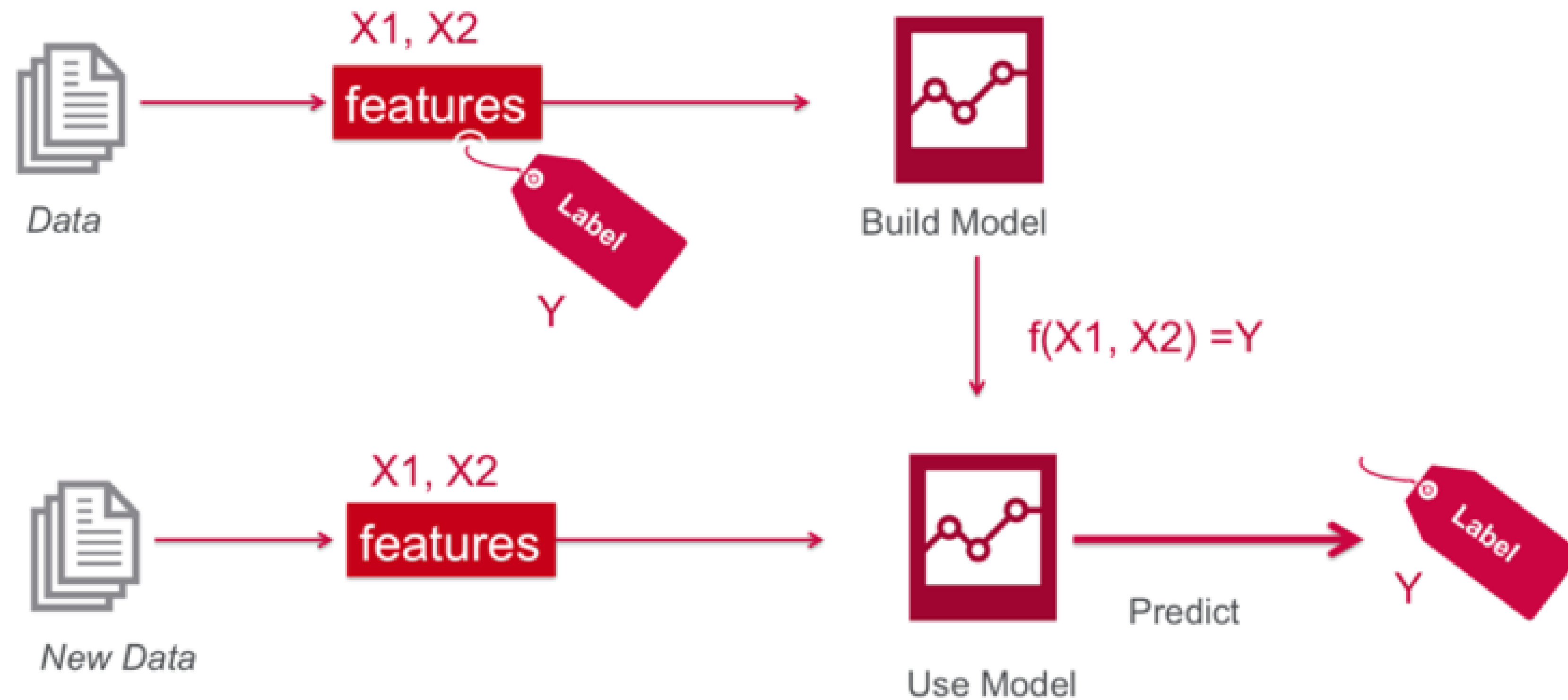




Decision tree Classification

Supervised learning

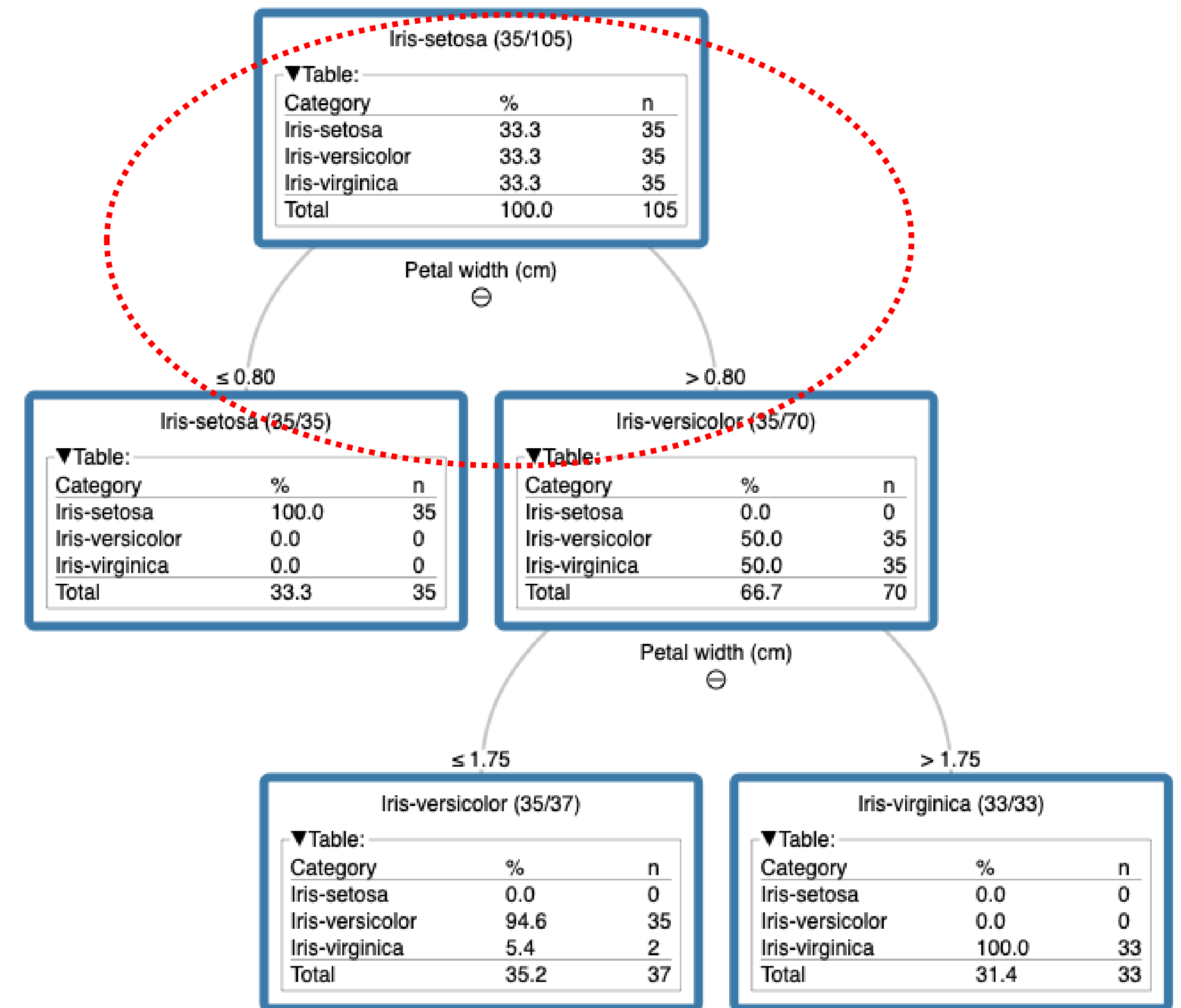
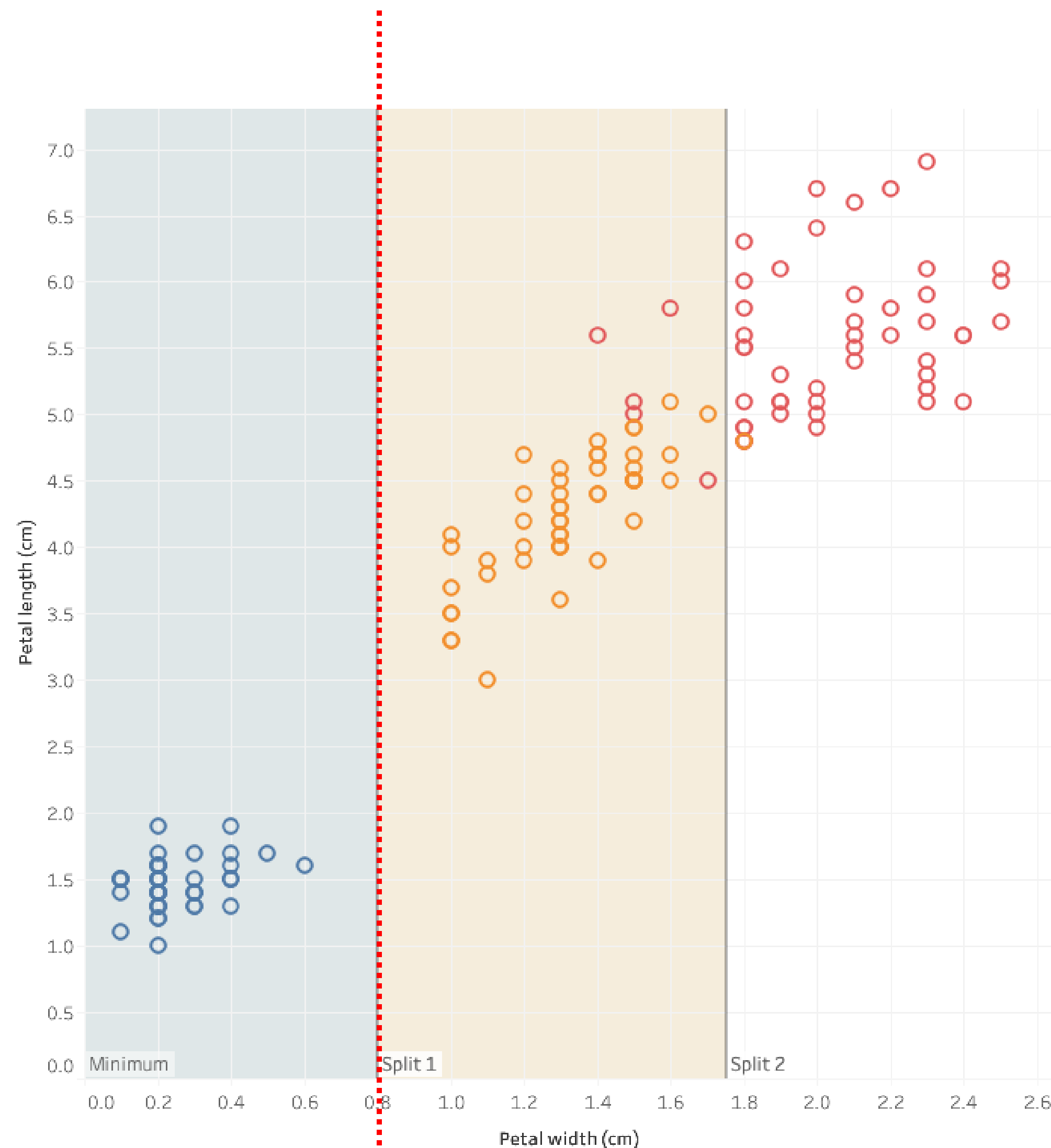
2



Demo in Tableau & KNIME

Back to our Iris example

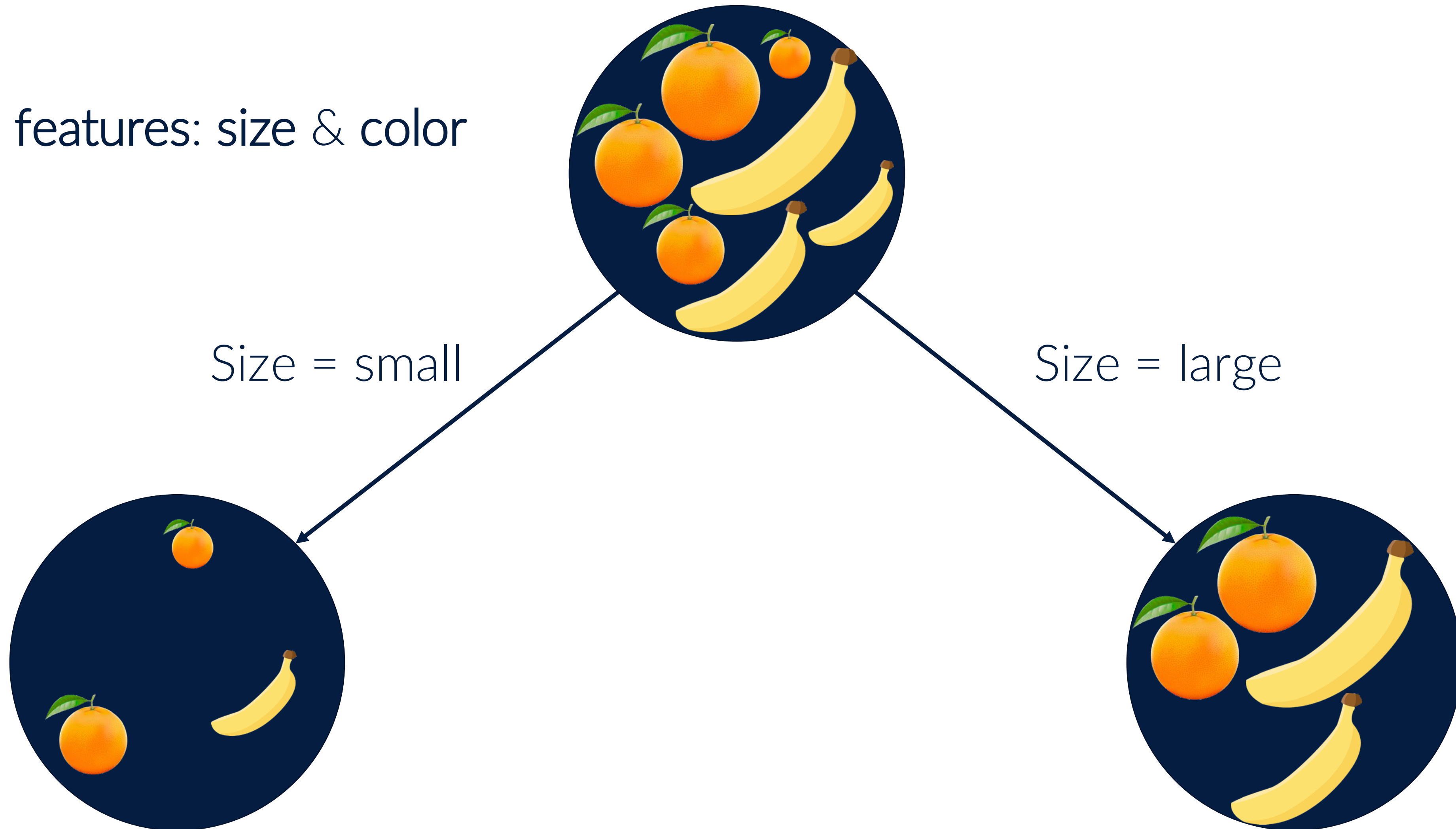
How is this first split decided?
Feature and split point of 0.8?



