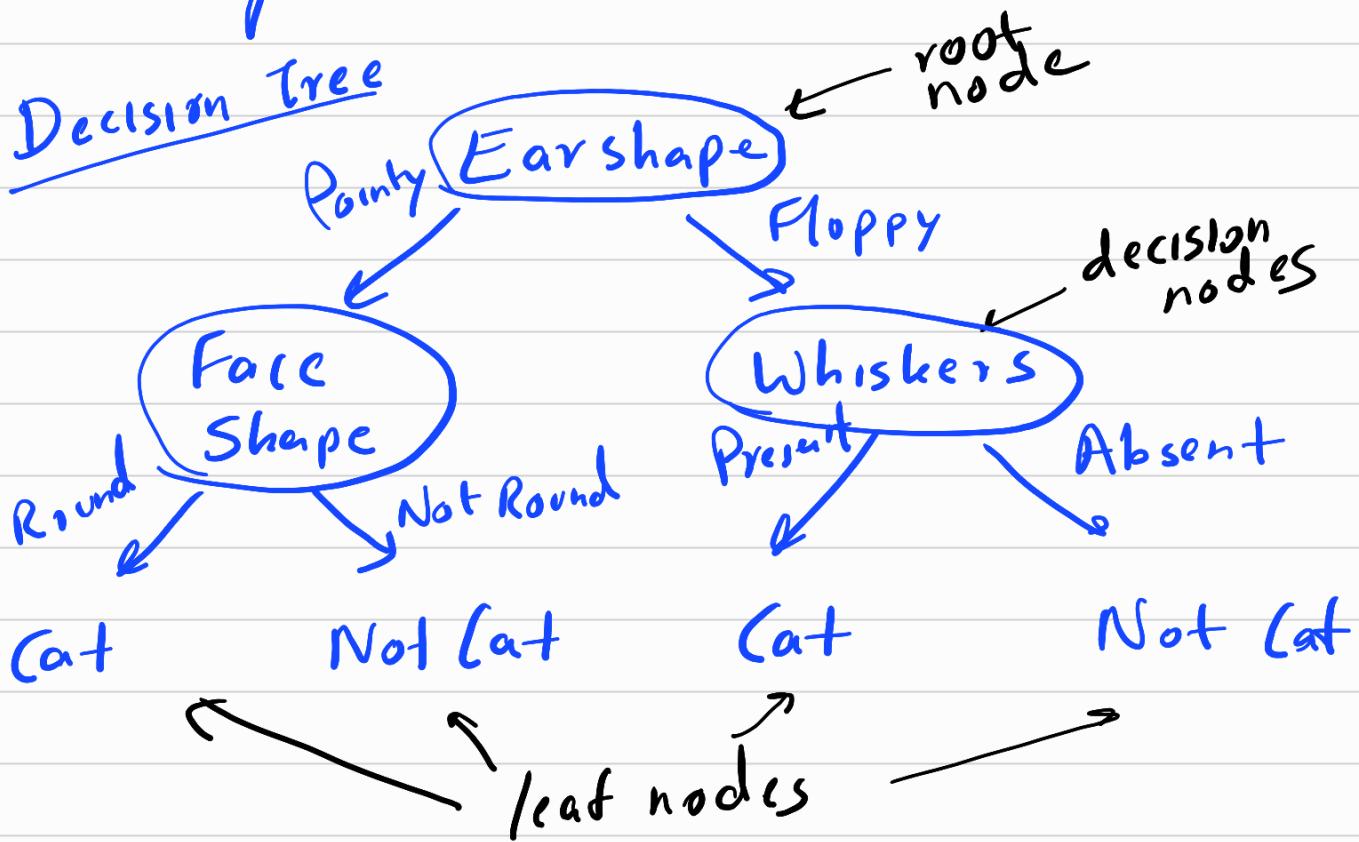


# Cat classification example;

(categorical (discrete) value)



Decisions?

① What feature to split at each node?

↳ Maximize purity

② When to stop splitting?

