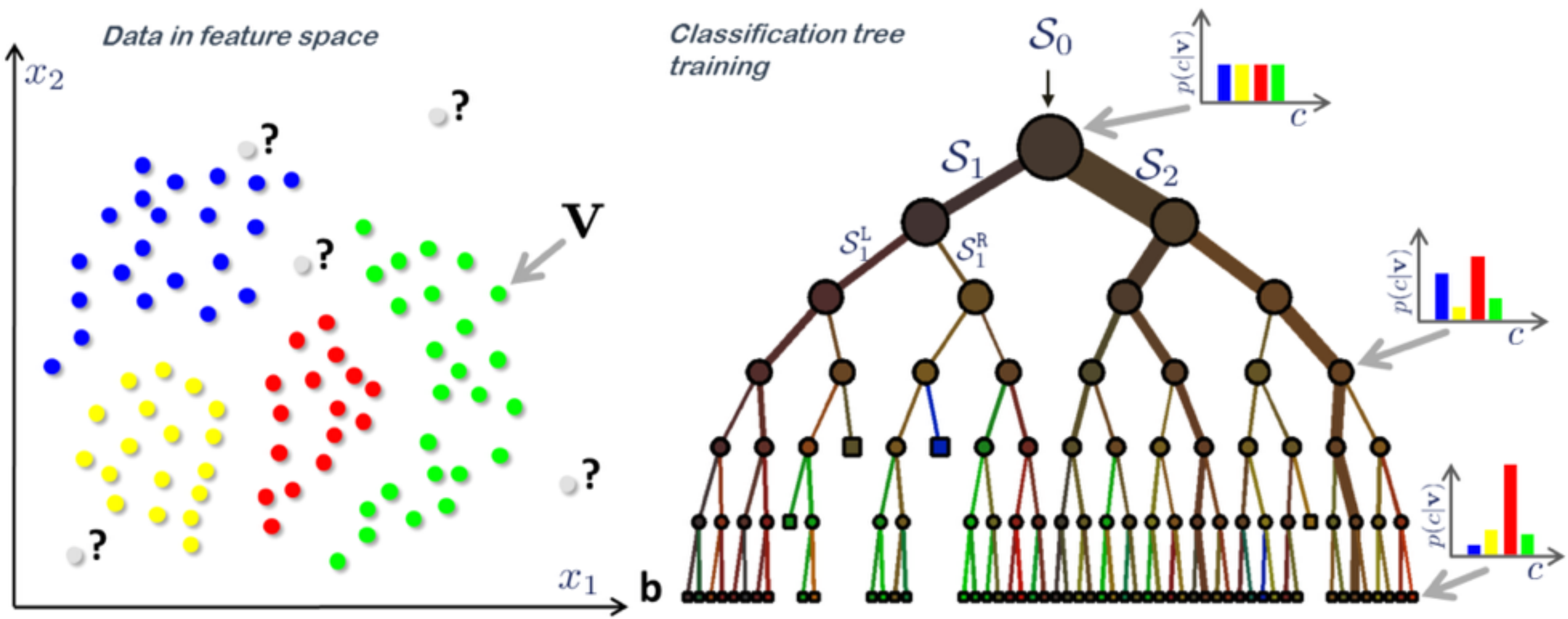
Random Forests

Sourav Sen Gupta

CDS 2015 | PGDBA | 9 Oct 2015

Classification Tree



Ref. — “Decision Forests” — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Information Gain

Information Gain = Parent Entropy — E(Child Entropy)

$$I = H(\mathcal{S}) - \sum_{i \in \{\mathtt{L},\mathtt{R}\}} \frac{|\mathcal{S}^i|}{|\mathcal{S}|} H(\mathcal{S}^i)$$

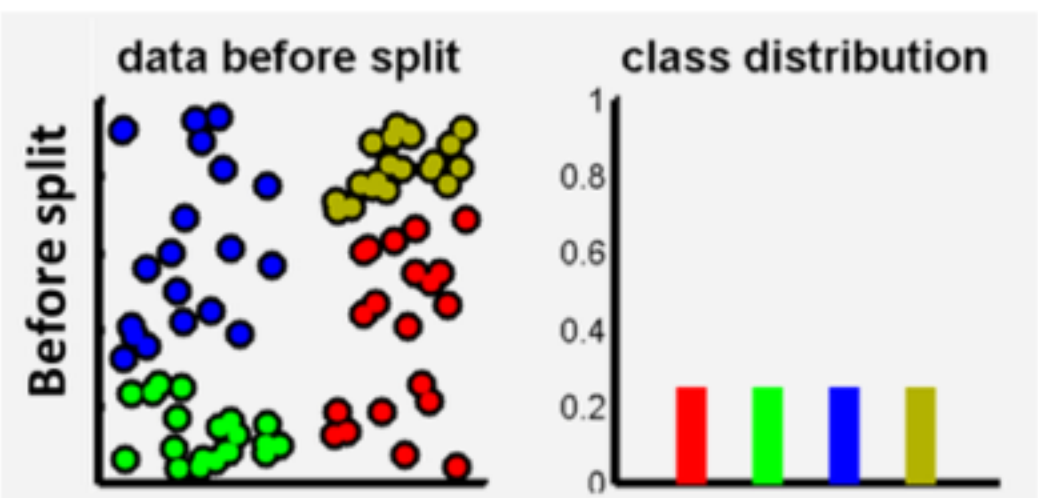
One notion of entropy is that of Shannon Entropy

