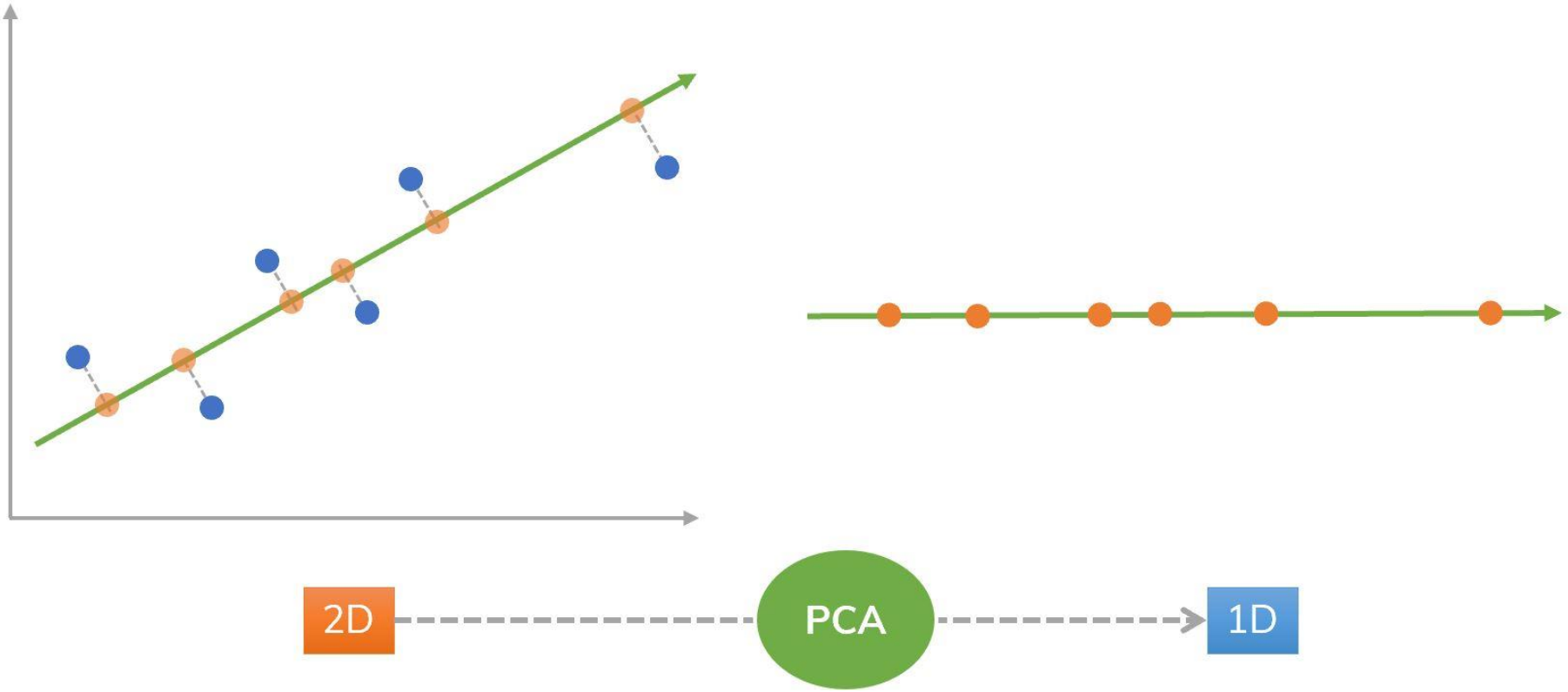
Day 05

Data Science with Python

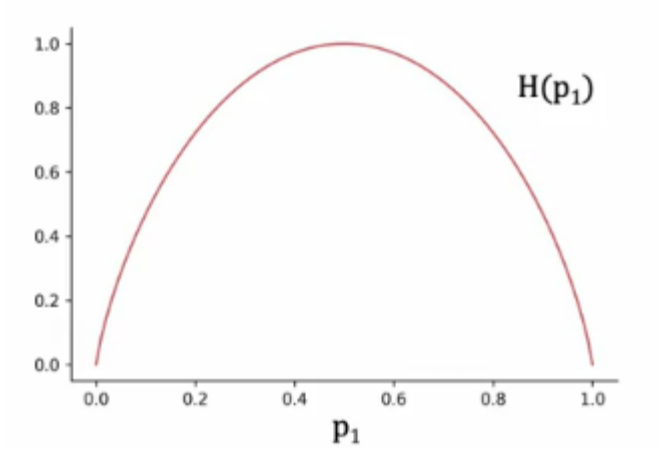
Pasindu Marasinghe
ppm@ucsc.cmb.ac.lk

Dimensionality Reduction


2D
PCA
1D

Decision Trees

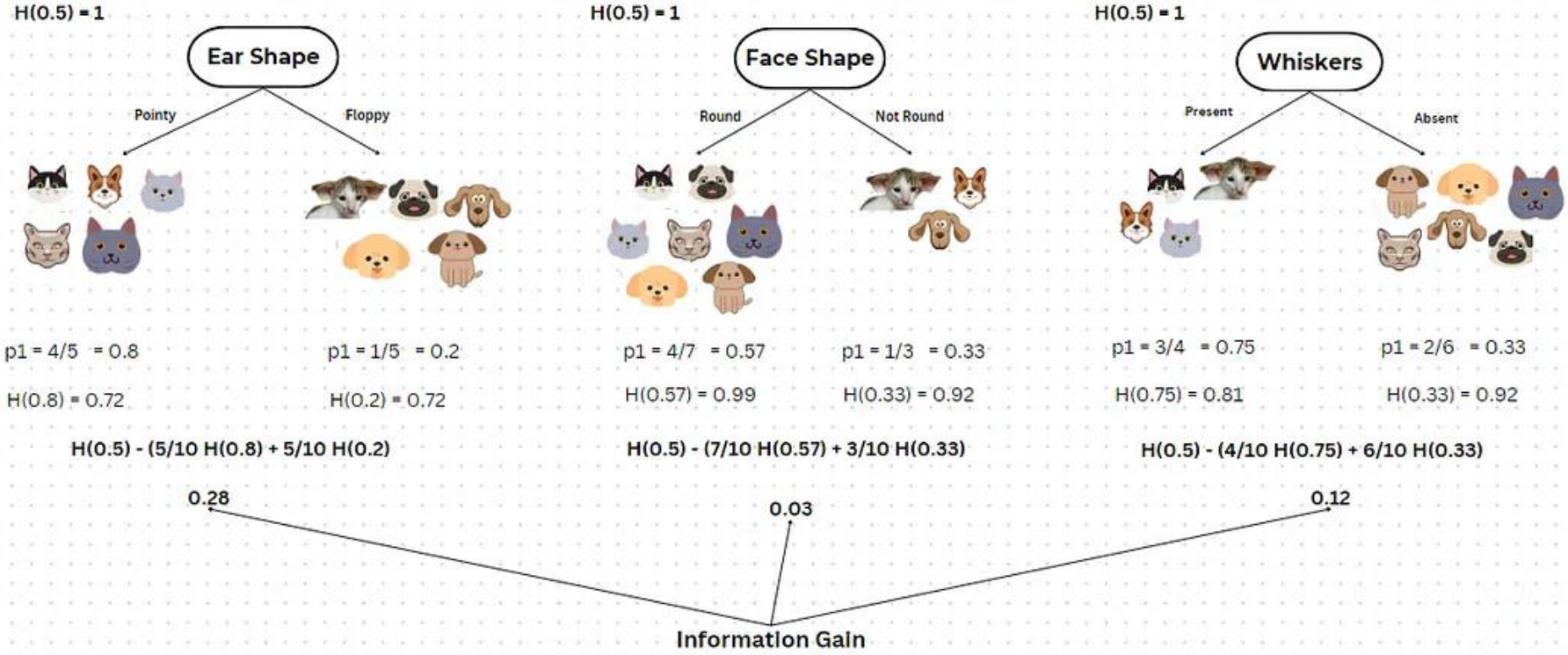
Ear Shape	Face Shape	Whiskers	Cat
 Pointy	Round	Present	1
 floppy	Not Round	Present	1
 Floppy	Round	Absent	0
 Pointy	Not Round	Present	0
 Pointy	Round	Present	1
 Pointy	Round	Absent	1
 Floppy	Not Round	Absent	0
 Pointy	Round	Absent	1
 Floppy	Round	Absent	0
 Floppy	Round	Absent	0



