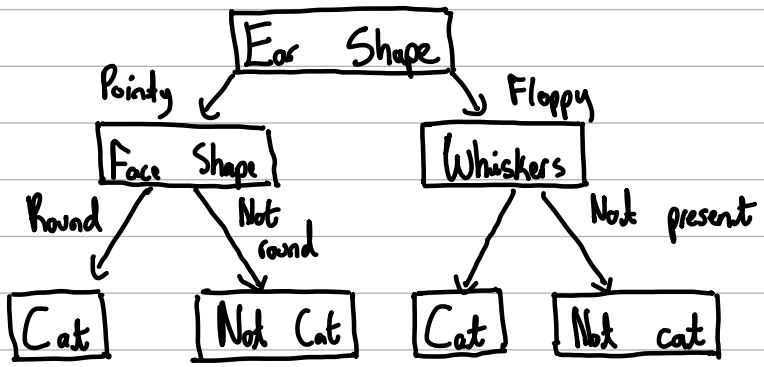
Reference Example: Cat Classification

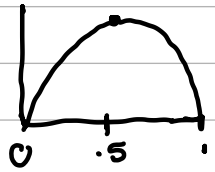
Decision Tree



Decision Tree learning

1. How to choose on what feature to split
2. When To Stop Splitting?

Entropy - measure of how "impure" dataset is (how one-sided it is)



50/50 = most impure → At 0 & 1, H = 0

Choosing Split → H(root node) = .5

Ear Shape

Pointy	Floppy
4 cats	1 cat
1 dog	4 dogs
H(4/5) = .72	H(.2) = .72

.5 - weighted avg = .28

Face Shape

Round	Not round
4 cats	1 cat
3 dogs	2 dogs
H(4/7) = .99	H(1/3) = .92

.5 - weighted avg = .03

Whiskers

3 cats	2 cats
1 dog	4 dogs
H(.75) = .81	H(.33) = .92

.5 - weighted avg = 0.12

Ear Shape has lowest reduction / highest info gain

Information Gain

$H(S)$ = starting entropy

H_{after} = How much entropy remains after split

$H(S) - H_{after}$ = entropy eliminated by splitting on A

Splitting on Continuous Variables (i.e. weight)

- Try multiple different thresholds
- Choose threshold with highest IG

Tree Ensemble

- Using multiple decision trees to make a vote on what a feature is
  - ↓ How to make each tree different?

Sampling w/ Replacement

- Keep picking random values from dataset, but don't replace them
- This will cause duplicates, and some features may not even appear
- Each tree will have bias towards a particular example's features
- Reduces OVERFITTING