Measuring the quality of a split

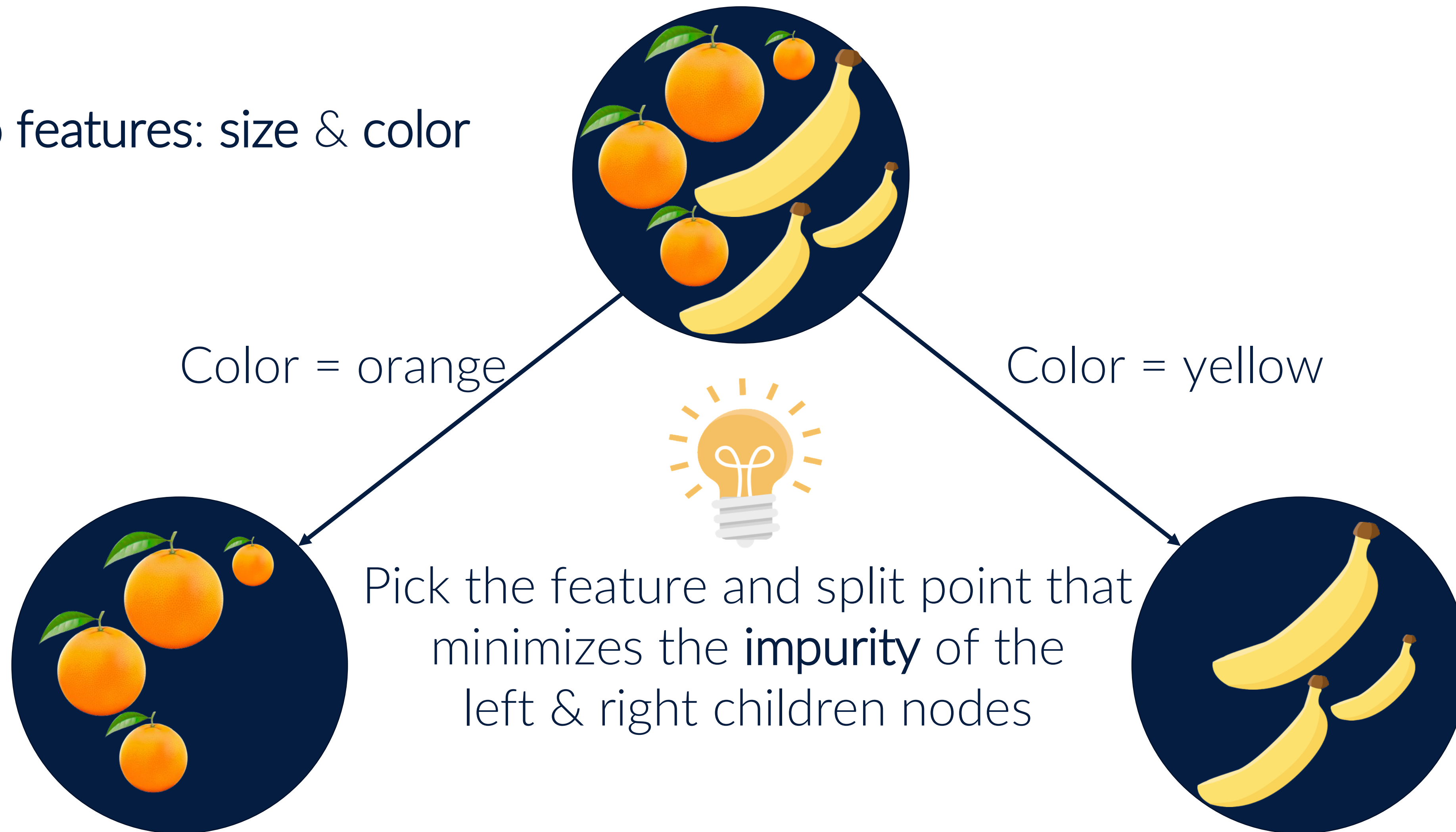
5

Say, two features: size & color



Measuring the quality of a split

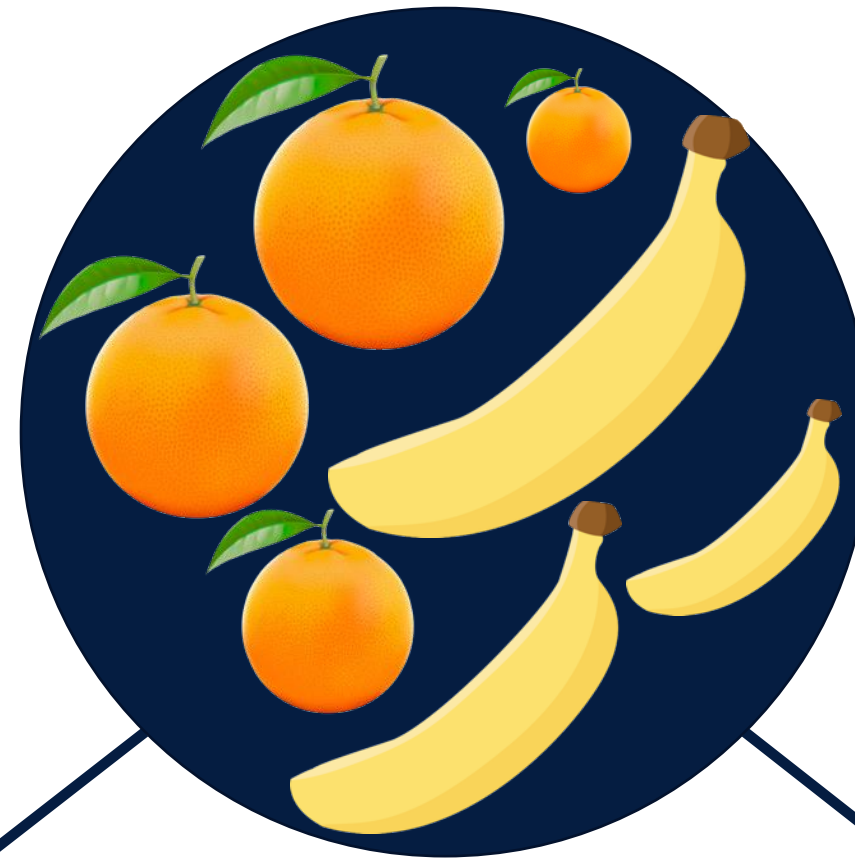
Say, two features: size & color



Entropy as a measure of impurity

7

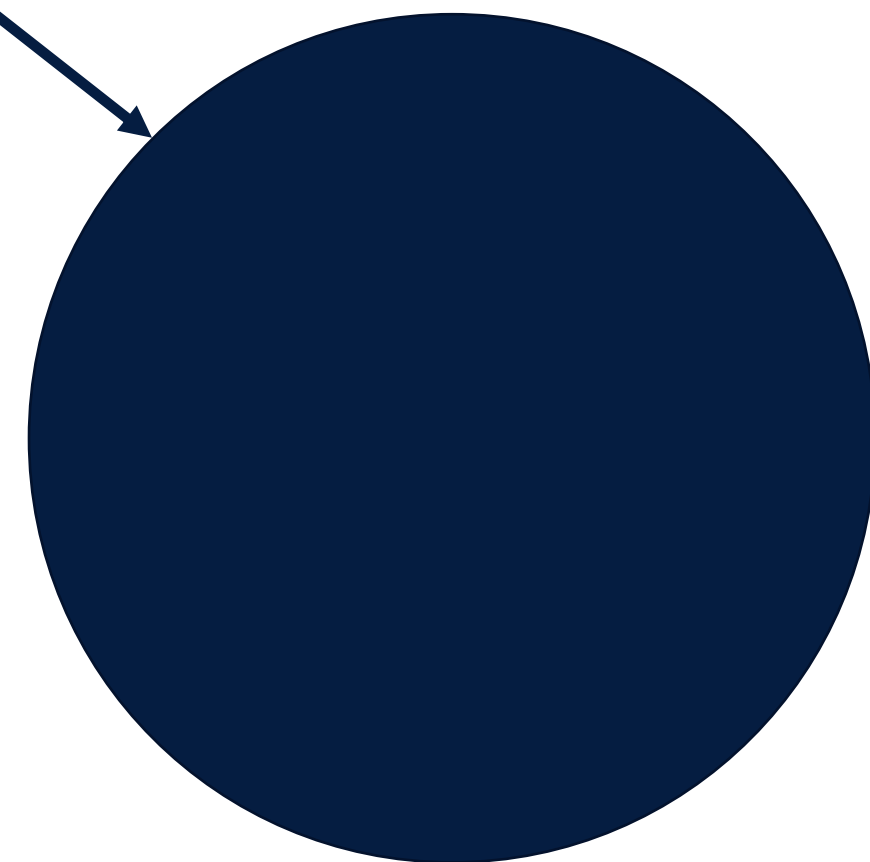
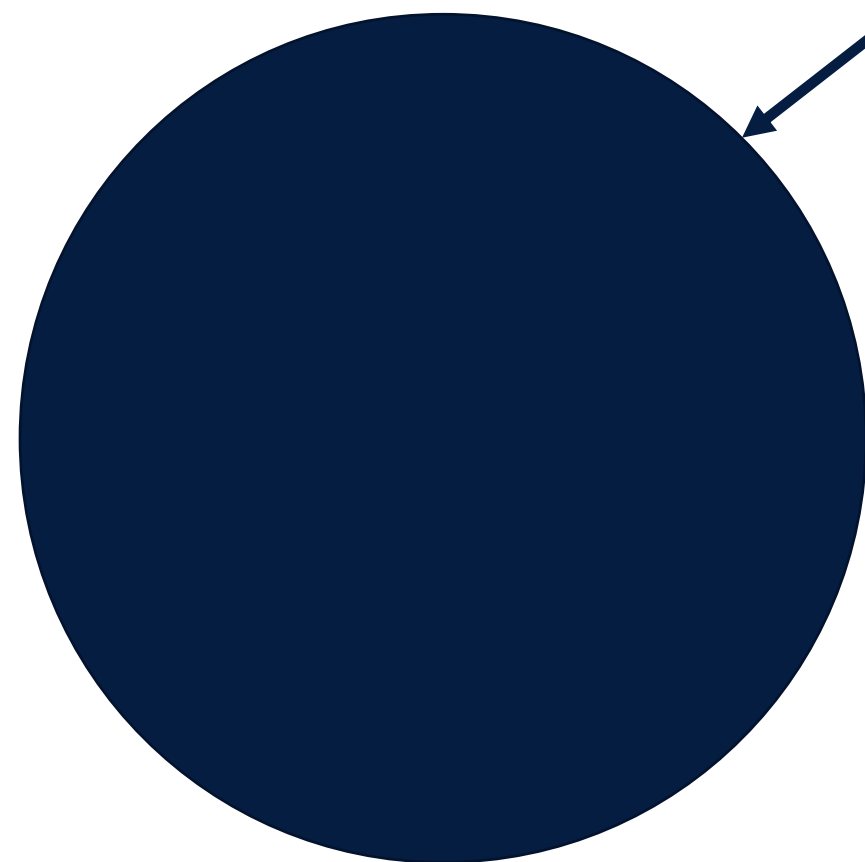
- $Entropy = - \sum_{i=1}^k p_i \log_2(p_i)$
- p_i : relative frequency of class i



