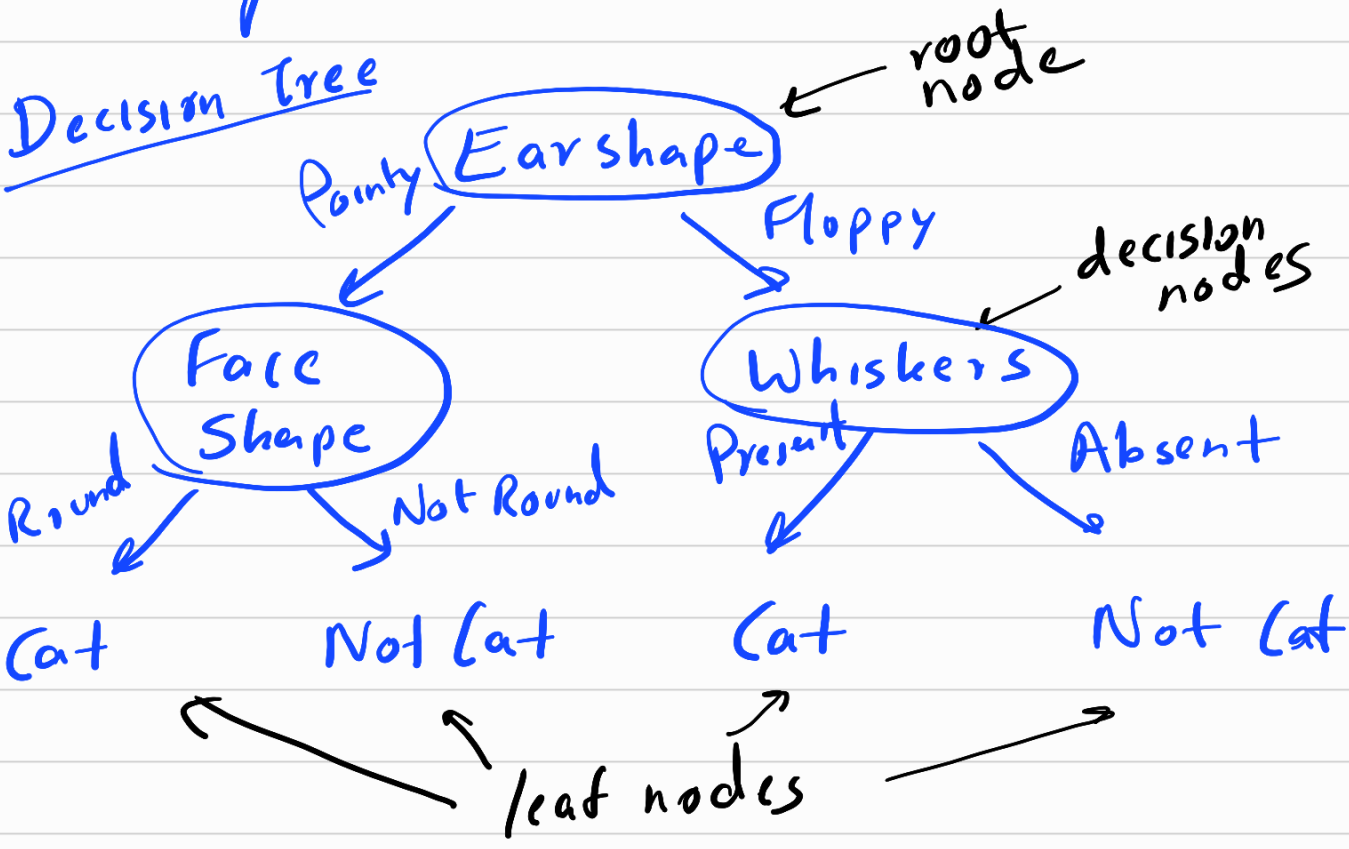
Cat classification example;