$$H(\mathcal{S}) = -\sum_{c \in \mathcal{C}} p(c) \log(p(c))$$

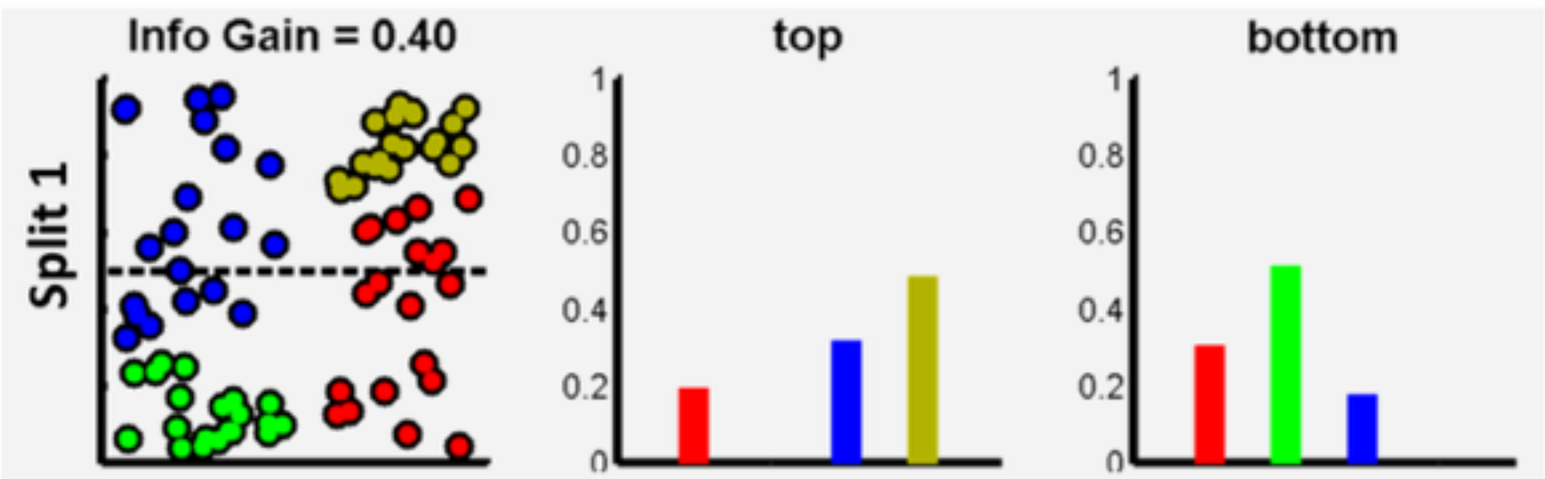
Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Choosing Split

