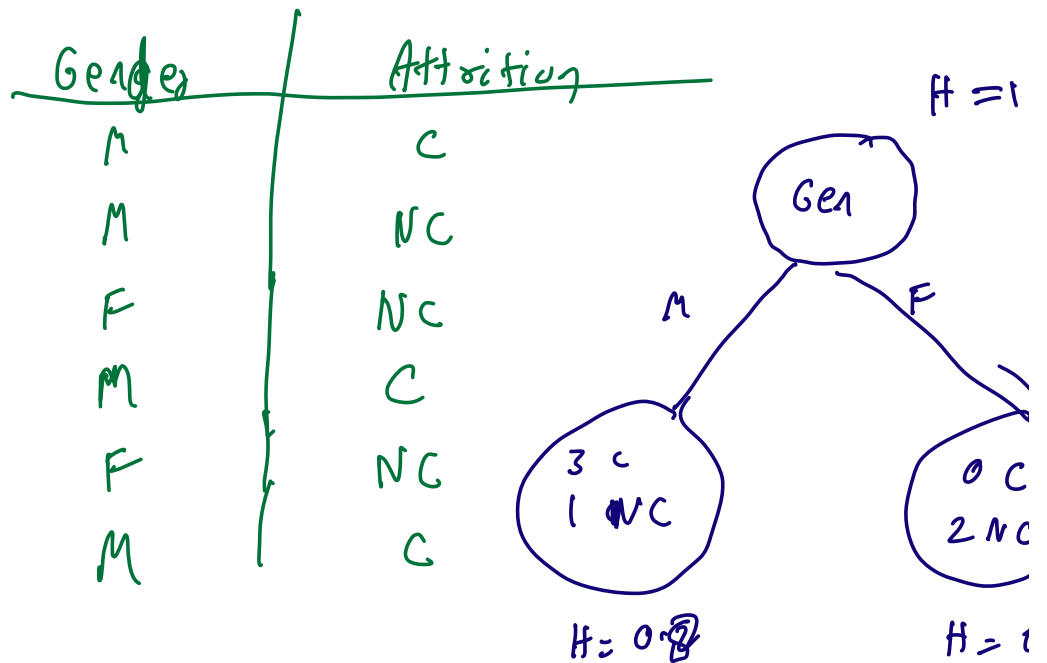
$$H(a_2) = 0.6$$

$$H(a_3) = 0.8$$

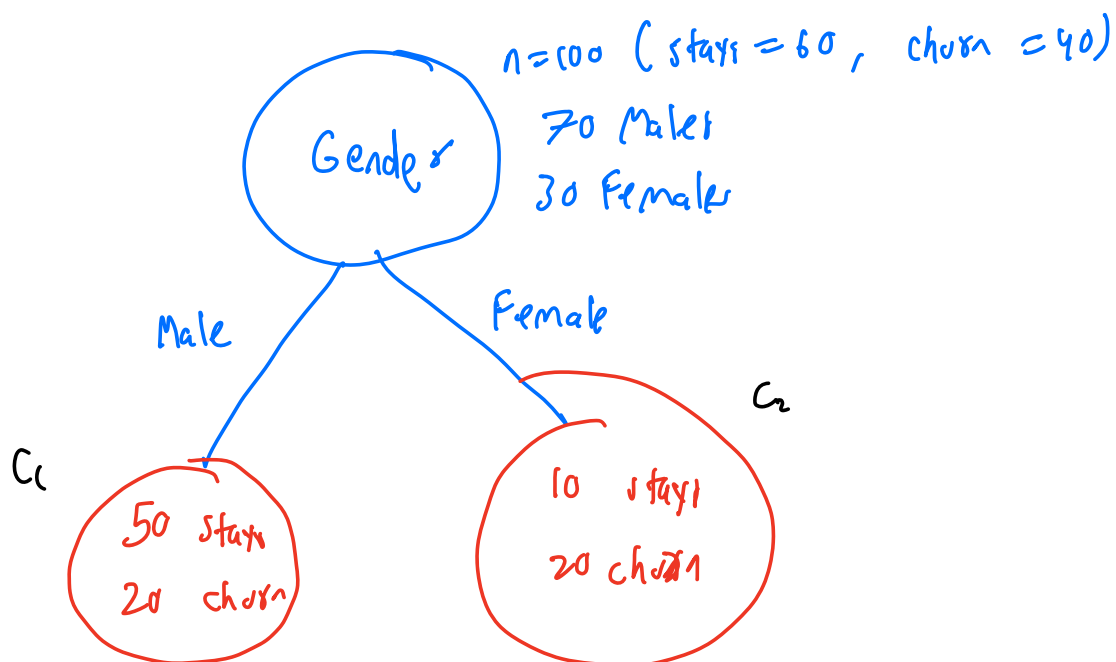
$$n = b_1 + b_2 + b_3$$

$$\frac{b_1}{n} * H(a_1) + \frac{b_2}{n} * H(a_2)$$

$$+ \frac{b_3}{n} * H(a_3)$$



Information Gain =  $H(\text{Parent}) - \text{Weighted entropy of child node}$





$$P(y=0) = \frac{60}{100} = \frac{3}{5}$$

$$P(y=1) = \frac{40}{100} = \frac{2}{5}$$

$$H(P) \Rightarrow -\frac{3}{5} \log\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) \log\left(\frac{2}{5}\right) \\ = 0.971$$

