Advanced Learning Algorithms - Week 4

Decision Tree Model

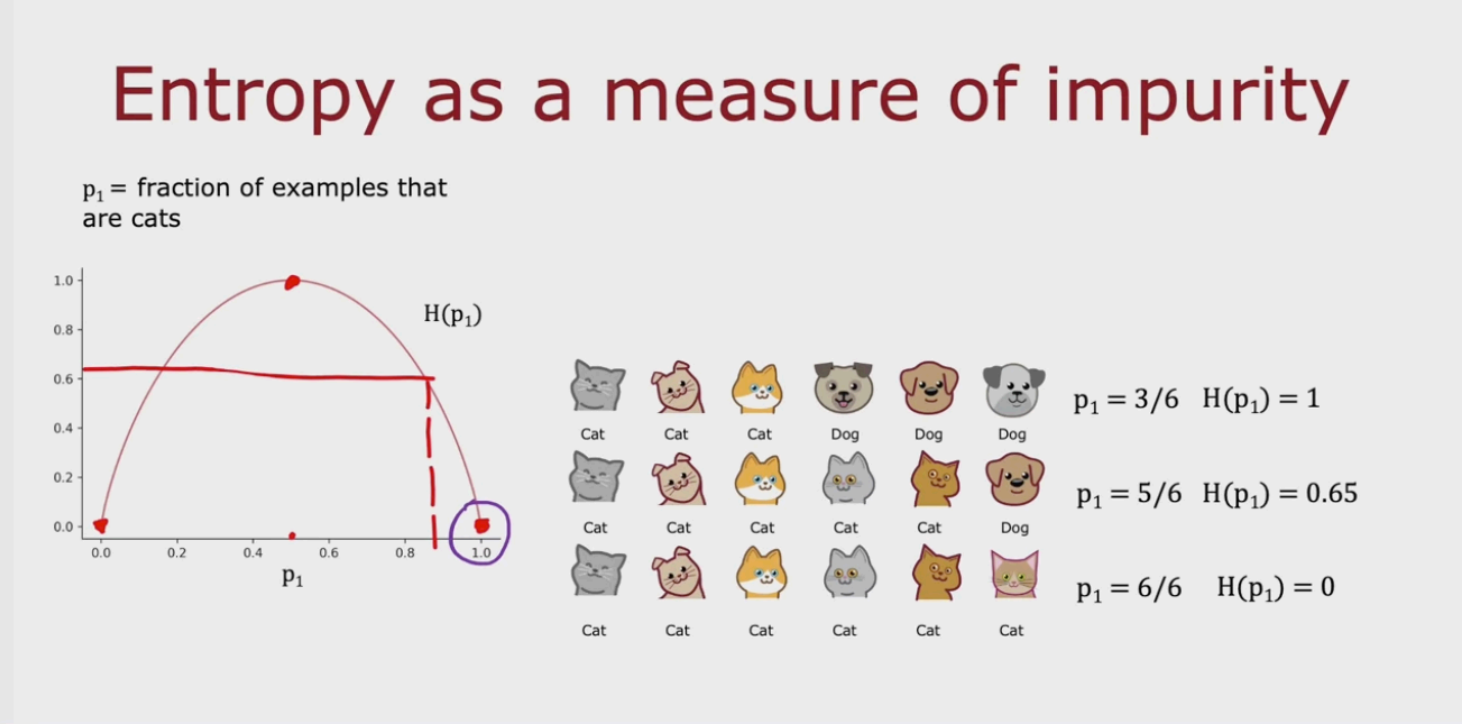
Every 'object' have some values. Face: Round, Square - Ear: Pointy, Floppy...
Algorithm iterates over the tree to classify an object.

Pick One Tree That Fits The Best

How to build:

Features : Aim is to maximize purity.
When to Stop Splitting? : Max depth of the tree (overfitting), When a node becomes %100 one class, threshold of minimum # of examples in node.