Categorical (discrete) value



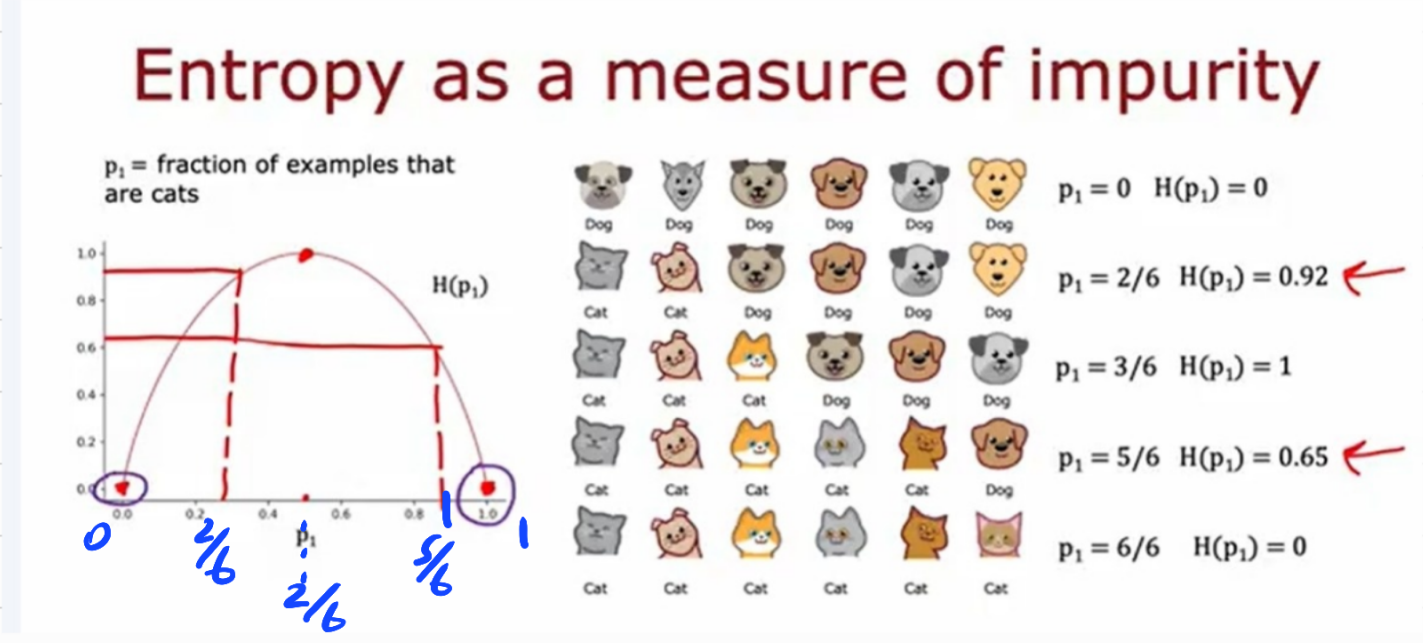
Decisions?

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 threshhold)

* tree small ⟶ reduce overfitting
  - no. of examples in node below a threshhold.