↳ When node 100%

- tree exceeding maximum depth
- (improvements are below threshold)

\* tree small → reduce overfitting

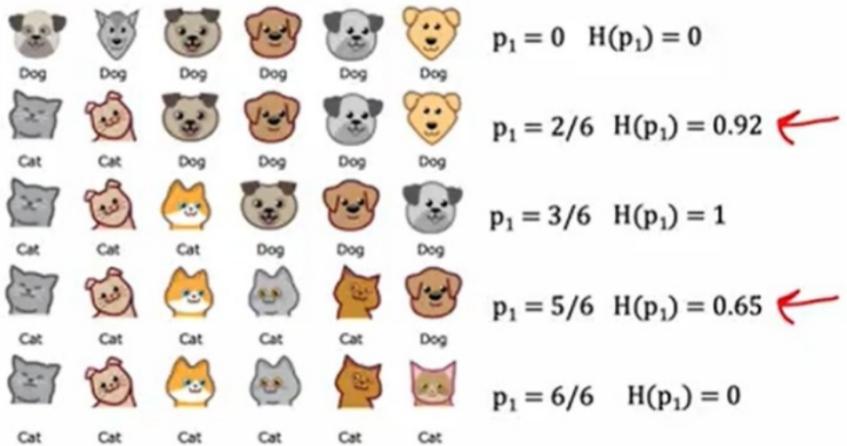
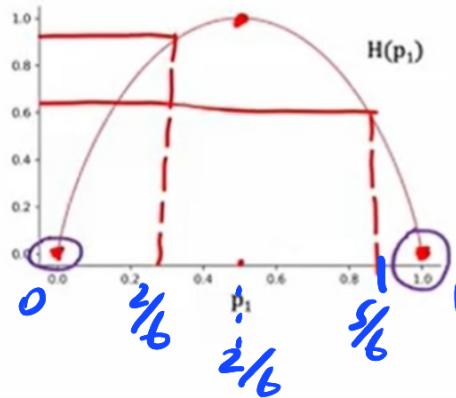
- no. of examples in node below a threshold.

# Measuring Impurity

## Entropy as a measure

### Entropy as a measure of impurity

$p_1$  = fraction of examples that are cats



$p_1$  = fraction of example that are cats

$$p_0 = 1 - p_1$$

$$H(p_1) = -p_1 \log_2(p_1) - p_0 \log_2(p_0)$$

$$= -p_1 \log_2(p_1) - (1-p_1) \log_2(1-p_1)$$

$(0 \overset{1}{\curvearrowright} 1) \rightarrow \text{genie criteria (lookalike func)}$

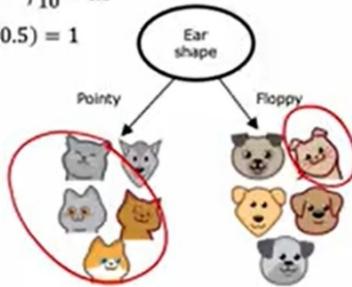
# Choosing a Split

entropy at root node

## Choosing a split

$$p_1 = 5/10 = 0.5$$

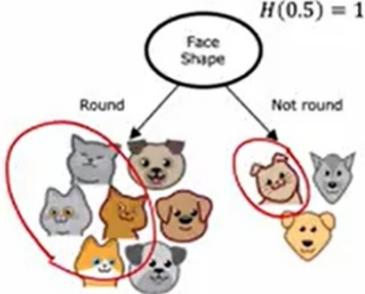
$$H(0.5) = 1$$



$$p_1 = 4/5 = 0.8 \quad p_1 = 1/5 = 0.2$$

$$H(0.8) = 0.72 \quad H(0.2) = 0.72$$

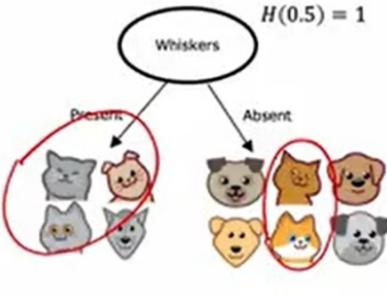
$$H(0.5) - \left( \frac{5}{10} H(0.8) + \frac{5}{10} H(0.2) \right) \\ = 0.28$$



$$p_1 = 4/7 = 0.57 \quad p_1 = 1/3 = 0.33$$

$$H(0.57) = 0.99 \quad H(0.33) = 0.92$$

$$H(0.5) - \left( \frac{7}{10} H(0.57) + \frac{3}{10} H(0.33) \right) \\ = 0.03$$



$$p_1 = 3/4 = 0.75 \quad p_1 = 2/6 = 0.33$$

$$H(0.75) = 0.81 \quad H(0.33) = 0.92$$

$$H(0.5) - \left( \frac{4}{10} H(0.75) + \frac{6}{10} H(0.33) \right) \\ = 0.12$$