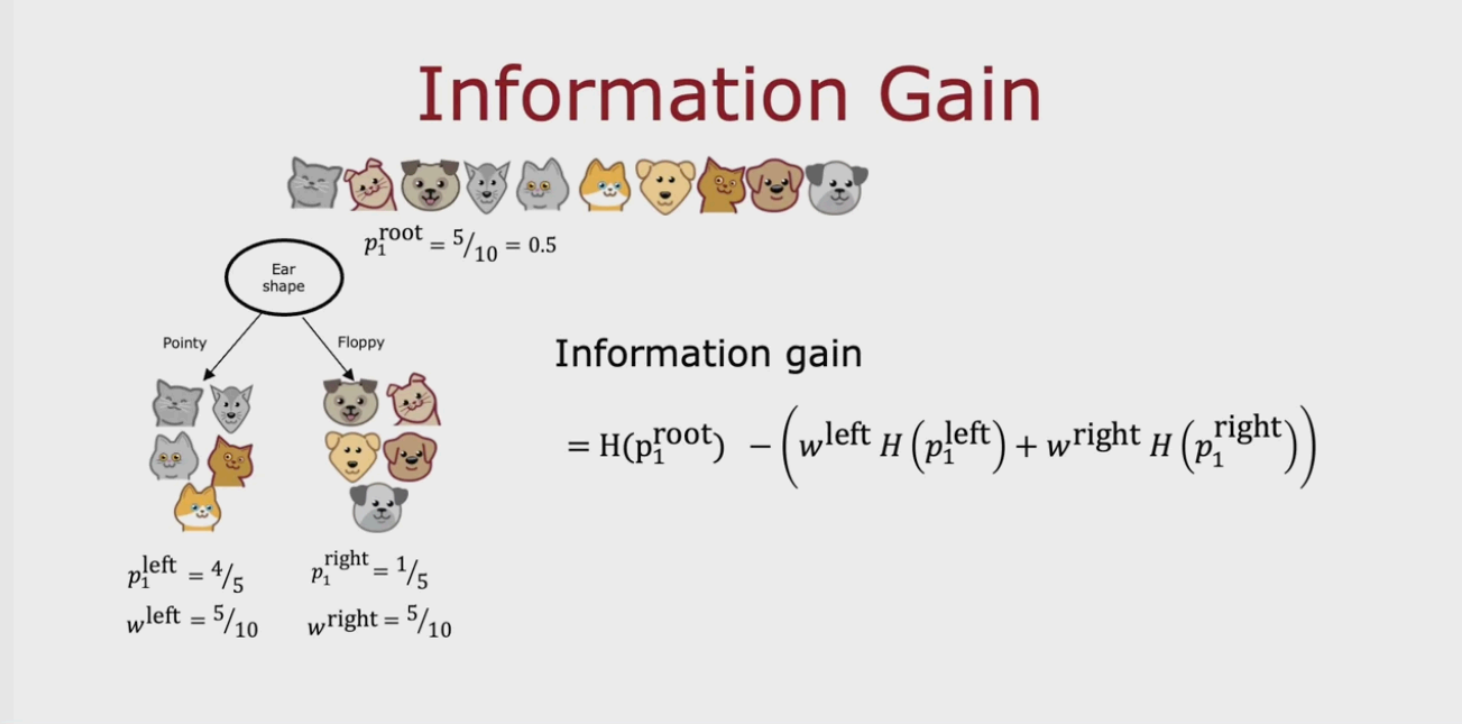
Entropy as a measure of impurity



Entropy Function

Reduction of Entropy / Increase in Purity : **Choosing a split**
Reduction in Entropy = Information Gain

Should bother to increase the size of the tree?



- At the first step: Everything is at root.
- Check the highest information gain among the possible splitting options (features).
- Repeat - until hitting one of the thresholds.