$$p_{\text{orange}}^0 = 4/7$$

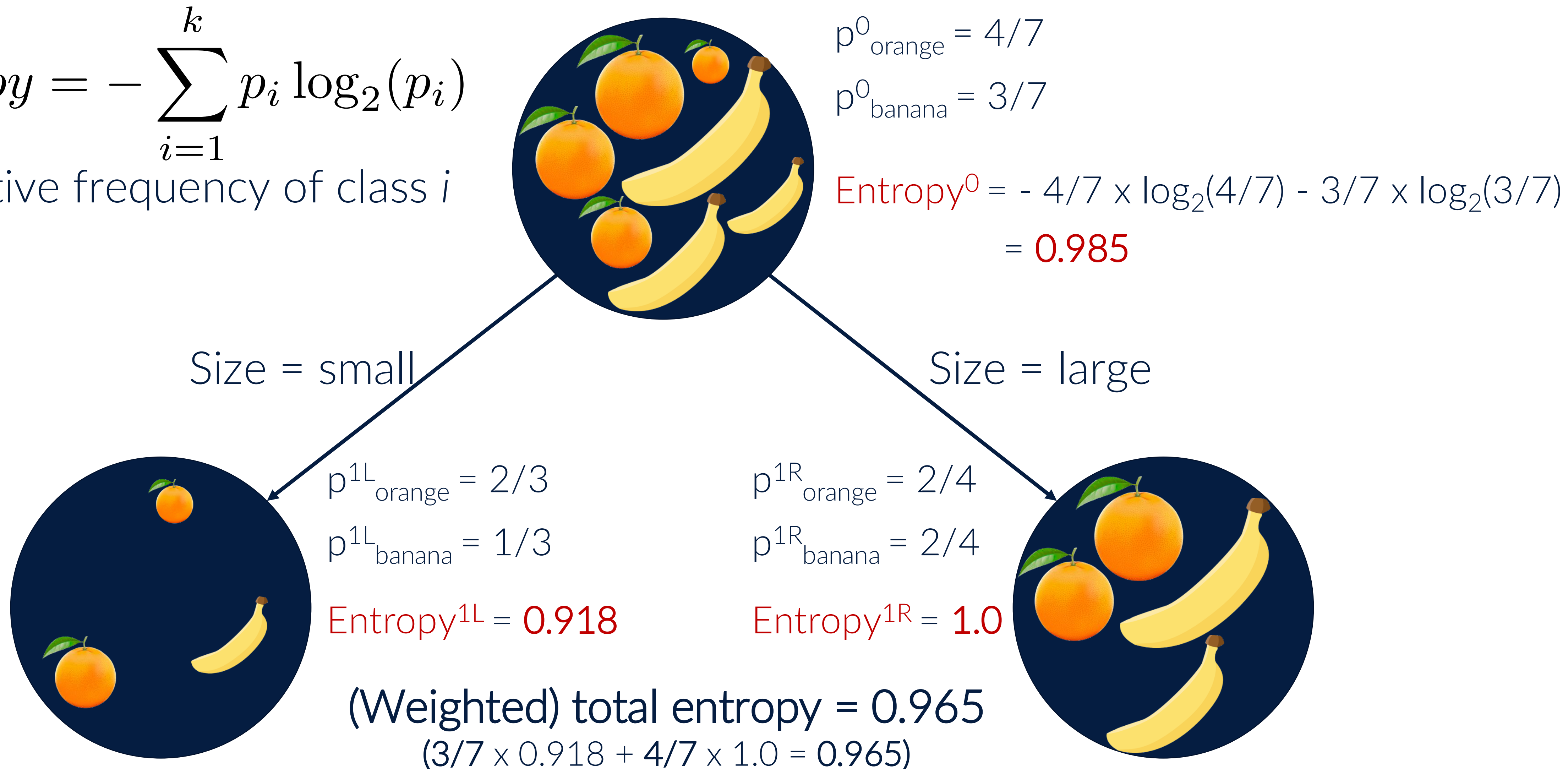
$$p_{\text{banana}}^0 = 3/7$$

$$Entropy^0 = - 4/7 \times \log_2(4/7) - 3/7 \times \log_2(3/7) \\ = 0.985$$



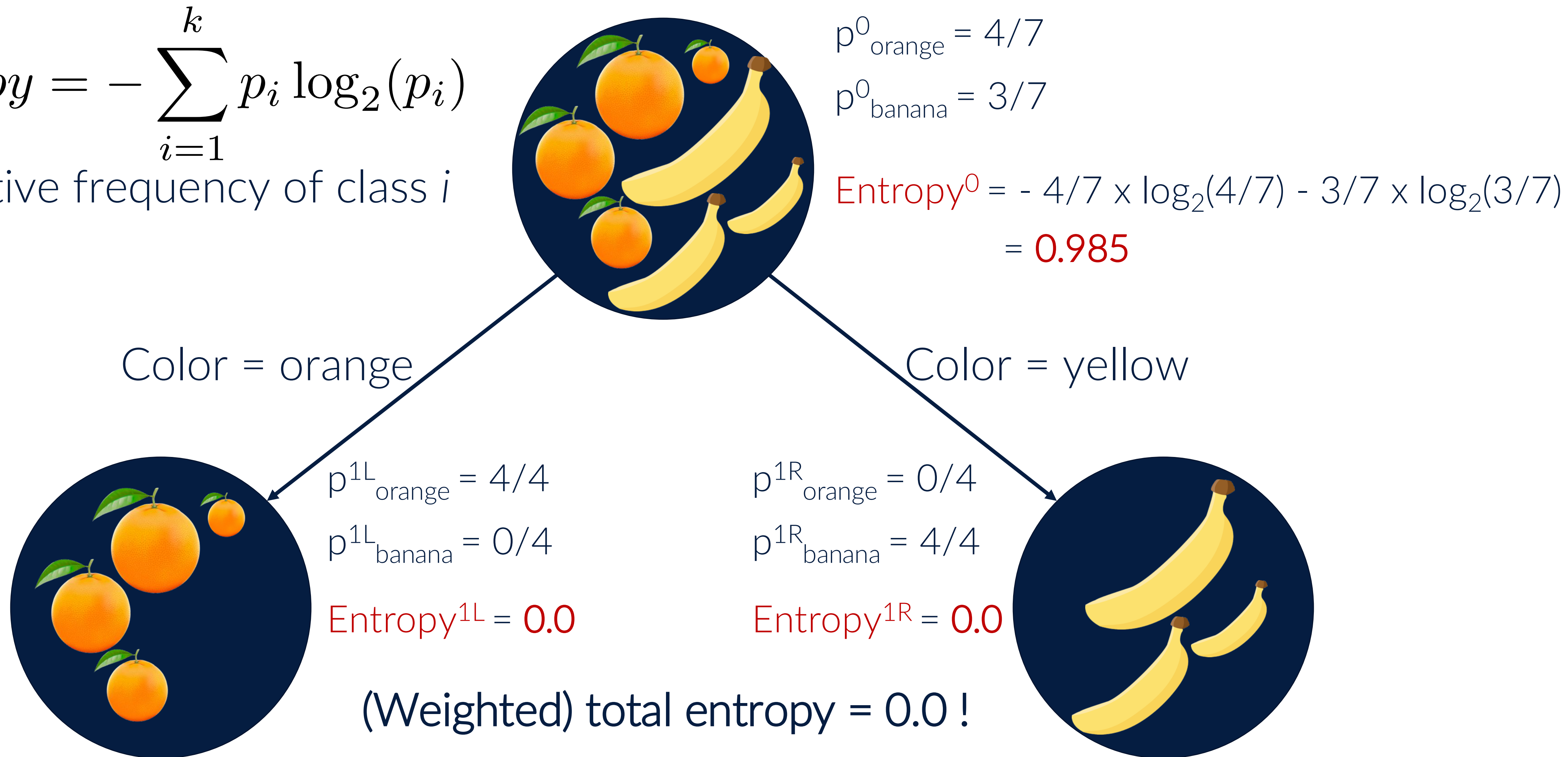
Entropy as a measure of impurity

- $Entropy = - \sum_{i=1}^k p_i \log_2(p_i)$
- p_i : relative frequency of class i



Entropy as a measure of impurity

- $Entropy = - \sum_{i=1}^k p_i \log_2(p_i)$
- p_i : relative frequency of class i



Information Gain as split decision

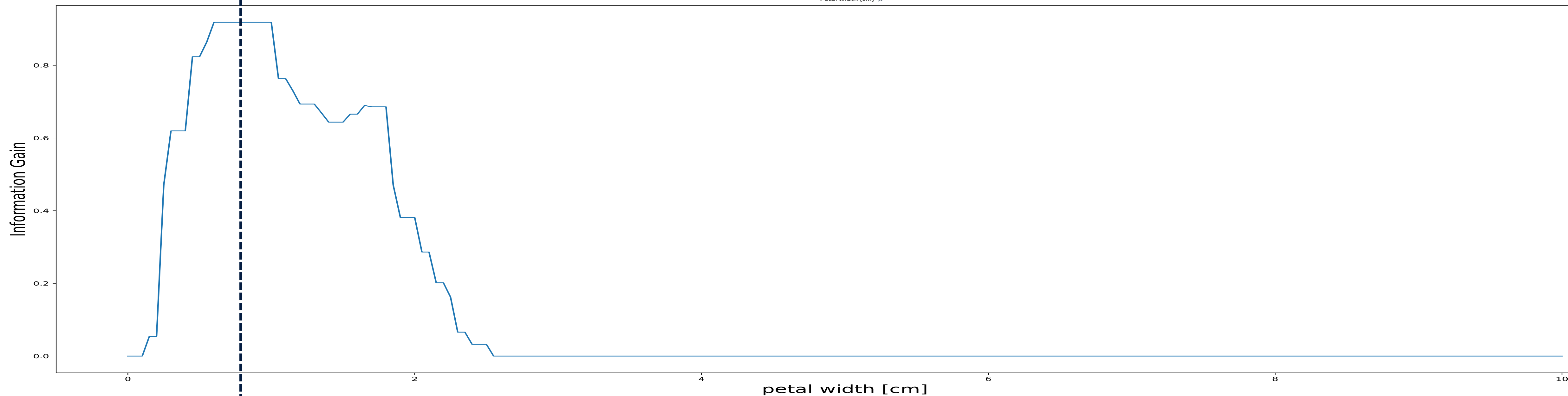
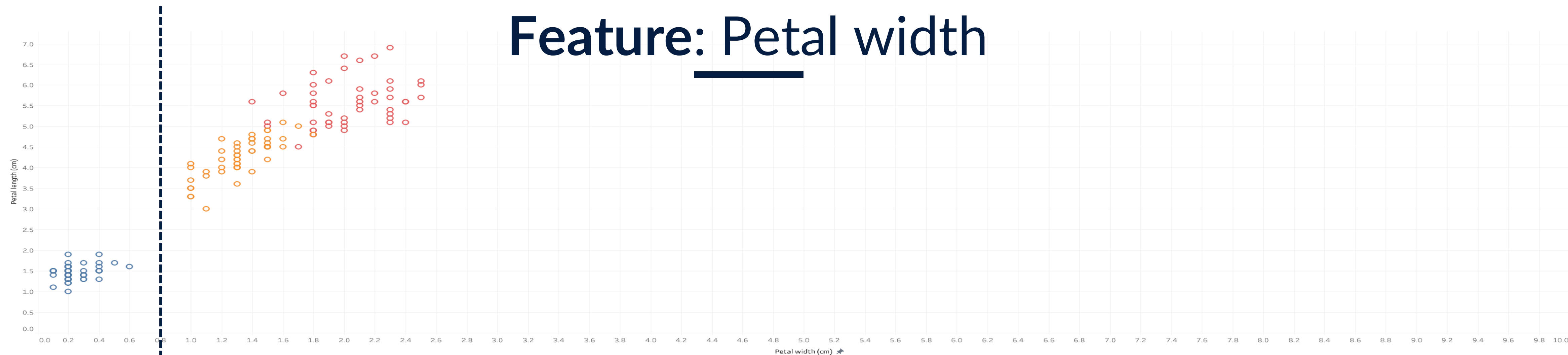
10

- Information Gain = How much entropy we removed by the split
- The split (feature & value) that maximizes the Information Gain is chosen

