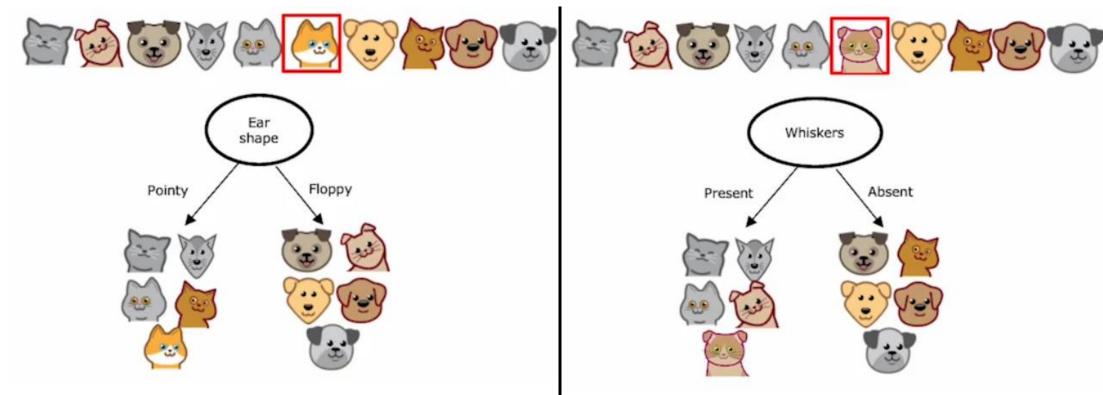
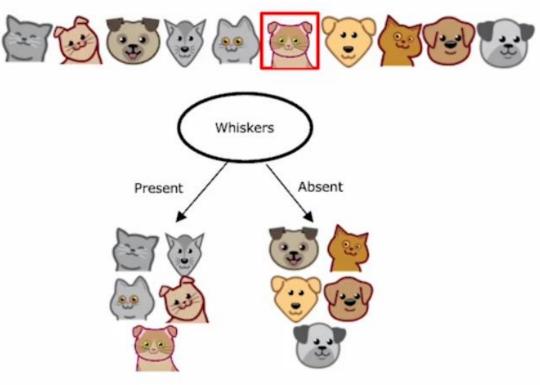
DECISION TREE LEARNING