One Hot Encoding

-> One feature that can take up 3 values - Ear: Pointy/Floppy/Round

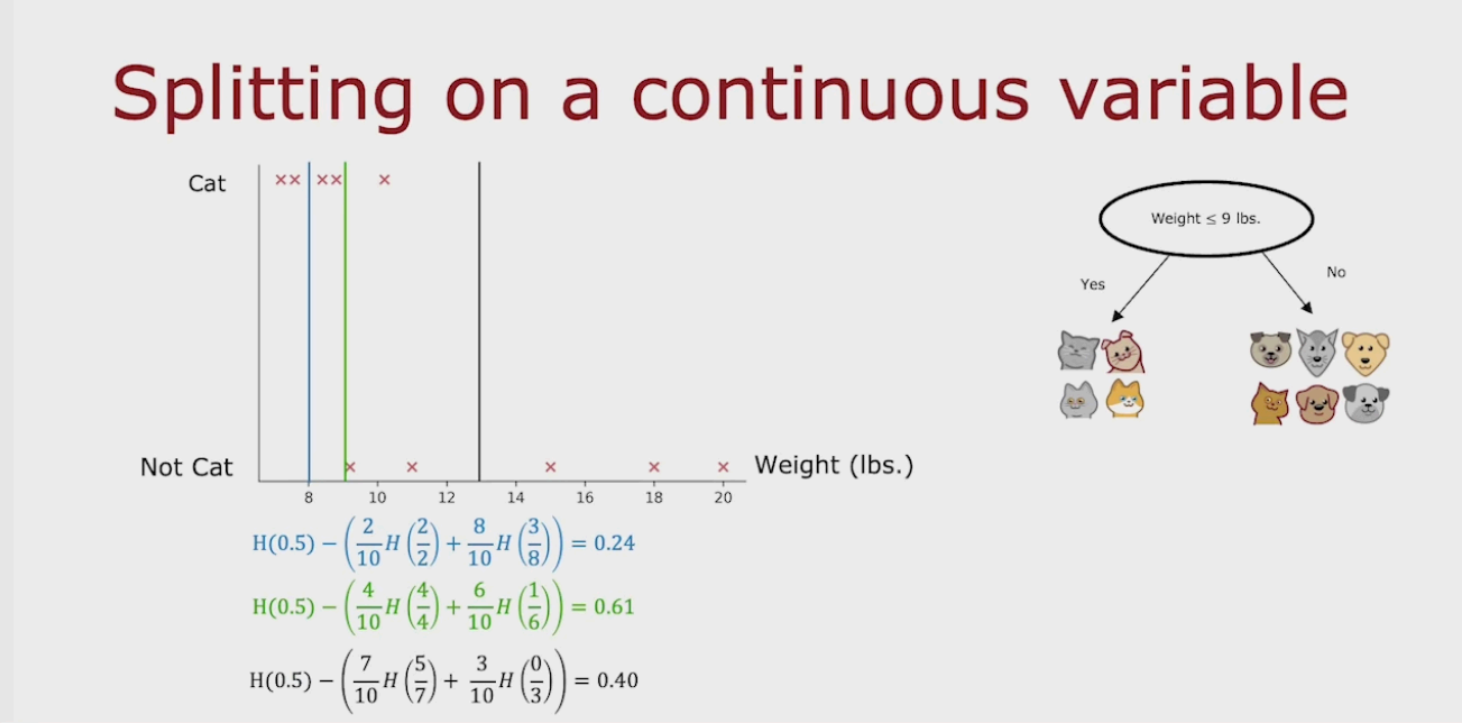
-> Three feature that can take up only 2 values (0 or 1) - Pointy Ear: 0/1

Also works with regression/neural

Continuous Features

-> Weight (5.0 - 10.0)