Split by size: $0.985 - 0.965 = 0.02$
Split by color: $0.985 - 0.0 = 0.985$

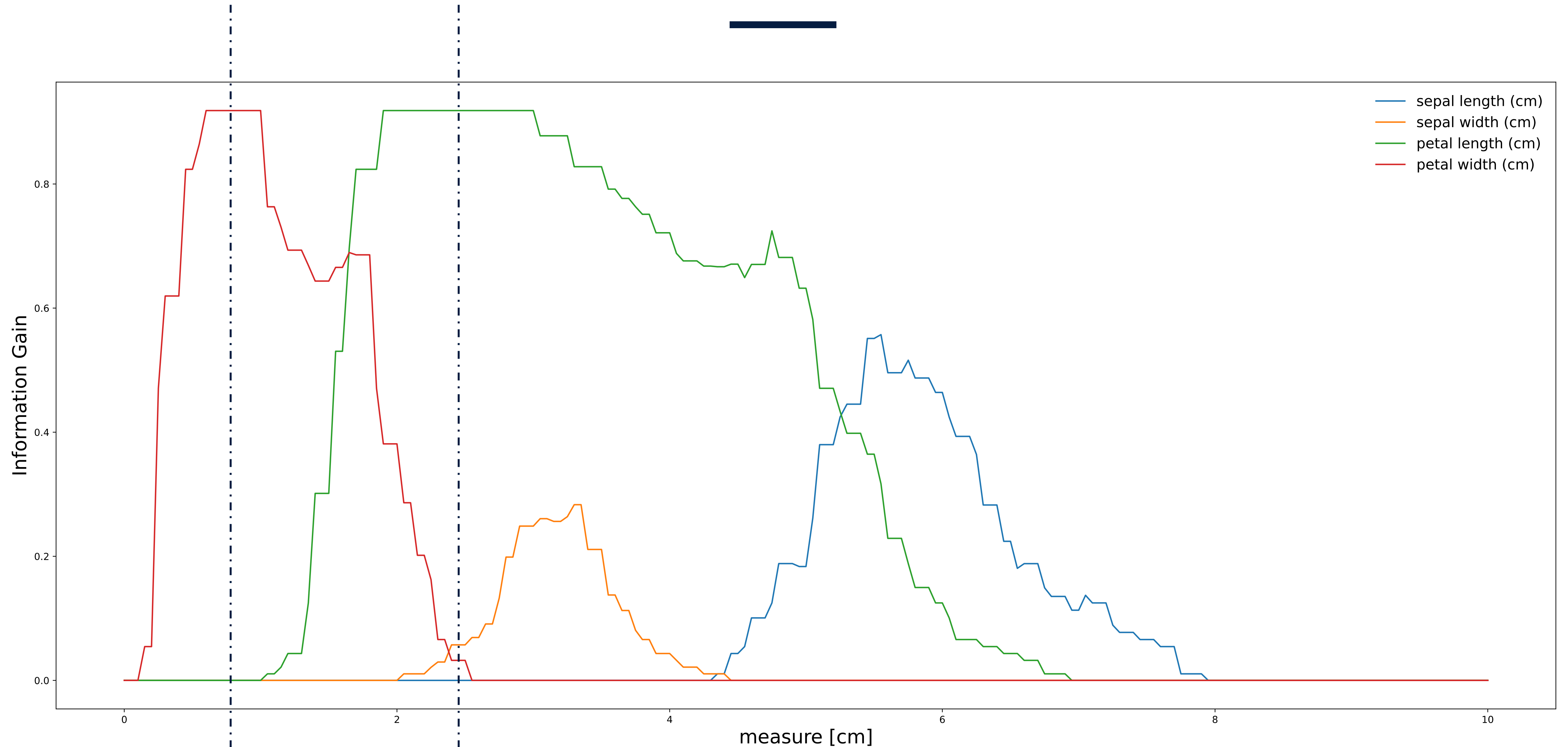
Back to our Iris example

Feature: Petal width



Over all 4 features

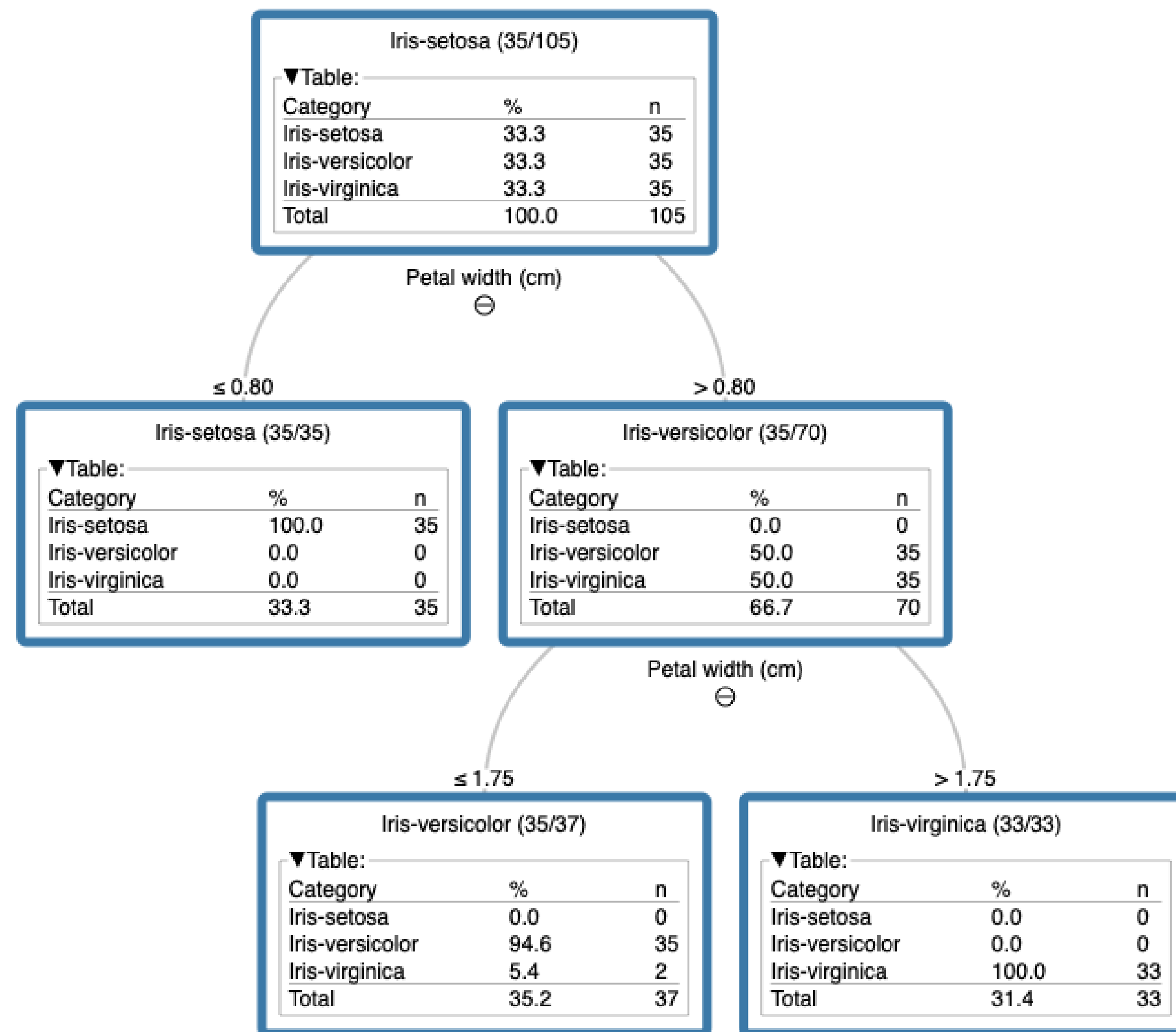
12



Recursively splitting

Stopping criterion (overfitting!)

- $\text{purity} > \text{threshold}$
- $\text{number of samples} < \text{threshold}$
- $\text{depth} > \text{threshold}$



Another split criterion

Gini index (or Gini impurity)

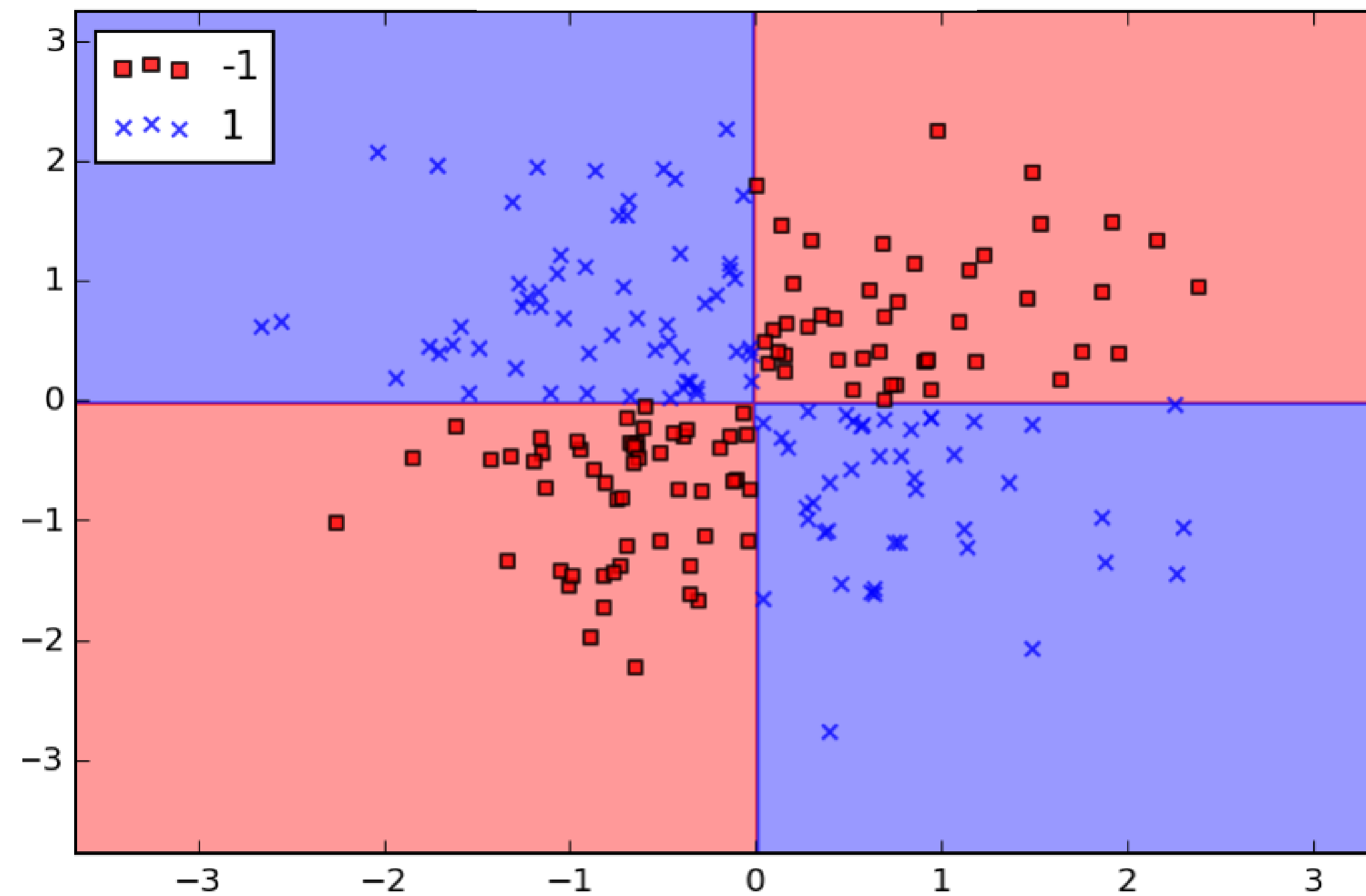
$$Gini = \sum_{i=1}^k p_i(1 - p_i) = 1 - \sum_{i=1}^k p_i^2$$

- The Gini Impurity of a dataset is a number between 0 and 0.5
- Gini impurity measures how often a randomly chosen element of a set (of cardinality k) would be incorrectly labeled if it were labeled randomly and independently according to the distribution of labels in the set.