Measuring Impurity

Entropy as a measure



p_1 = fraction of example that are cats

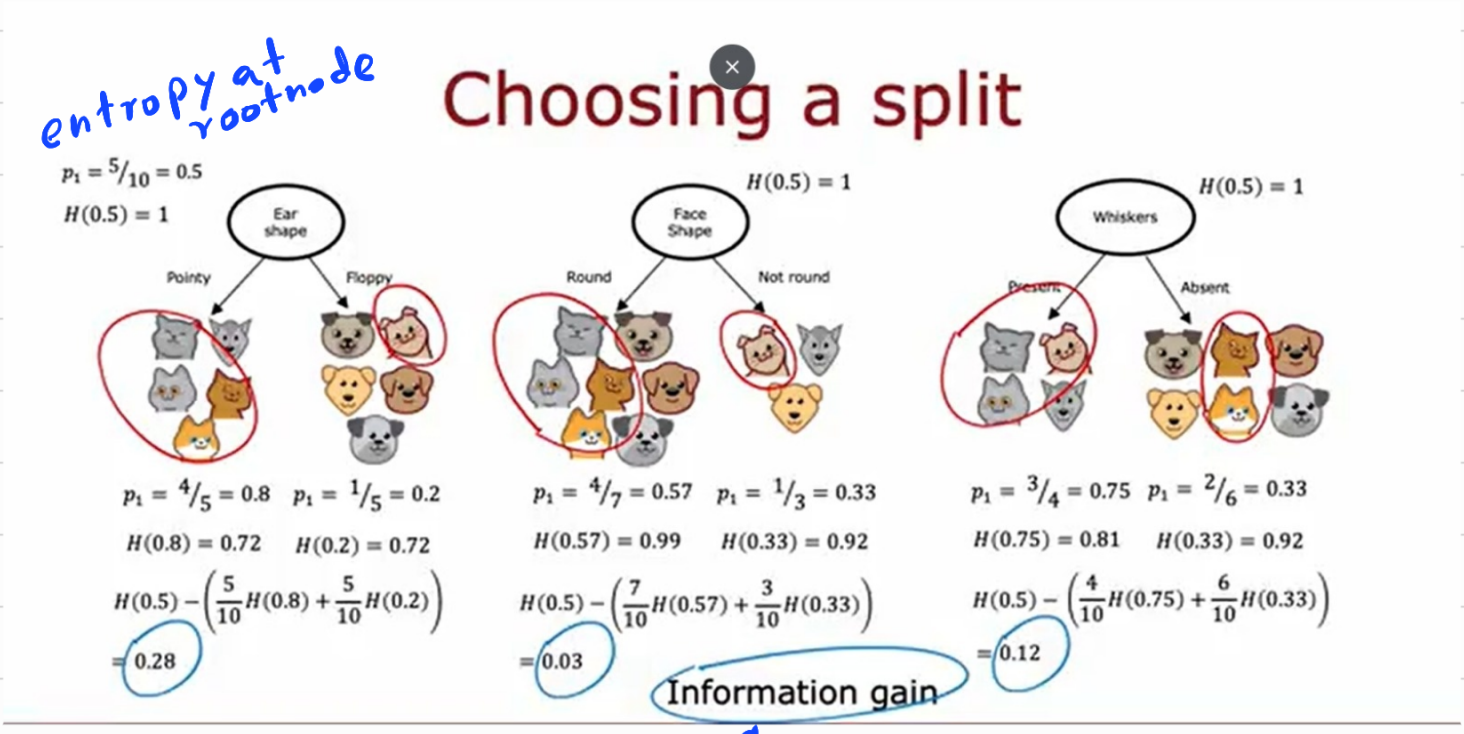
$p_0 = 1 - p_1$

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

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

→ genie criteria lookalike func

Choosing a Split



reduction of entropy by doing a split

$p_1^{left} = 4/5$ $\quad p_1^{right} = 1/5$

$w^{left} = 5/10$ $\quad w^{right} = 5/10$

$p_1^{root} = 5/10$

Information Gain:

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

Decision Tree Learning;

info gain pick highest feature

create left - right branches

split process until stop criteria

↳ 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

Continuos features

Weight

7.2

8.8

15

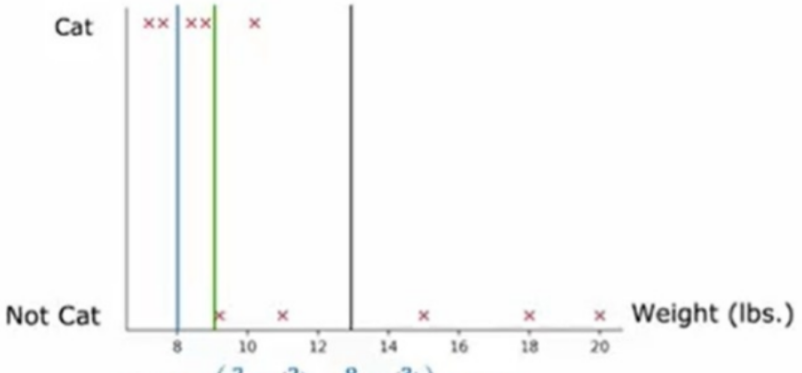
9.2

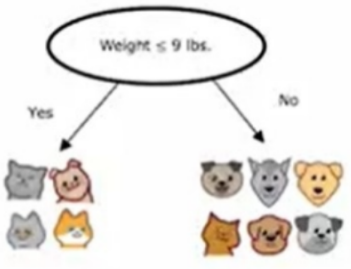
!