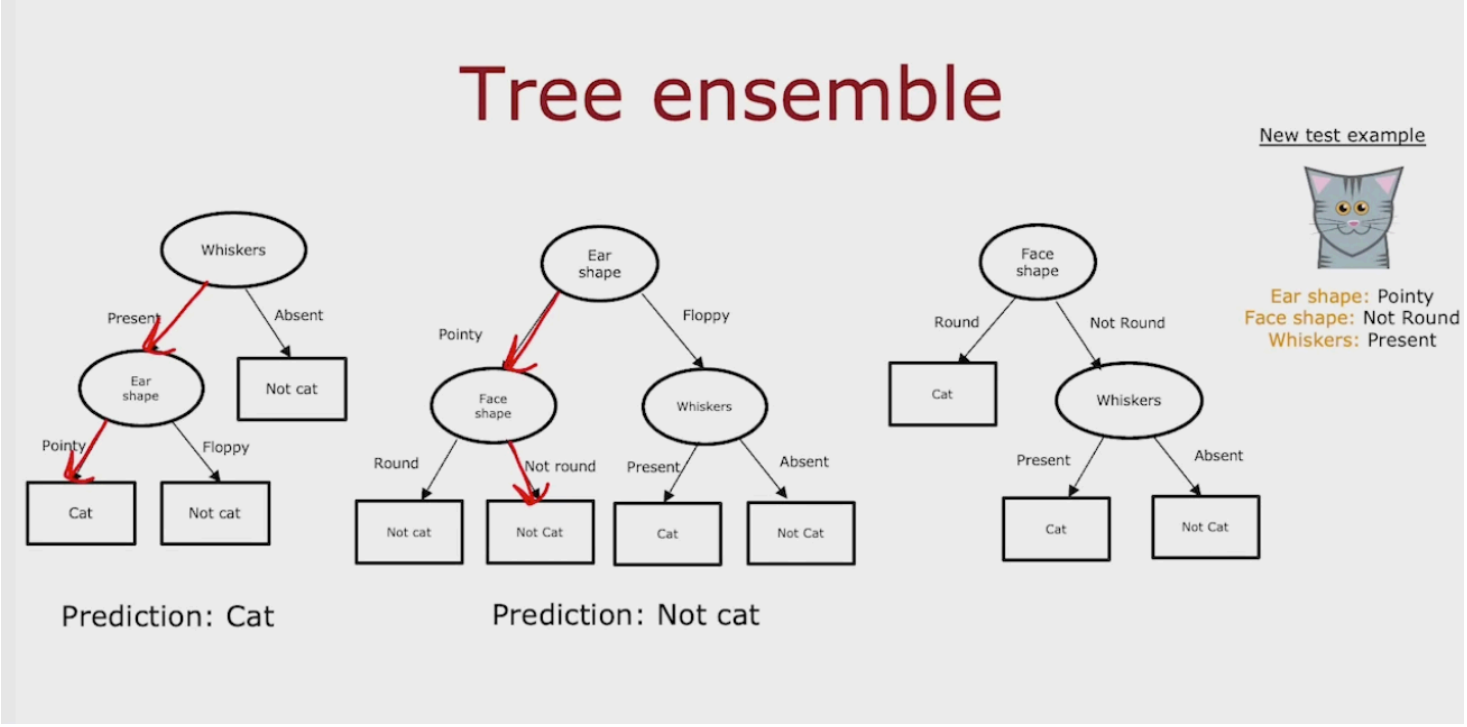
Pick a value for threshold. Split the examples. (Weight < 8 lbs)



One Decision Tree : Highly sensitive to small changes to the data.