More complex datasets

15

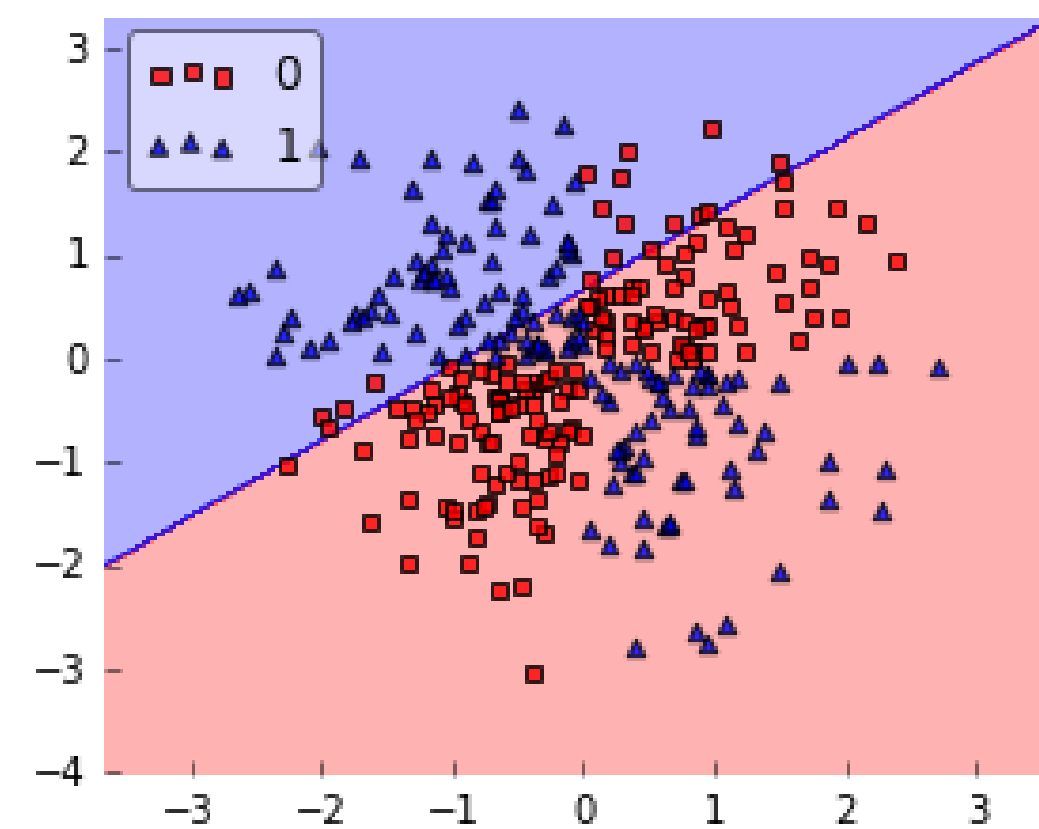
XOR



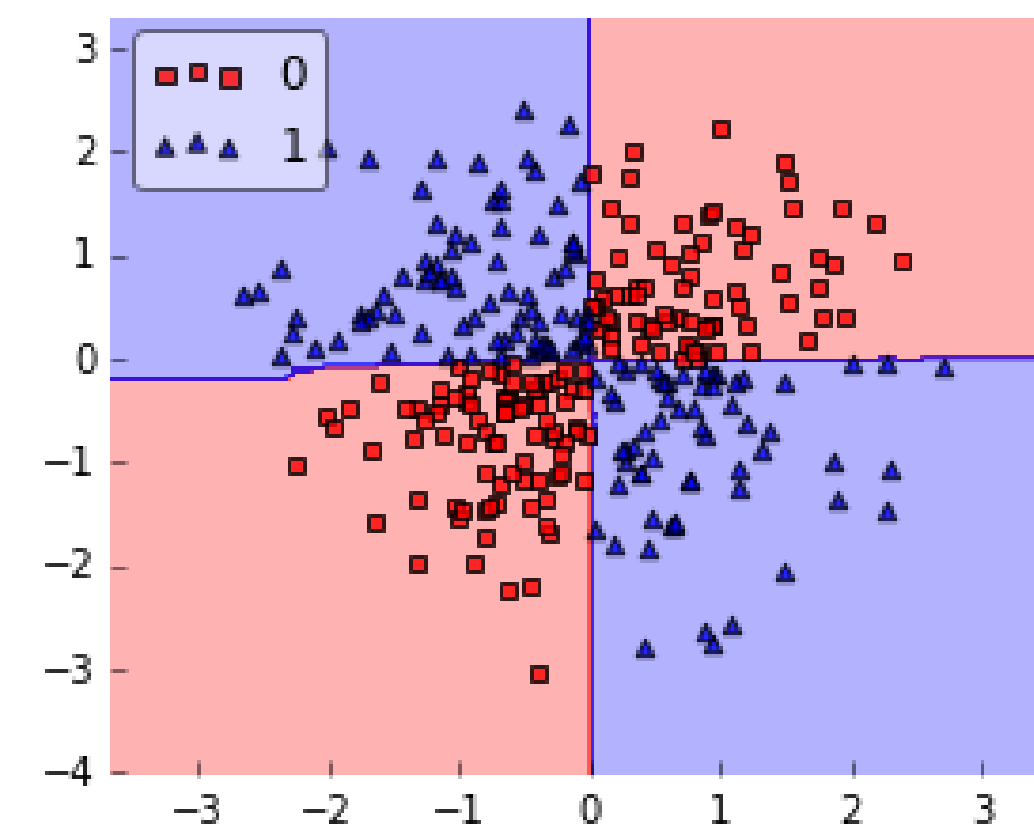
More complex datasets

XOR

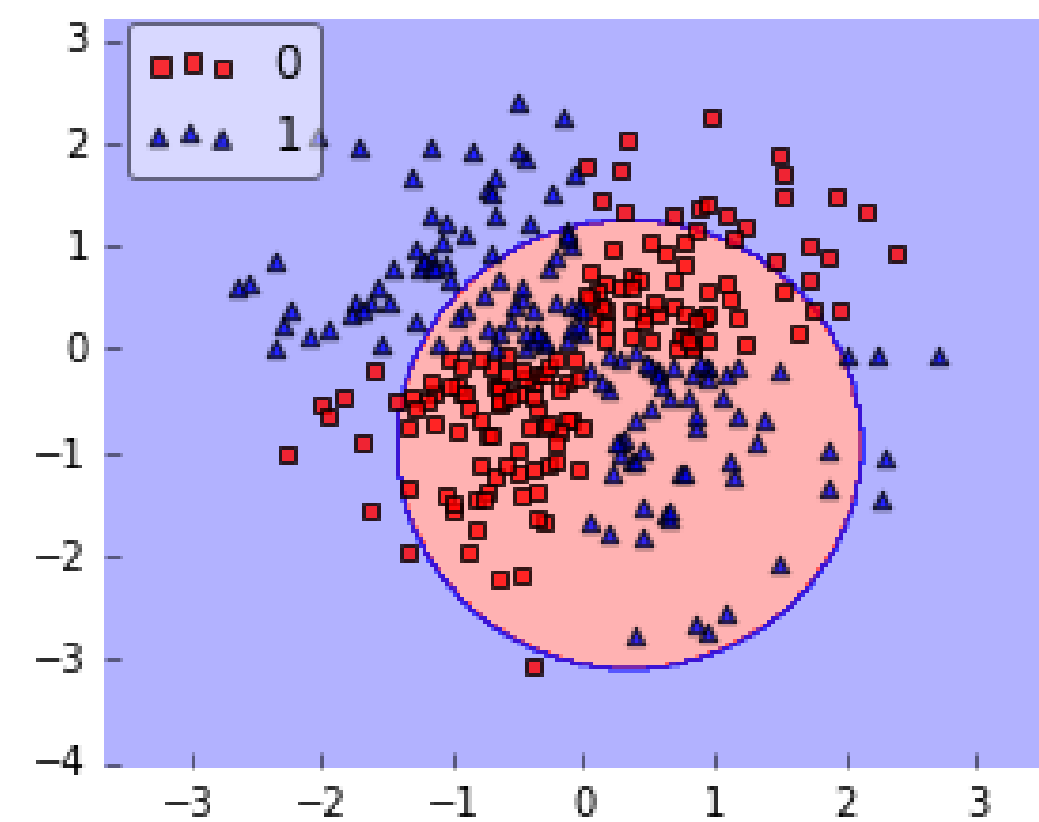
Logistic regression



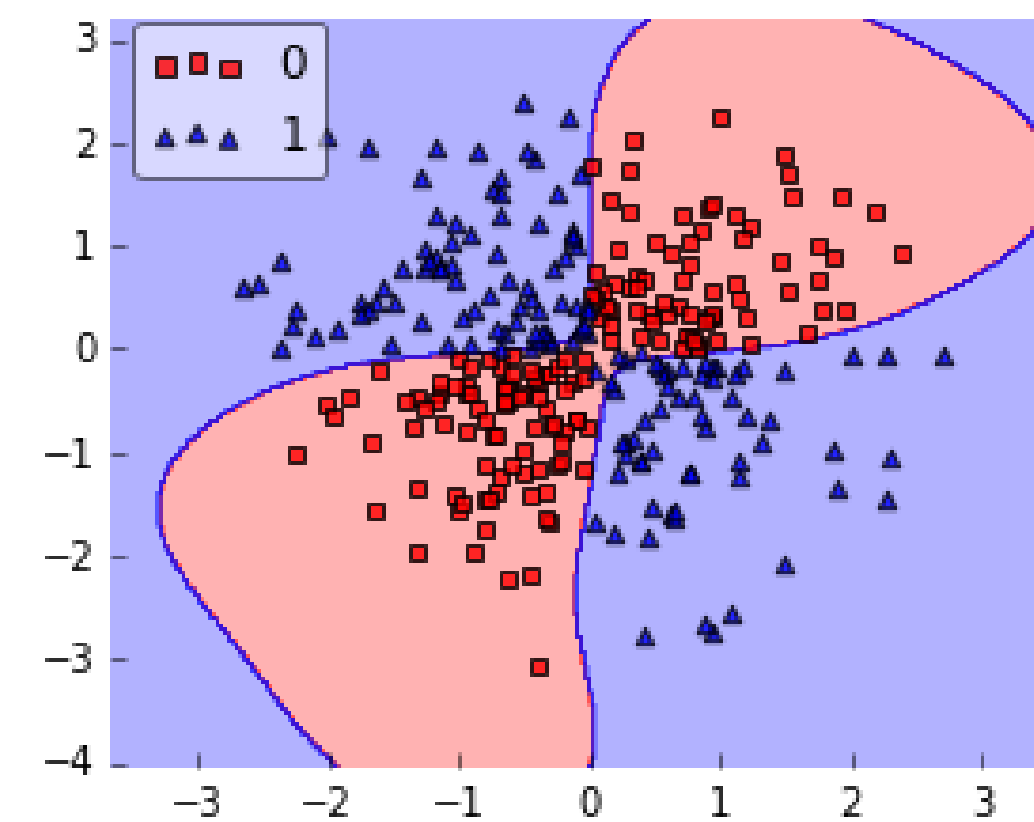
Deeper decision tree



Naïve Bayes



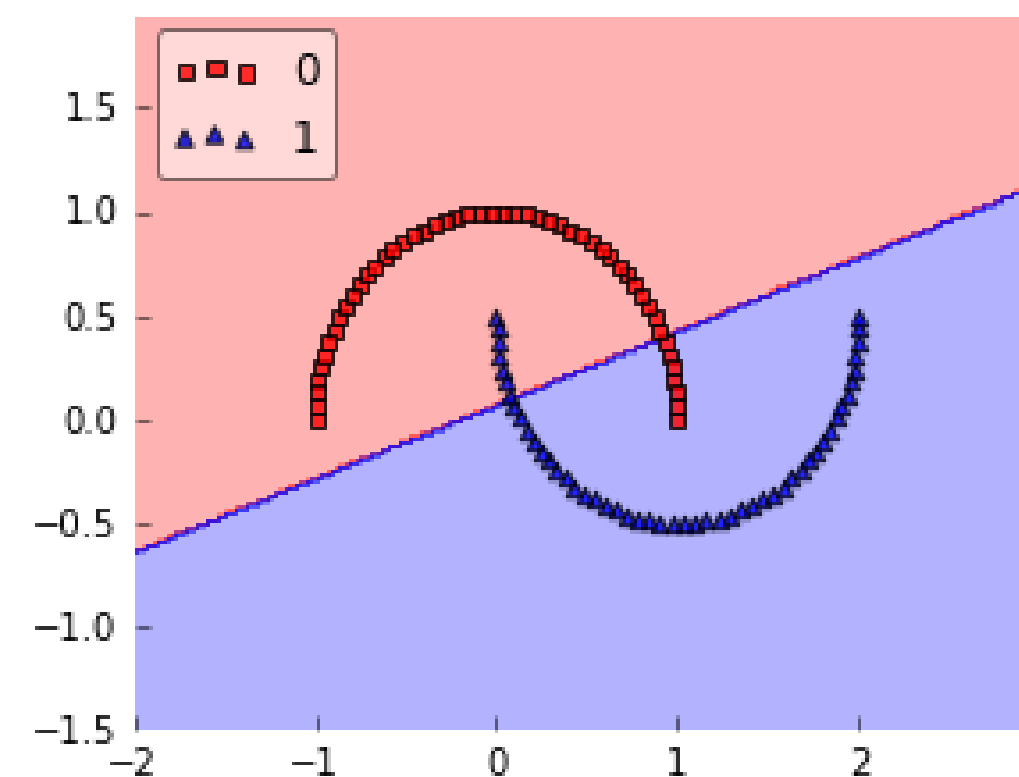
Kernel SVM



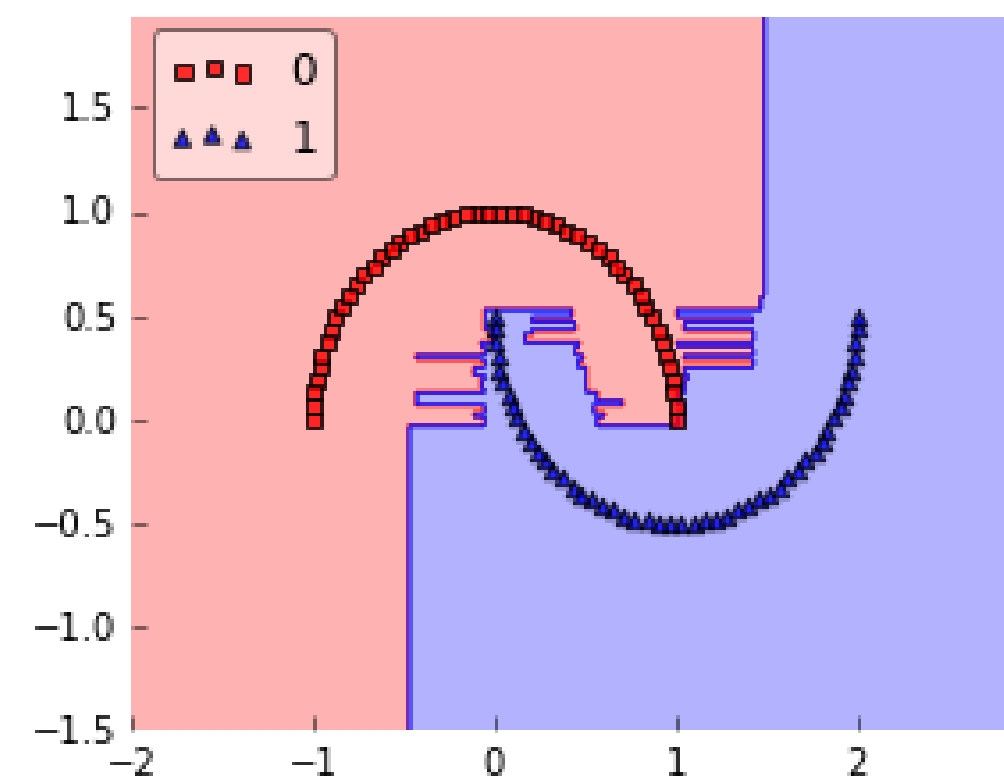
More complex datasets

Half-Moons

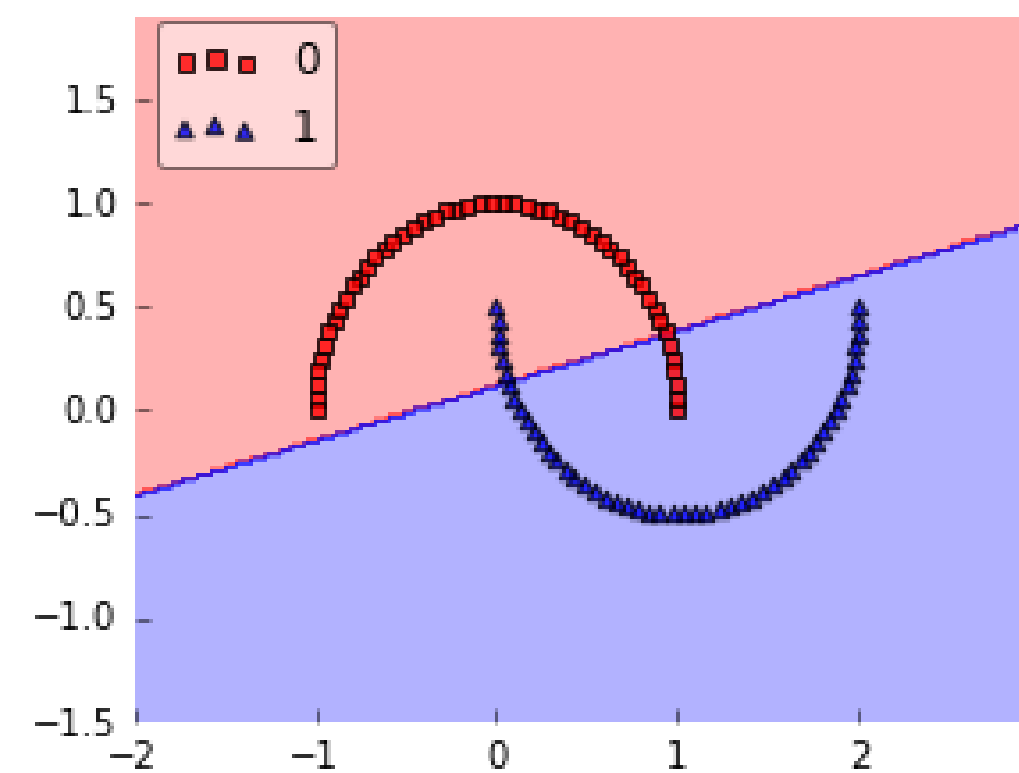
Logistic regression



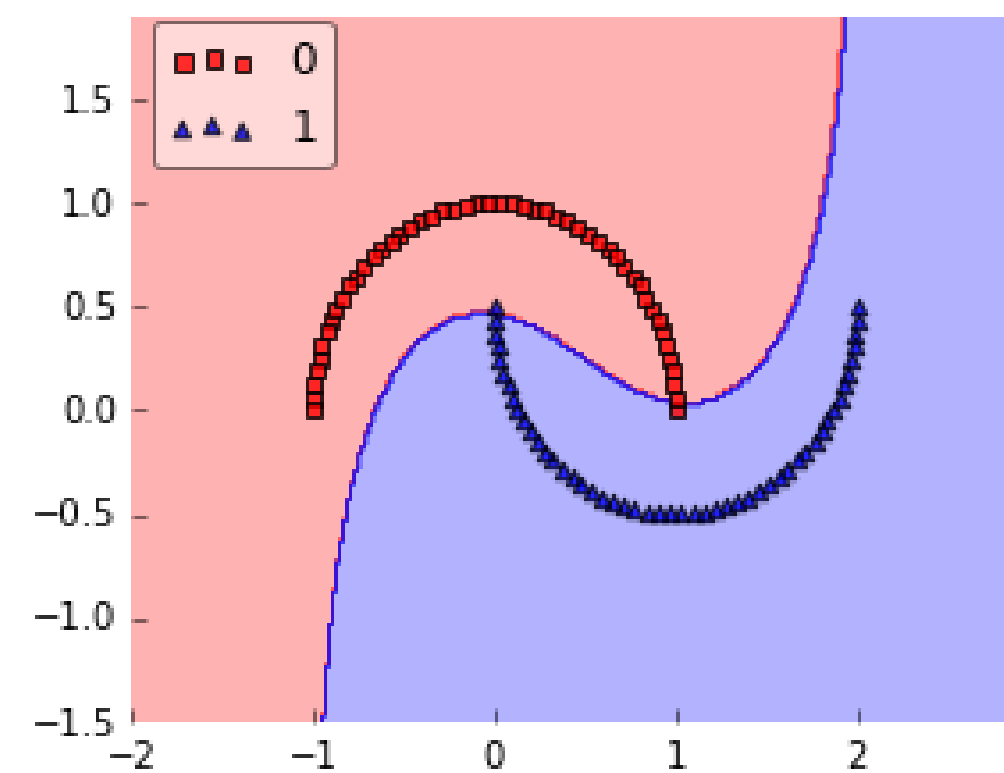
Deeper decision tree



Naïve Bayes



Kernel SVM



(Dis)Advantages of Decision Trees

18

Advantages:

1. Interpretability
2. Less Data Preparation
3. Non-Parametric
4. Non-Linearity

Disadvantages:

1. Overfitting
2. Unstability



Random forest classifier

- More stable
- Less prone to overfitting

Random forest classifier

19

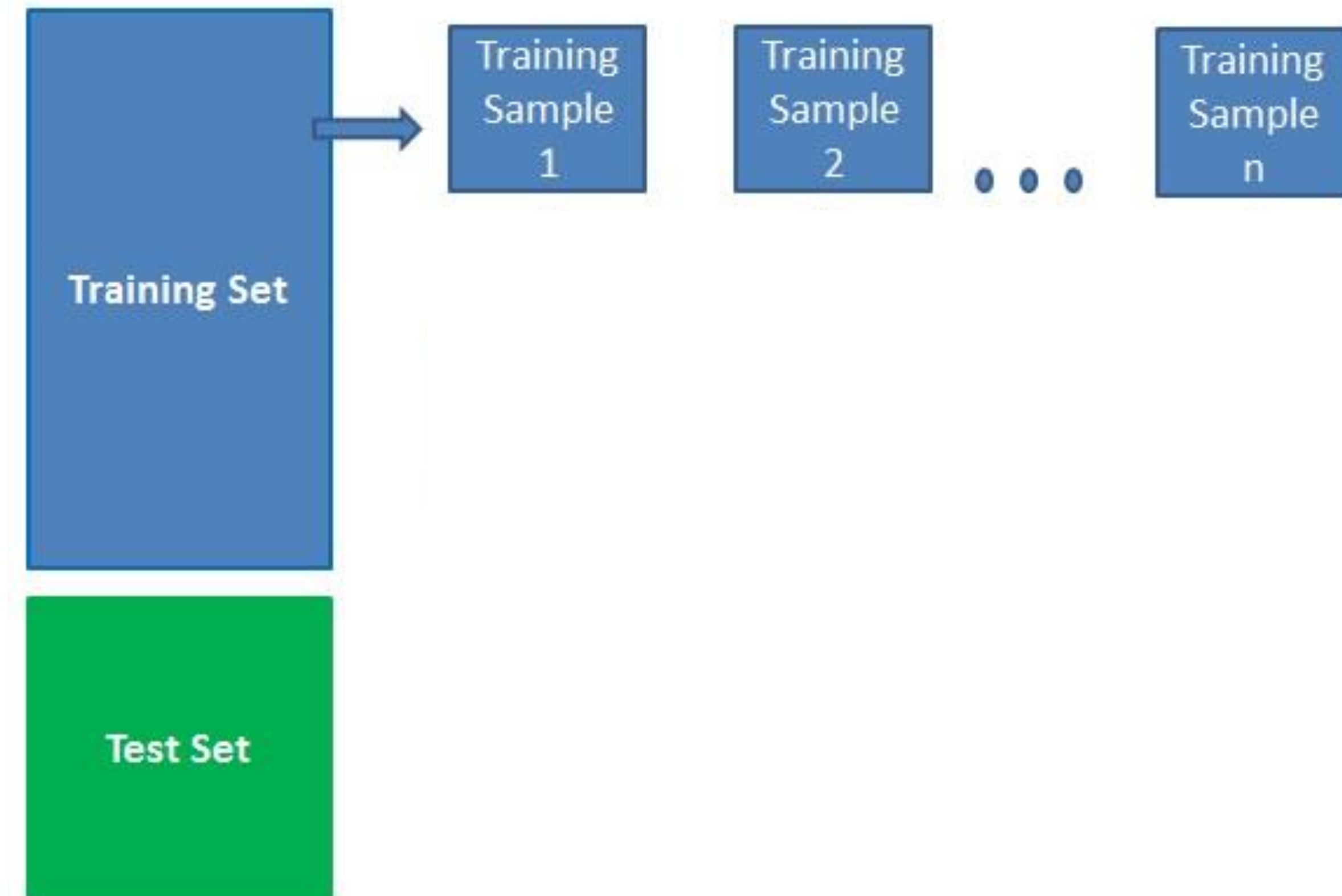
- Works in 4 steps



Random forest classifier

20

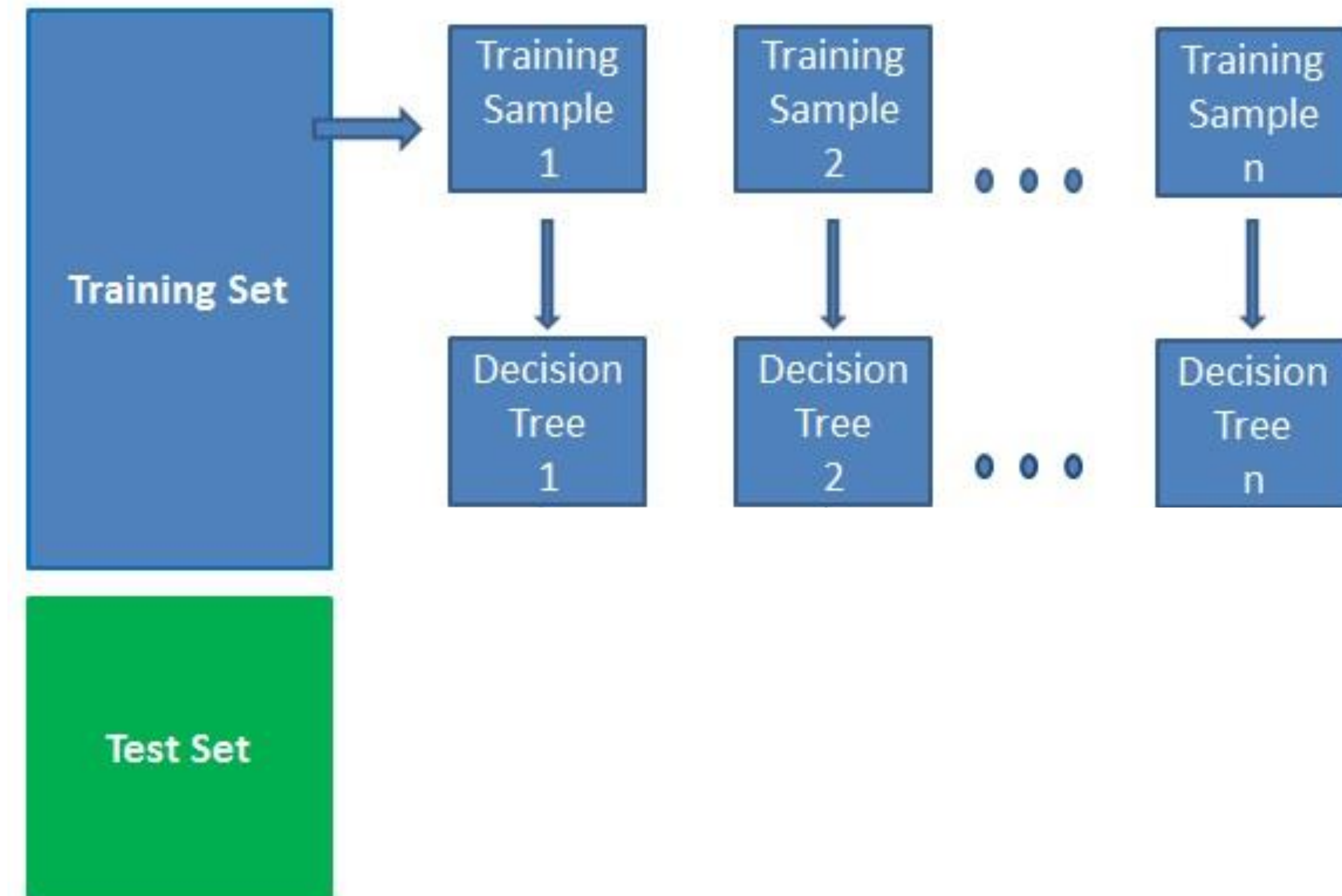
- Works in 4 steps:
 1. Select random samples from a given dataset



Random forest classifier

21

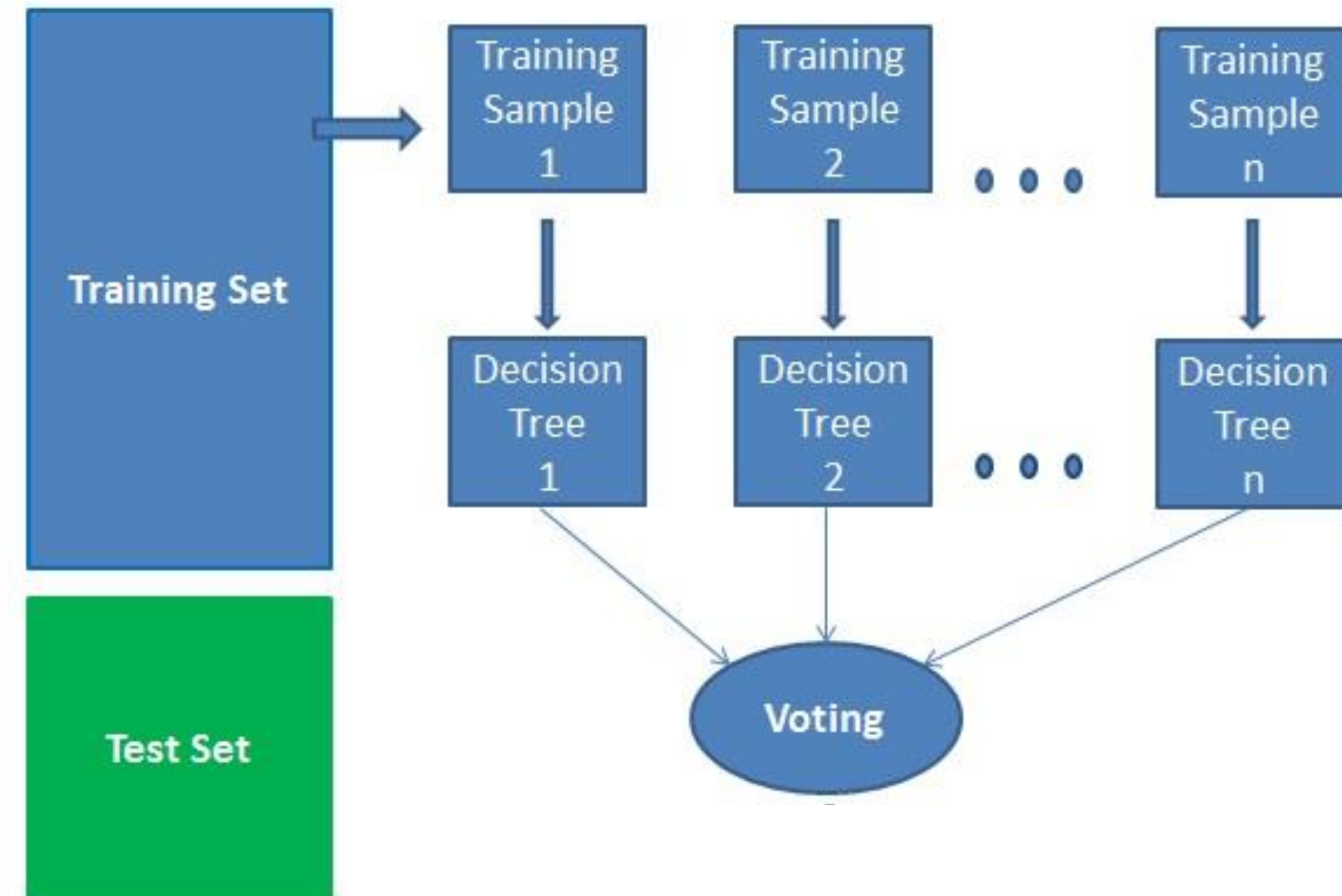
- Works in 5 steps:
 1. Select random samples from a given dataset
 2. Construct a decision tree for each sample