eg:-
Weight ≤ 8 lbs

splitting on a cts variable;

Splitting on a continuous variable





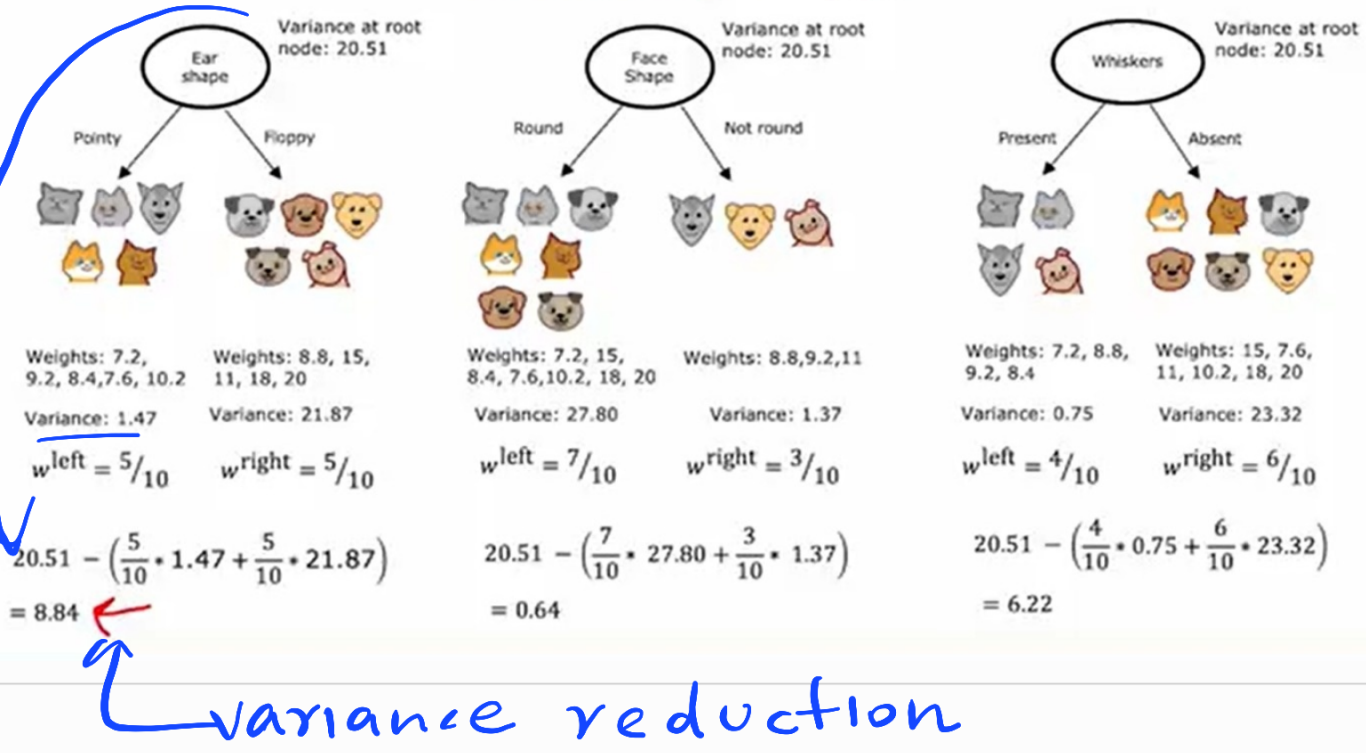
$$H(0.5) - \left(\frac{2}{10}H\left(\frac{2}{2}\right) + \frac{8}{10}H\left(\frac{3}{8}\right)\right) = 0.24$$

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

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

Regression Trees;

Choosing a split



variance reduction