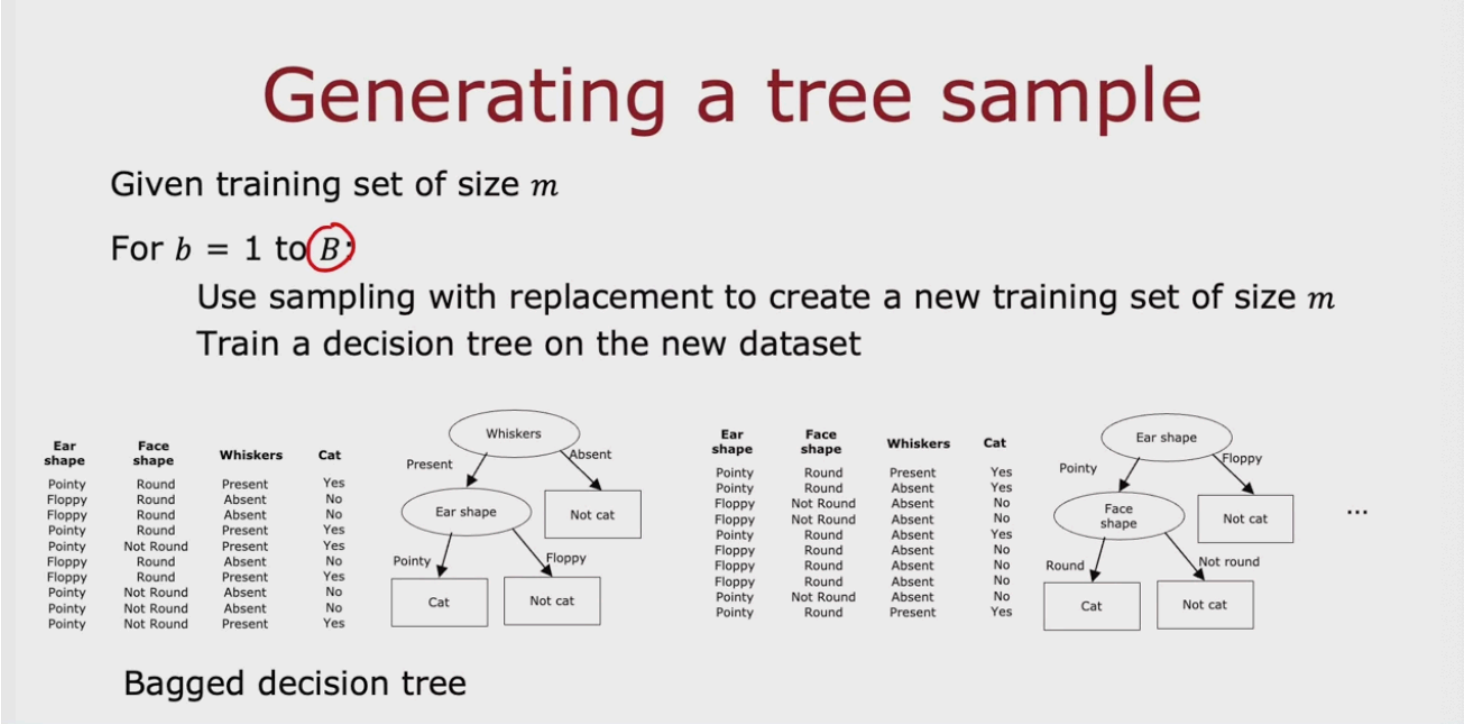
Better Option: Train whole bunch of different decision trees. - **Tree Ensemble**

Make them vote for the output.



Sampling with replacement -> replacing makes it possible to get the same output again.
Tokens.