Random forest classifier

22

- Works in 5 steps:
 1. Select random samples from a given dataset
 2. Construct a decision tree for each sample
 3. Get a prediction from each tree
 4. Perform a vote for each predicted result



Random forest classifier

23

- Works in 5 steps:
 1. Select random samples from a given dataset
 2. Construct a decision tree for each sample
 3. Get a prediction from each tree
 4. Perform a vote for each predicted result
 5. Select the prediction result with the most votes as the final prediction

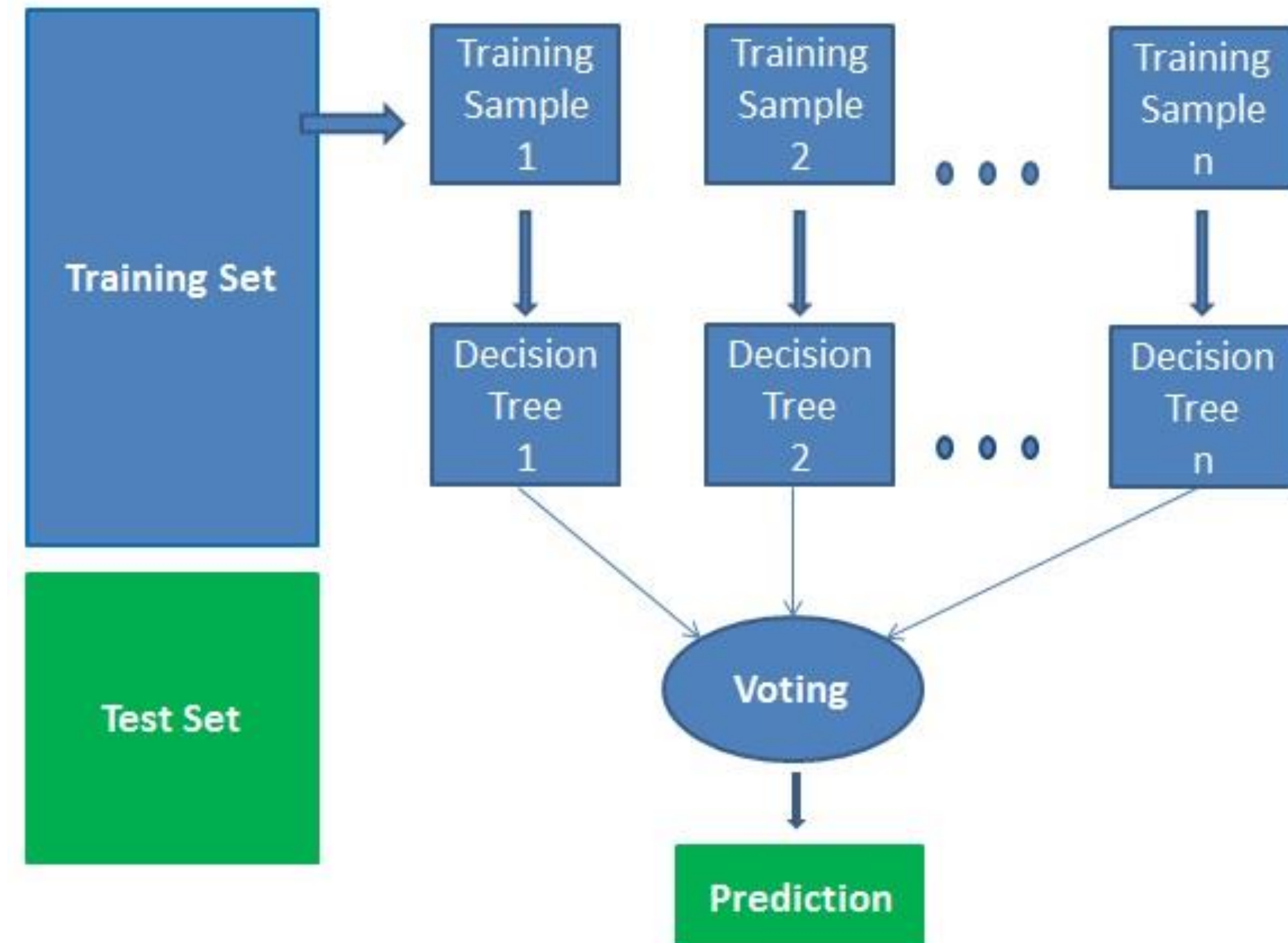
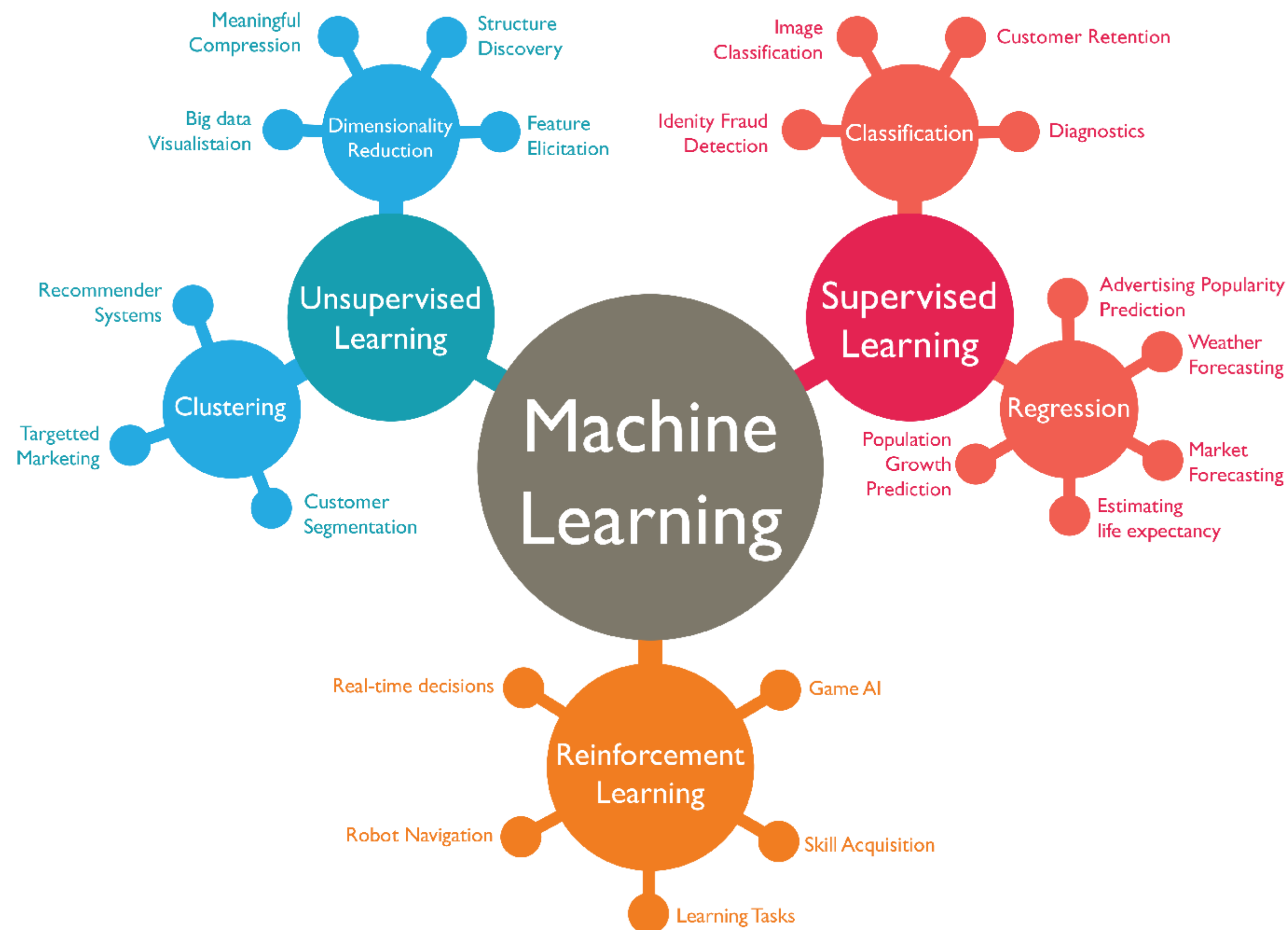


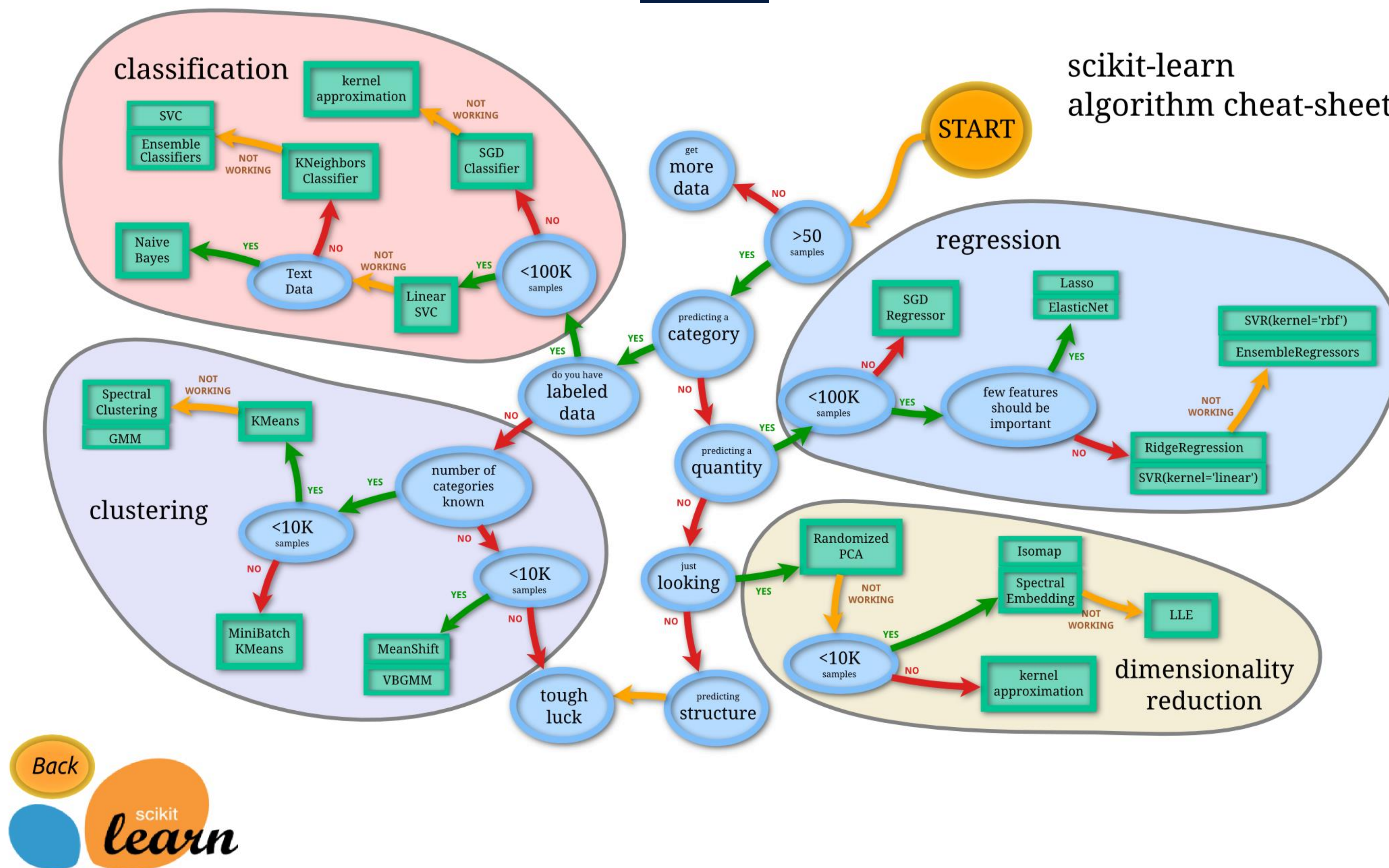
Illustration in Python

(cf notebook)