Information gain

reduction of entropy  
by doing a split

$$p_1^{\text{left}} = \frac{4}{5}$$

$$p_1^{\text{right}} = \frac{1}{5}$$

$$w^{\text{left}} = \frac{5}{10}$$

$$w^{\text{right}} = \frac{5}{10}$$

$$p_1^{\text{root}} = \frac{5}{10}$$

Information Gain :

$$H(p_1^{\text{root}}) - \left( w^{\text{left}} H(p_1^{\text{left}}) + w^{\text{right}} H(p_1^{\text{right}}) \right)$$

# Decision Tree Learning ;

info gain pick highest feature

create left - right branches

split process until stop criterion

→ Recursive algorithm

## One-hot encoding

Earshape	pointy	flop.	oval
pointy	1	0	0
oval	0	0	1
floppy	0	1	0

→

binary

Hot feature takes  
value = 1

# Continuous features

Weight

7 - 2

8 - 8

15

9 - 2

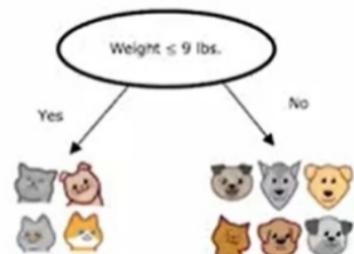
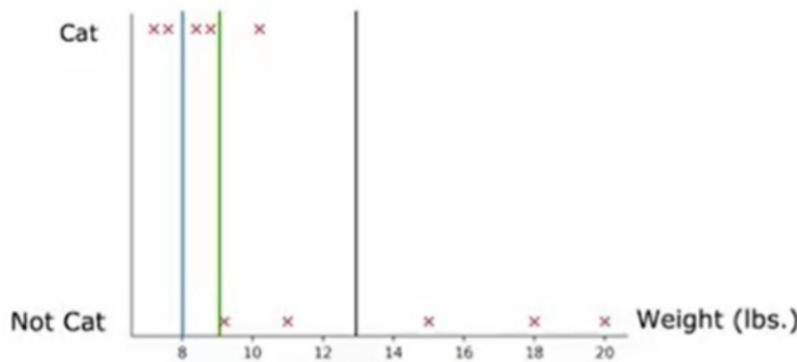
eg:-

Weight  $\leq$  8 lbs

splitting on a  
cts variable ;

;

## Splitting on a continuous variable



$$H(0.5) - \left( \frac{2}{10} H\left(\frac{2}{2}\right) + \frac{8}{10} H\left(\frac{3}{8}\right) \right) = 0.24$$

$$H(0.5) - \left( \frac{4}{10} H\left(\frac{4}{4}\right) + \frac{6}{10} H\left(\frac{1}{6}\right) \right) = 0.61$$

$$H(0.5) - \left( \frac{7}{10} H\left(\frac{5}{7}\right) + \frac{3}{10} H\left(\frac{0}{3}\right) \right) = 0.40$$

# Regression Trees ;

## Choosing a split

Variance at root node: 20.51



Weights: 7.2, 9.2, 8.4, 7.6, 10.2, 8.8, 15, 11, 18, 20

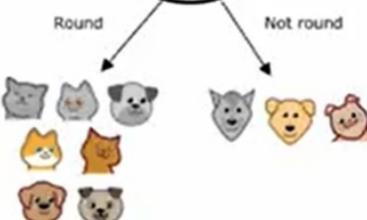
Variance: 1.47

$$w_{\text{left}} = \frac{5}{10} \quad w_{\text{right}} = \frac{5}{10}$$

$$20.51 - \left( \frac{5}{10} \cdot 1.47 + \frac{5}{10} \cdot 21.87 \right)$$

$$= 8.84$$

Variance at root node: 20.51



Weights: 7.2, 15, 8.4, 7.6, 10.2, 18, 20

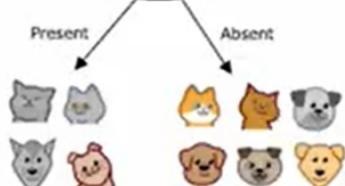
Variance: 27.80

$$w_{\text{left}} = \frac{7}{10} \quad w_{\text{right}} = \frac{3}{10}$$

$$20.51 - \left( \frac{7}{10} \cdot 27.80 + \frac{3}{10} \cdot 1.37 \right)$$

$$= 0.64$$

Variance at root node: 20.51



Weights: 7.2, 8.8, 9.2, 8.4

Variance: 0.75

$$w_{\text{left}} = \frac{4}{10} \quad w_{\text{right}} = \frac{6}{10}$$

$$20.51 - \left( \frac{4}{10} \cdot 0.75 + \frac{6}{10} \cdot 23.32 \right)$$

$$= 6.22$$

variance reduction