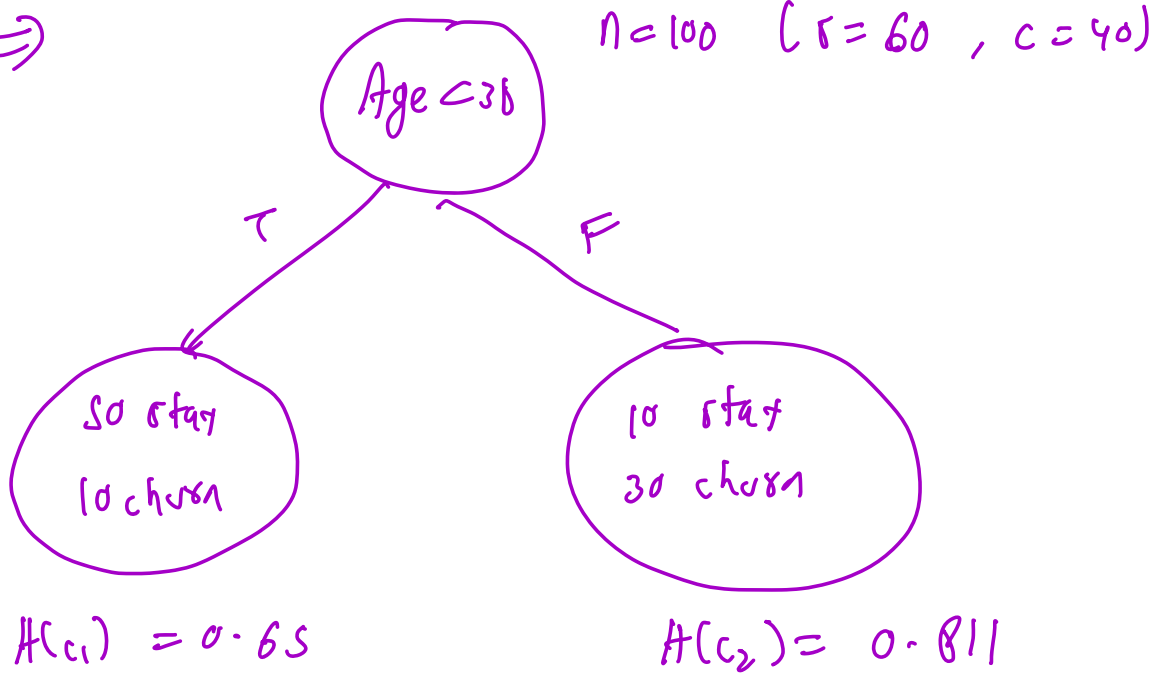
$$H(C_1) = -\frac{20}{70} \log\left(\frac{20}{70}\right) - \frac{50}{70} \log\left(\frac{10}{70}\right) \\ = 0.8631$$

$$H(C_2) = -\frac{10}{30} \log\left(\frac{10}{30}\right) - \frac{20}{30} \log\left(\frac{20}{30}\right) \\ = 0.918$$

$$W \Rightarrow \frac{70}{100} \times 0.8631 + \frac{30}{100} \times 0.918 \\ = 0.8796$$

$$IG = 0.971 - 0.8796 \\ = 0.0914$$

$\Rightarrow$



$w \in \Rightarrow 0.714$

$IG(\text{age}) \Rightarrow 0.971 - 0.714$   
 $\Rightarrow 0.257$