Types of learning



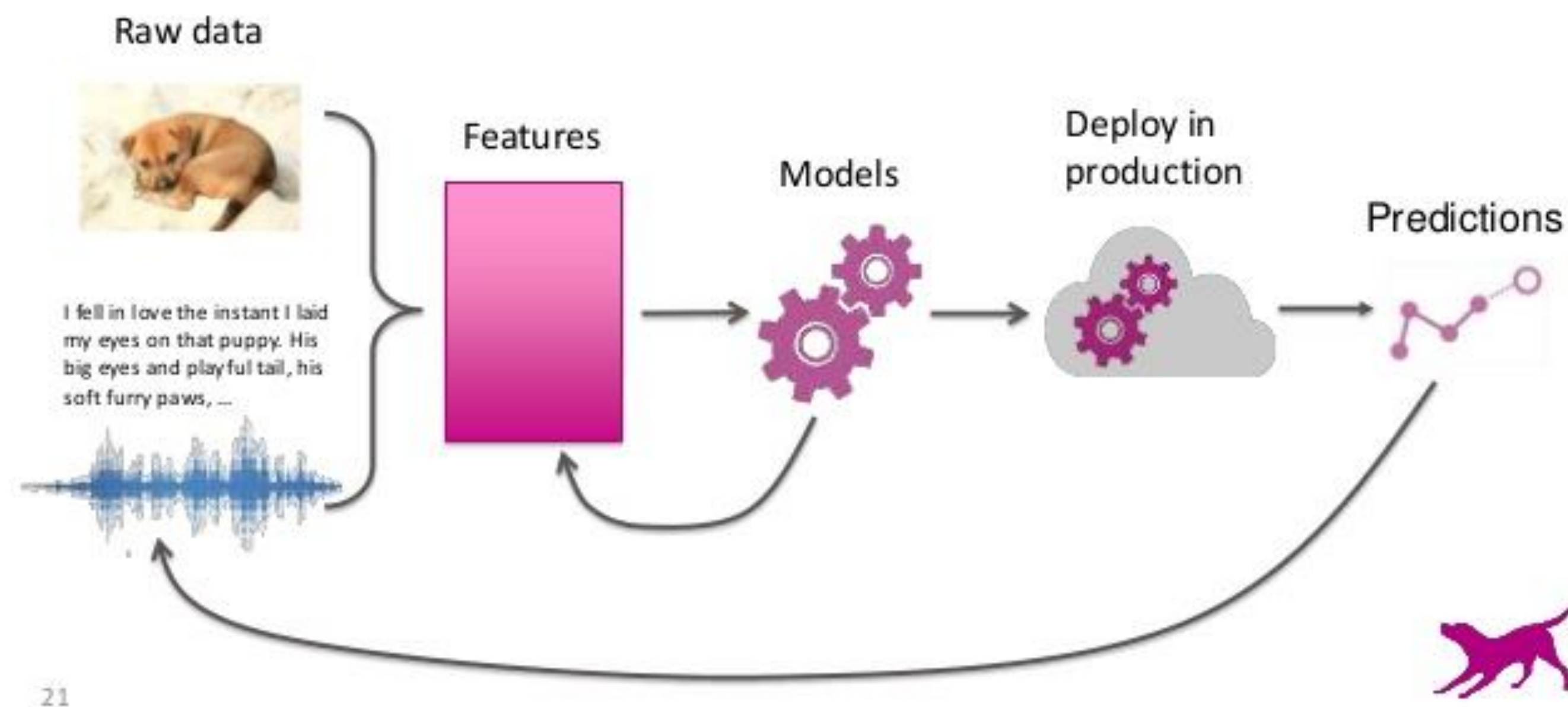
Choosing the right model



Final (important) words

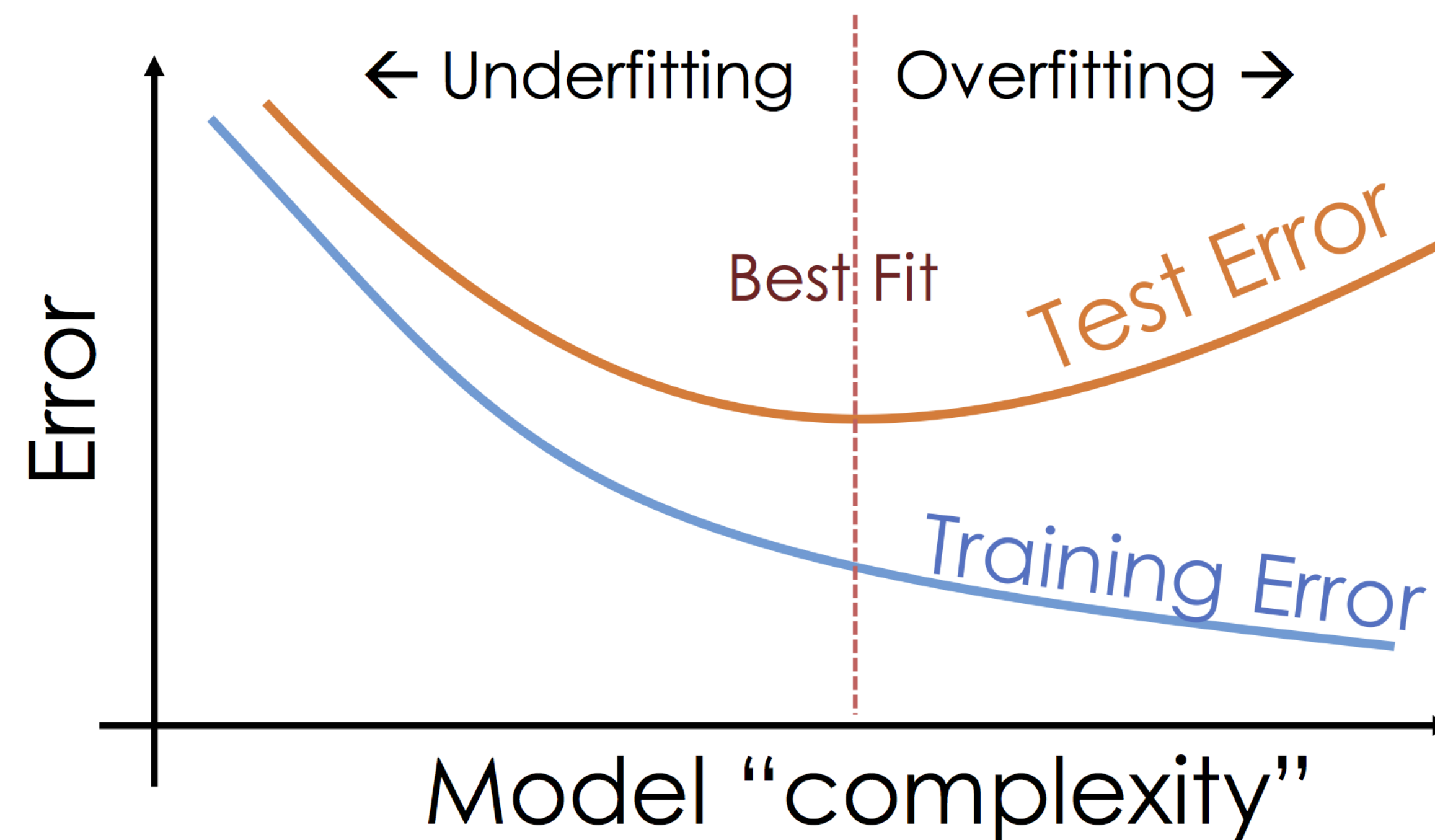
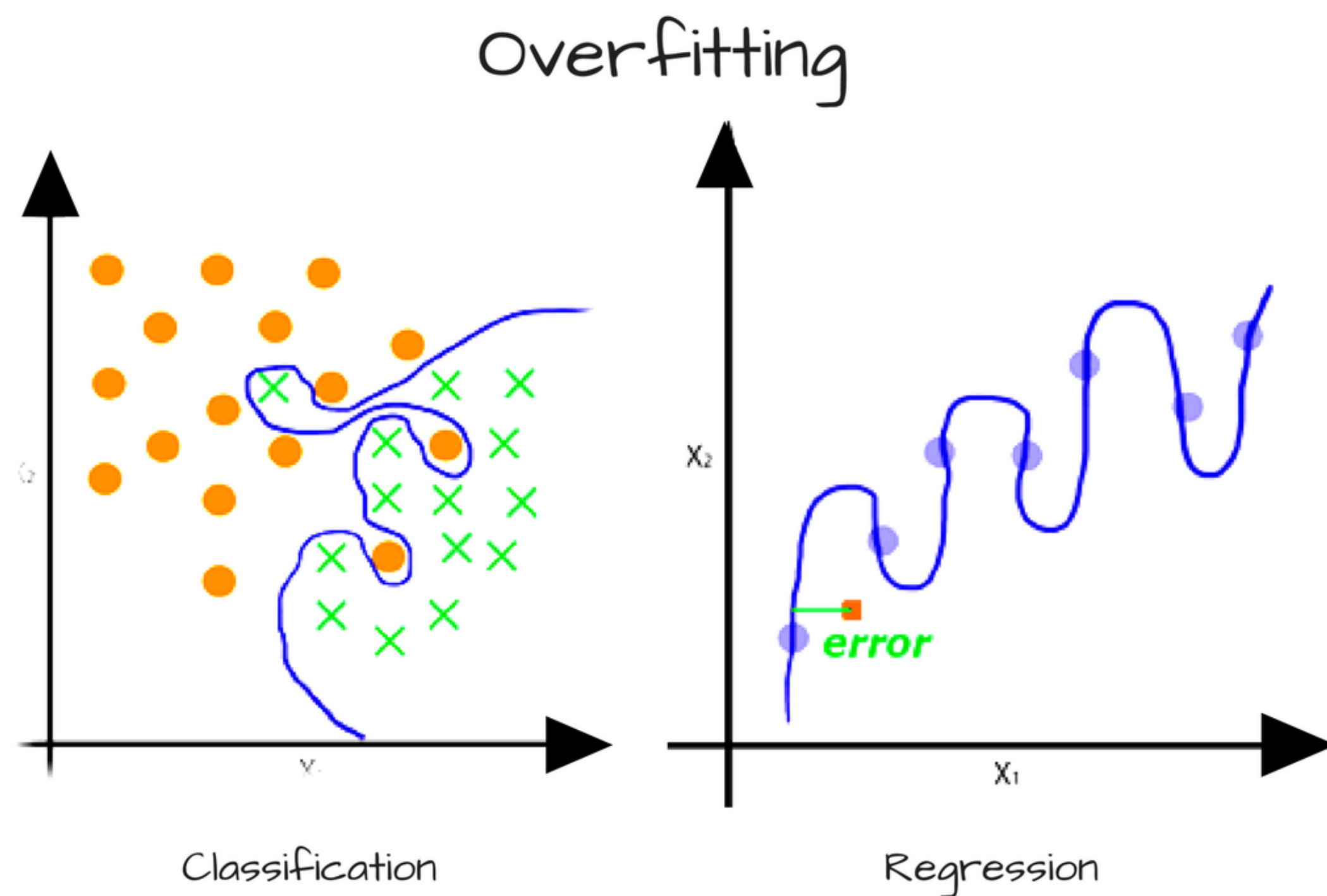
Feature engineering & selection

The machine learning pipeline



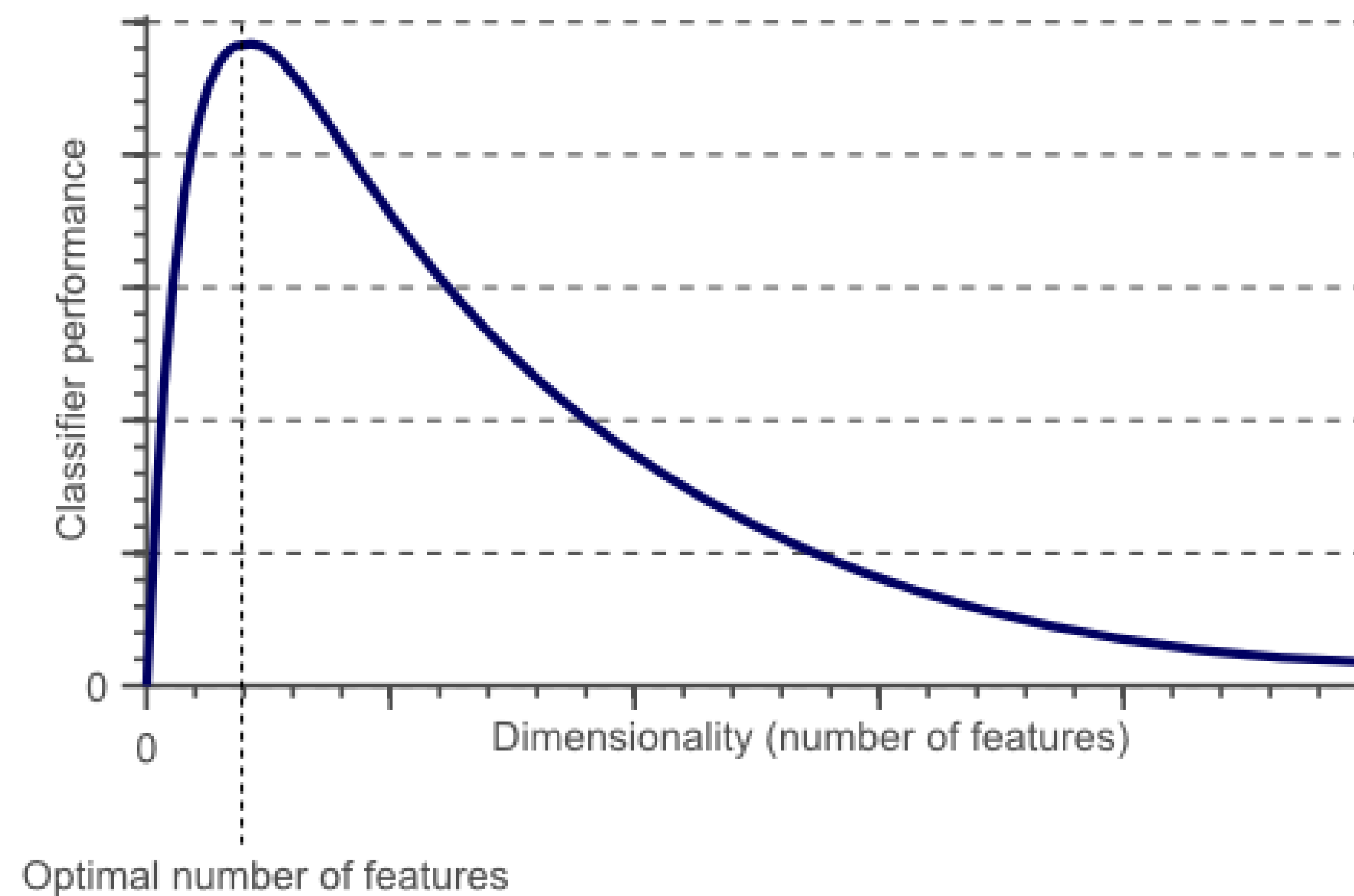
Final (important) words

Bias-Variance trade-off



Final (important) words

Curse of dimensionality



QUESTIONS ?

