













Decision Tree

Decision tree model

Example: Cat Classification

- A cat adoption center uses features like ear shape, face shape, and whiskers to classify animals as cats or not.
- The dataset consists of 10 examples, with features taking on categorical values, making it a binary classification task.
 - The feature X takes on **categorical (discrete values)**

	Ear shape	Face shape	Whiskers	Cat
	Pointy	Round	Present	1
	Floppy	Not round	Present	1
	Floppy	Round	Absent	0
	Pointy	Not round	Present	0
	Pointy	Round	Present	1
	Pointy	Round	Absent	1
	Floppy	Not round	Absent	0
	Pointy	Round	Absent	1
	Floppy	Round	Absent	0
	Floppy	Round	Absent	0

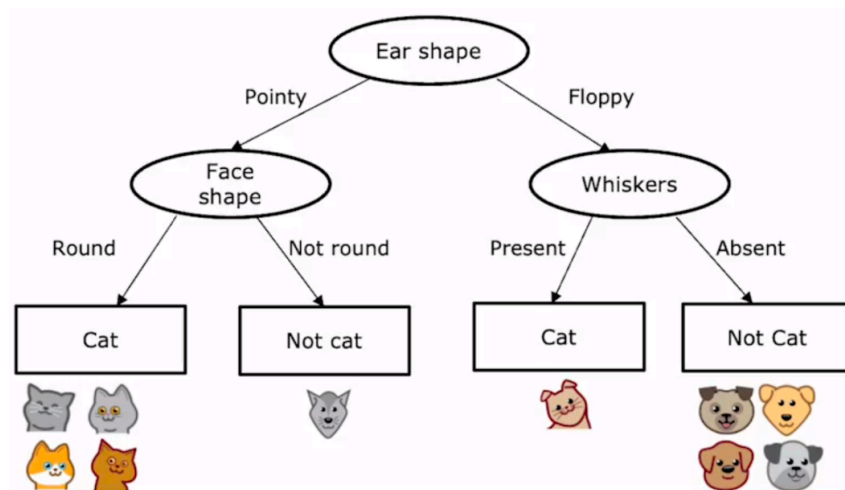
Structure of Decision Trees

- The top node is called the **root node**, and nodes that make decisions (**internal node**) are called **decision nodes** (including *root node*).
- **Leaf nodes** at the bottom make predictions, and the structure resembles a tree, with roots at the top and leaves at the bottom.
 - Leaf node: *Cat/Not Cat*

Learning Process

Decision Tree Construction Steps

- **Choosing the Root Feature:** The first step is selecting a feature for the root node, such as ear shape, to split the training examples into subsets.
- **Branching Decisions:** After the initial split, further features are chosen for subsequent nodes, like face shape, to continue dividing the examples into more specific groups.



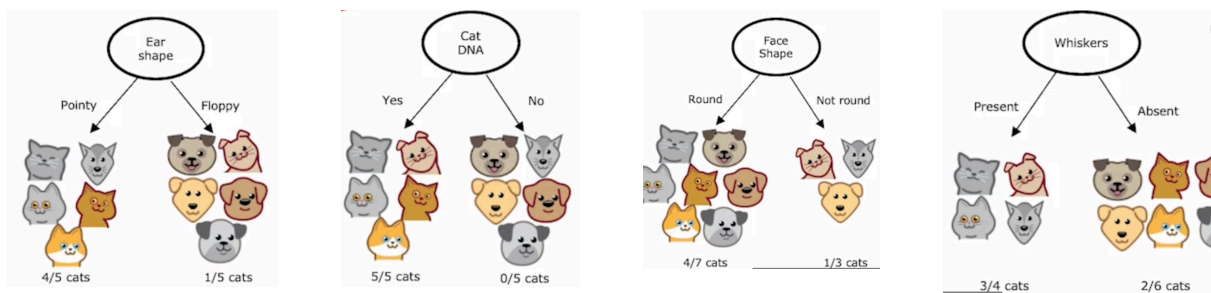
Key Decisions in Building a Decision Tree



Decision 1: *How to choose what feature to split on at each node?*

Maximize purity (or minimize impurity)

- **Feature Selection for Splitting:** The algorithm must determine which feature to use at each node to maximize the purity of the resulting subsets (i.e., groups that are as homogeneous as possible).



Decision 2: When do you stop splitting?

- When a node is 100% one class
 - When splitting a node will result in the tree exceeding a maximum depth.
 - When improvements in purity score are below a threshold.
-
- **Stopping Criteria for Splitting:** Decisions must be made on when to stop splitting nodes, which can be based on achieving pure subsets, reaching a maximum tree depth, or having too few examples to split further.