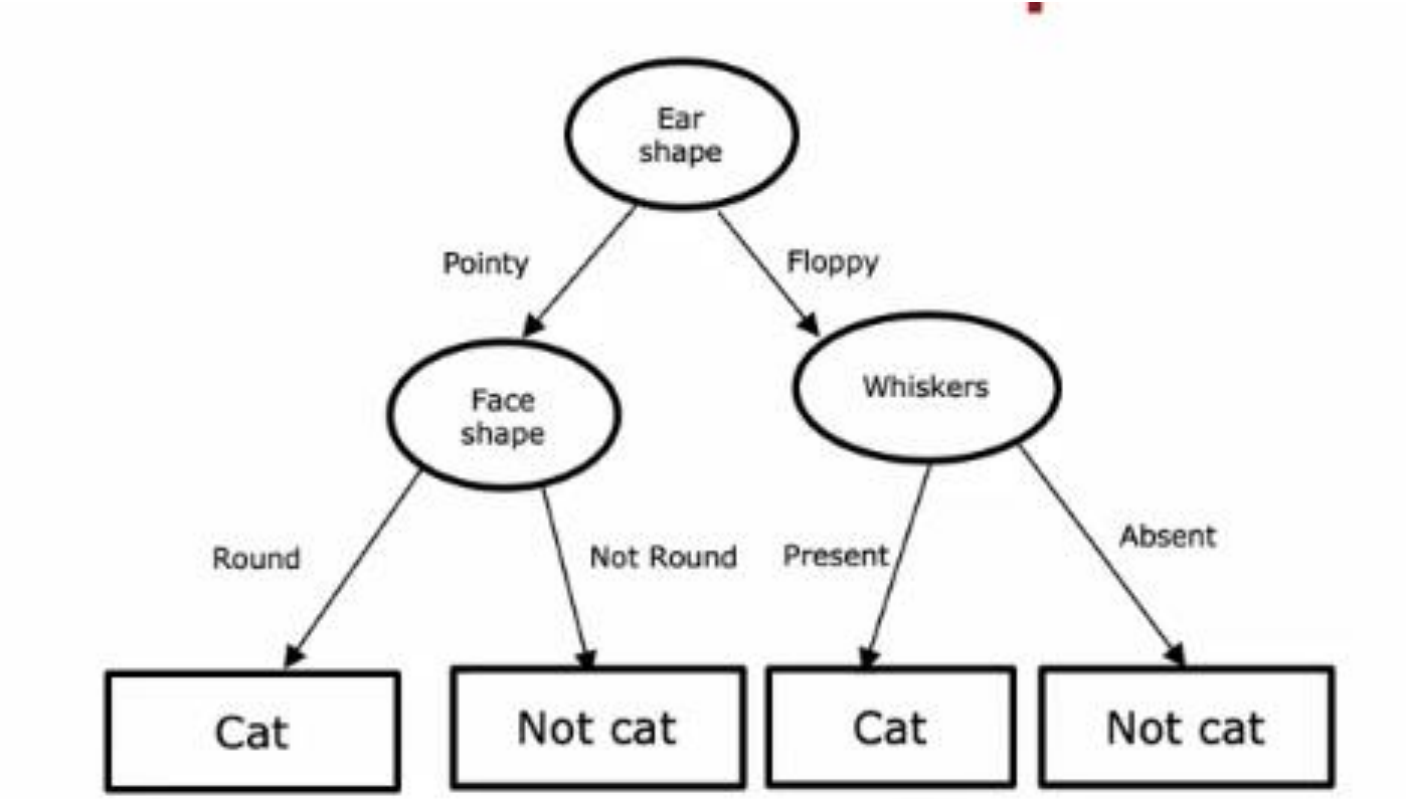
$$p_0 = 1 - p_1$$

$$H(p1) = -p_1 \log_2(p_1) - p_0 \log_2(p_0)$$

$$H(p1) = -p_1 \log_2(p_1) - (1 - p_1) \log_2(1 - p_1)$$

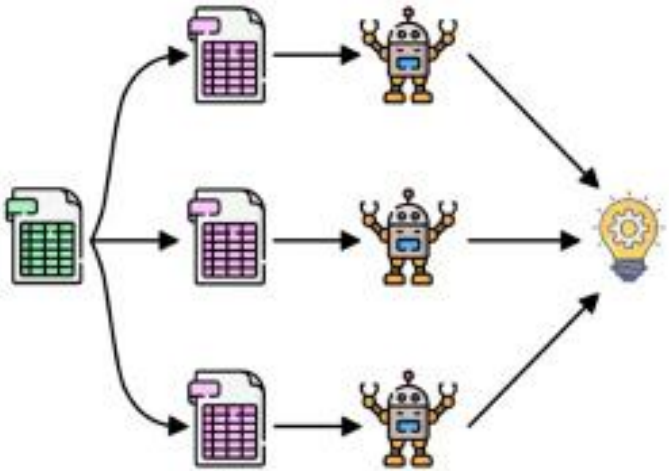


$$Information\ Gain: \\ = H(p_1^{root}) - (w^{left}\ H(p_1^{left}) + w^{right}\ H(p_1^{right}))$$


Ear shape
Pointy
Floppy
Face shape
Whiskers
Round
Not Round
Present
Absent
Cat
Not cat
Cat
Not cat