Sensitivity of Decision Tree on small change in data

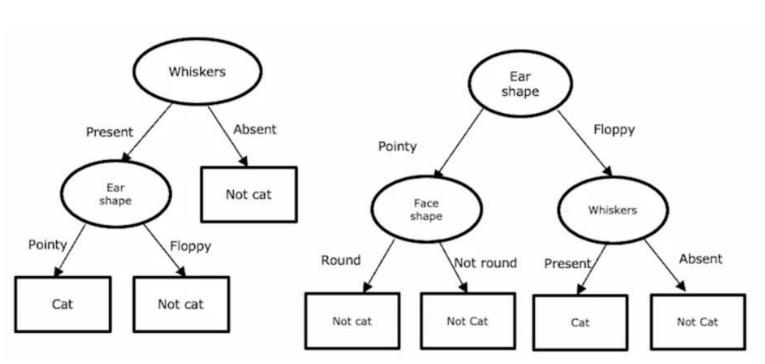




00

Face shape: Not Round

Whiskers: Present



Round Not Round

Cat Whiskers

Present Absent

Cat Not Cat

Prediction: Cat

O Tree Ensemble

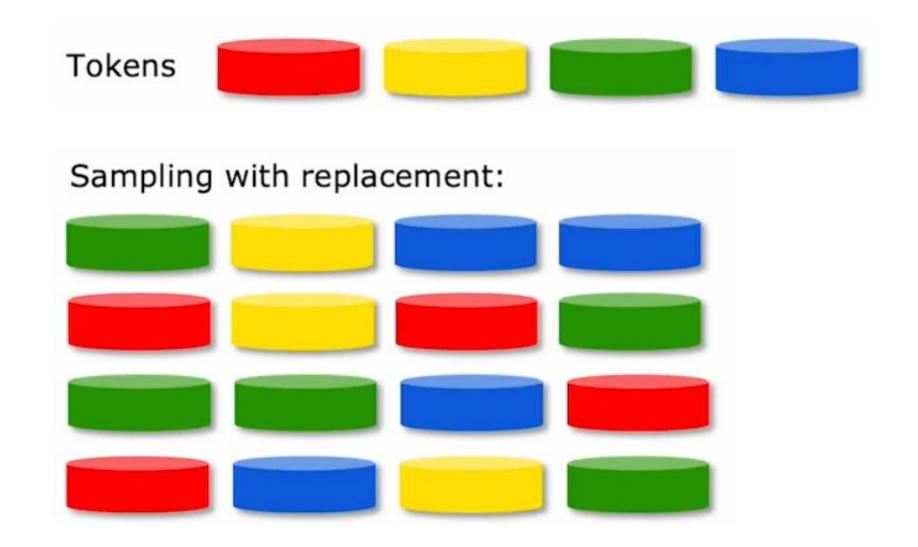
Prediction: Not cat

Prediction: Cat

Final prediction: Cat



Sampling with Replacement



O Sampling with Replacement



	Ear shape	Face shape	Whiskers	l Cat
3	Pointy	Round	Present	1
	Floppy	Not round	Absent	0
()	Pointy	Round	Absent	1
	Pointy	Not round	Present	0
	Floppy	Not round	Absent	0
(w)	Pointy	Round	Absent	1
(==7	Pointy	Round	Present	1
	Floppy	Not round	Present	1
•	Floppy	Round	Absent	0
(=)	Pointy	Round	Absent	T 1



Generating a tree sample (Bagged decision tree)

Given training set of size m

For b = 1 to B:

Use sampling with replacement to create a new training set of size m. Train a decision tree on the new dataset

Ear shape	Face shape	Whiskers	Cat	Present	Whiskers
Pointy	Round	Present	Yes	*	*
Floppy	Round	Absent	No		
Floppy	Round	Absent	No	(Ear shap	e) Not cat
Pointy	Round	Present	Yes		\prec
Pointy	Not Round	Present	Yes		
Floppy	Round	Absent	No	Pointy 1	Floppy
Floppy	Round	Present	Yes		
Pointy	Not Round	Absent	No		N-tt
Pointy	Not Round	Absent	No	Cat	Not cat
Pointy	Not Round	Present	Yes		



Generating a tree sample (Bagged decision tree)

Given training set of size m

For b = 1 to B:

Use sampling with replacement to create a new training set of size m. Train a decision tree on the new dataset

Ear shape	Face shape	Whiskers	Cat		Ear shape Floppy	
Pointy	Round	Present	Yes	Pointy /	(200)	
Pointy	Round	Absent	Yes	>	*	
Floppy	Not Round	Absent	No	Face		200
Floppy	Not Round	Absent	No	shape) Not cat	•••
Pointy	Round	Absent	Yes	Snape	\prec	
Floppy	Round	Absent	No	/		
Floppy	Round	Absent	No	Round J	Not round	
Floppy	Round	Absent	No		, _	
Pointy	Not Round	Absent	No			
Pointy	Round	Present	Yes	Cat	Not cat	

O Randomizing the feature choice

At each node, when choosing a feature to use to split, if n
features are available, pick a random subset of k < n features
and allow the algorithm to only choose from that subset of
features

Random forest algorithm

•
$$K = Sqrt(n)$$



Boosted trees intuition

Given training set of size *m*

For b = 1 to B:

Use sampling with replacement to create a new training set of size m

But instead of picking from all examples with equal (1/m) probability, make it more likely to pick misclassified examples from previously trained trees

Ear shape	Face shape	Whiskers	Cat	Whiskers		Ear shape	Face shape	Whiskers	Prediction	
		Dragant	Voc	Present	_	Pointy	Round	Present	Cat	$\overline{\mathbf{v}}$
Pointy	Round	Present	Yes			Floppy	Not Round	Present	Not cat	×
Floppy	Round	Absent	No		\	Floppy	Round	Absent	Not cat	
Floppy	Round	Absent	No	(Ear shape	Not cat	Pointy	Not Round	Present	Not cat	
Pointy	Round	Present	Yes							~
Pointy	Not Round	Present	Yes	/		Pointy	Round	Present	Cat	V
Floppy	Round	Absent	No	Round J	Not round	Pointy	Round	Absent	Not cat	
	Round	Present	Yes			Floppy	Not Round	Absent	Not cat	$\overline{\mathbf{v}}$
Floppy						Pointy	Round	Absent	Not cat	×
Pointy	Not Round	Absent	No	Cat	Not cat	Floppy	Round	Absent	Not cat	∇
Pointy	Not Round	Absent	No	Cat		Floppy	Round	Absent	Not cat	
Pointy	Not Round	Present	Yes			Порру	Round	Absent	not cut	•

XGBoost (eXtreme Gradient Boosting)

- Open-source implementation of boosted trees
- Fast efficient implementation
- Good choice of default splitting criteria and criteria for when to stop splitting
- Built in regularization to prevent overfitting
- Highly competitive algorithm for machine learning competitions (e.g. Kaggle competitions)