Random Forest Algorithm - Tree Ensembles

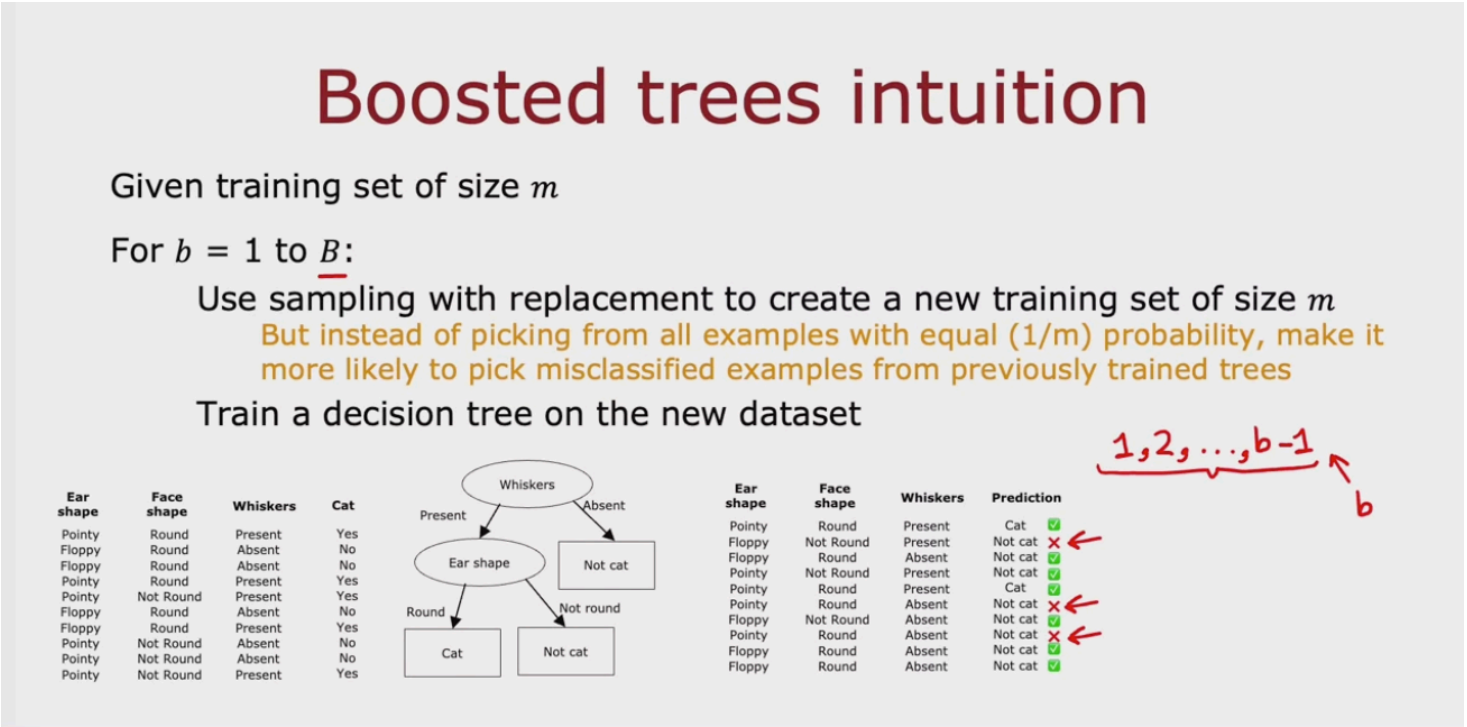


after some # of decision trees, computation is high, more trees return into diminishing returns.

Randomizing the feature choice.
There are n features available at some split, create a subset from the *n*, say *k*
make the algorithm choose from that subset of *k*
k = root of *n*

Boosted Decision Trees - (XGB eXtreme Gradient Boosting)

Work more on the parts which you perform (tree performs) not well - idea of boosting.



Ear shape	Face shape	Whiskers	Cat
Pointy	Round	Present	Yes
Floppy	Round	Absent	No
Floppy	Round	Absent	No
Pointy	Round	Present	Yes
Pointy	Not Round	Present	Yes
Floppy	Round	Absent	No
Floppy	Round	Present	Yes
Pointy	Not Round	Absent	No
Pointy	Not Round	Absent	No
Pointy	Not Round	Present	Yes

Ear shape	Face shape	Whiskers	Prediction
Pointy	Round	Present	Cat ✓
Floppy	Not Round	Present	Not cat ✗ ←
Floppy	Round	Absent	Not cat ✓
Pointy	Not Round	Present	Not cat ✓
Pointy	Round	Present	Cat ✓
Pointy	Round	Absent	Not cat ✗ ←
Floppy	Not Round	Absent	Not cat ✓
Pointy	Round	Absent	Not cat ✗ ←
Floppy	Round	Absent	Not cat ✓
Floppy	Round	Absent	Not cat ✓

Can use XGB both for classification and regression.

Conclusions

Decision trees and tree ensembles work well on **Tabular (Structured) Data**
Performs poorly for **unstructured data (images, audio, text)**

Neural Networks

Works well on **all types of data, including structured and unstructured data.**
May be slower.
Works with transfer learning, eliminate pre-training.
System of multiple models working together is better.