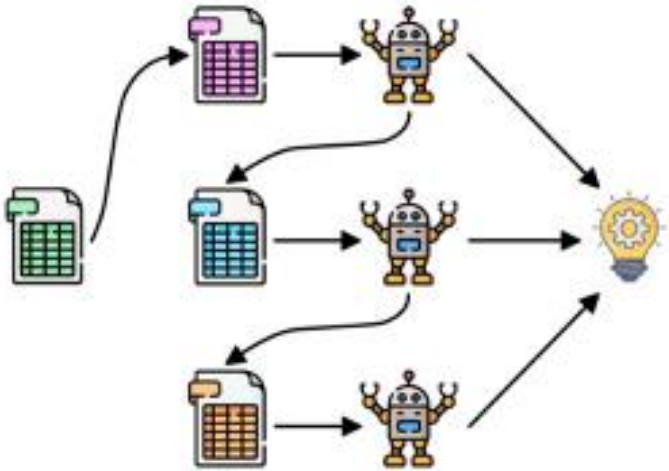
Bagging and Boosting

Bagging



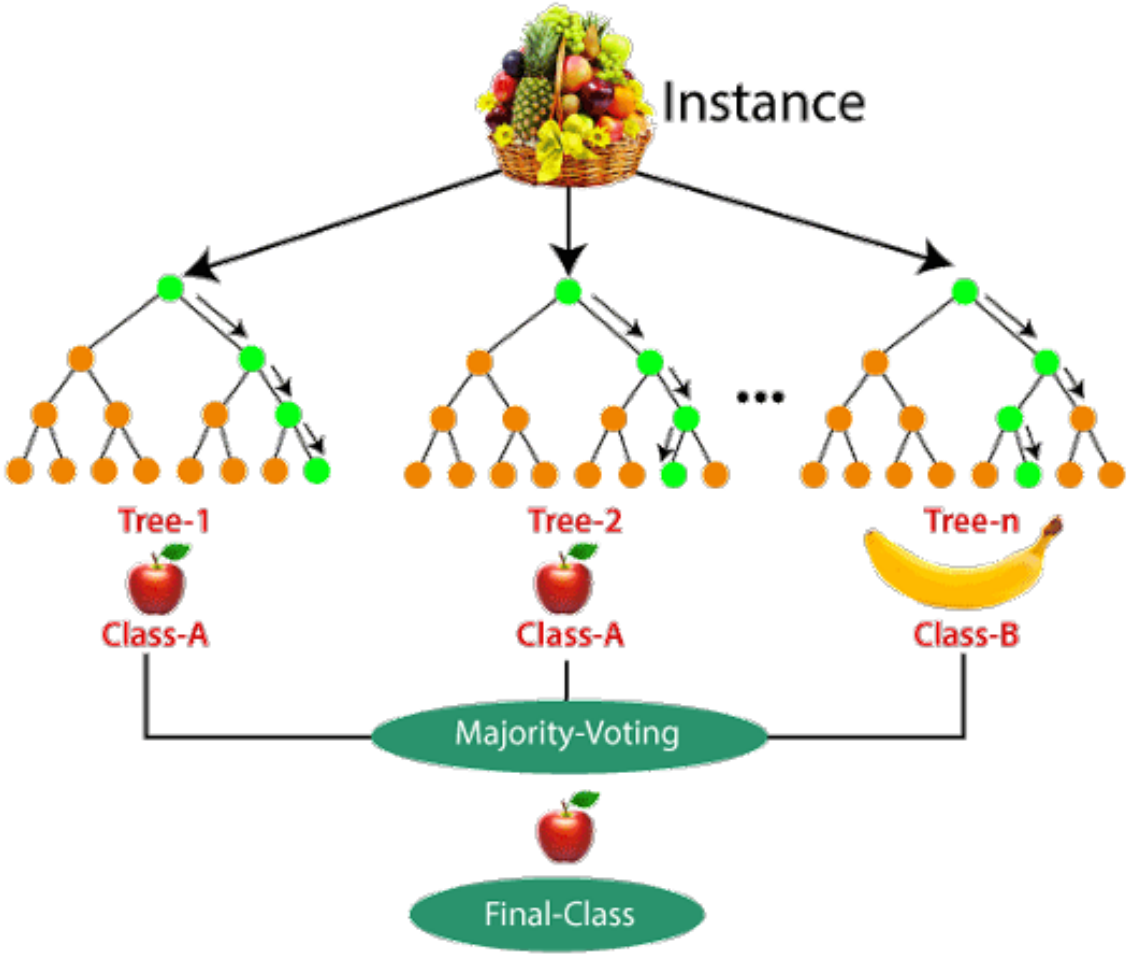
Parallel

Boosting



Sequential

Random Forest


Instance
Tree-1
Tree-2
Tree-n
Class-A
Class-A
Class-B
Majority-Voting
Final-Class