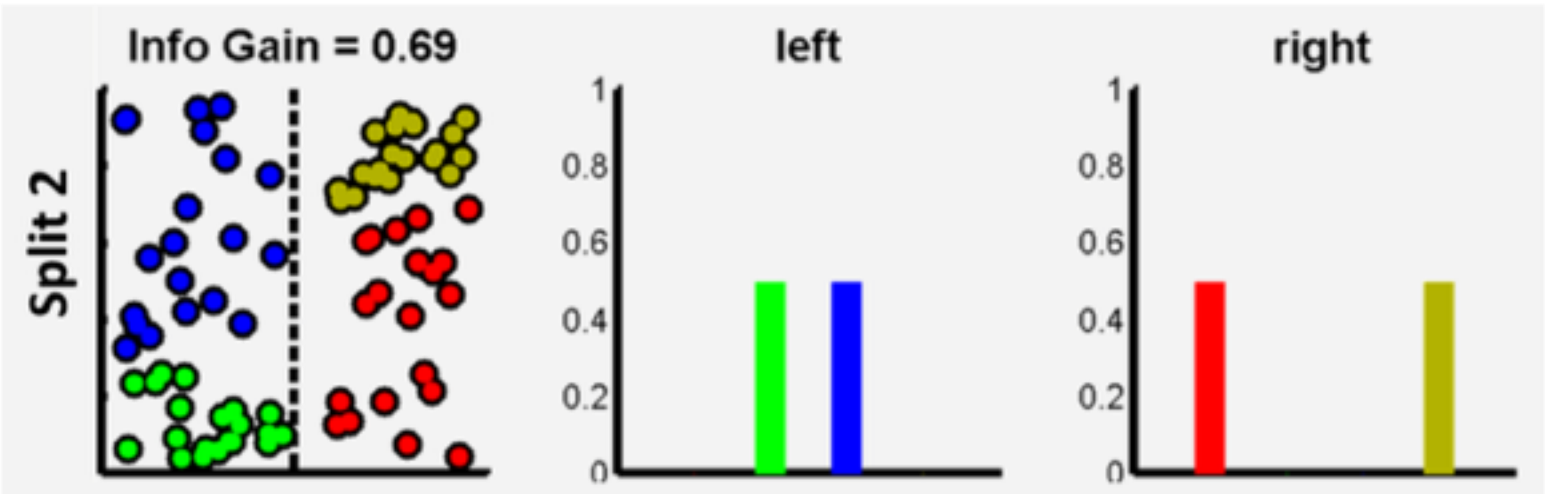
$$\boldsymbol{\theta}_j^* = \arg\max_{\boldsymbol{\theta}_j \in \mathcal{T}_j} I_j$$



(a)



(b)



(c)

Ref. — “Decision Forests” — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Random Tree

First Source of Randomness : Training Features

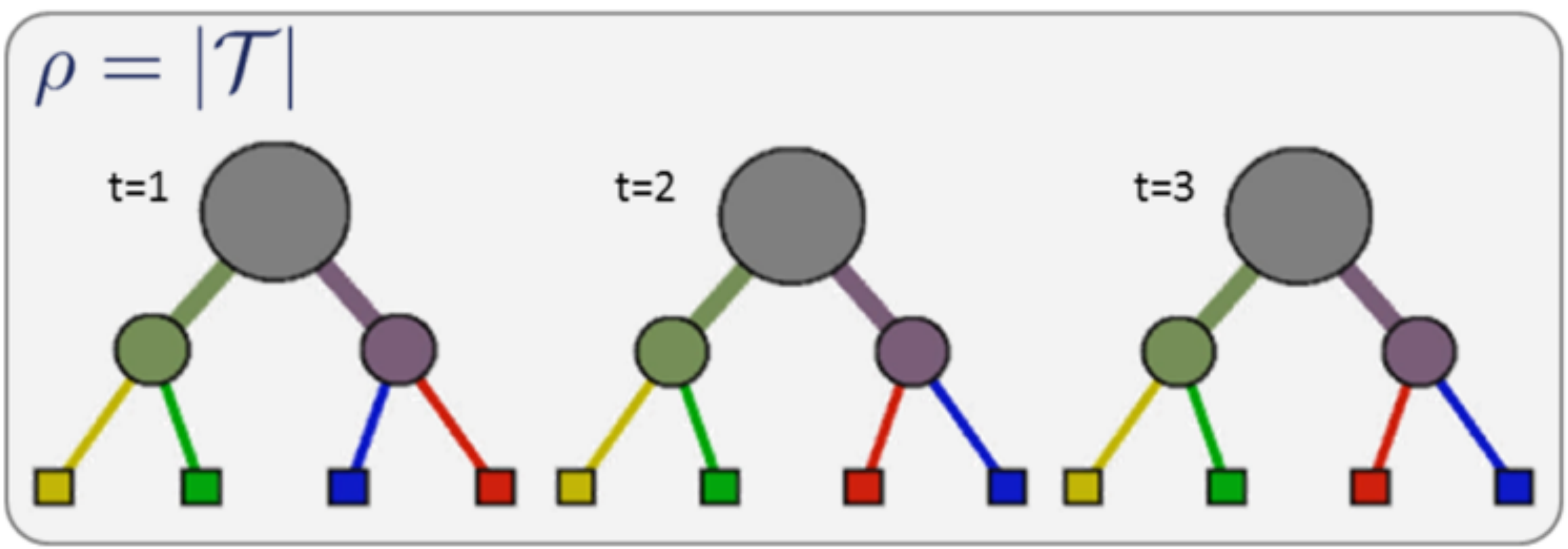
Option 1 : Choose all features for splitting every node

Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Random Tree

First Source of Randomness : Training Features

Option 1 : Choose all features for splitting every node



Low randomness, high tree correlation

Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Random Tree

First Source of Randomness : Training Features

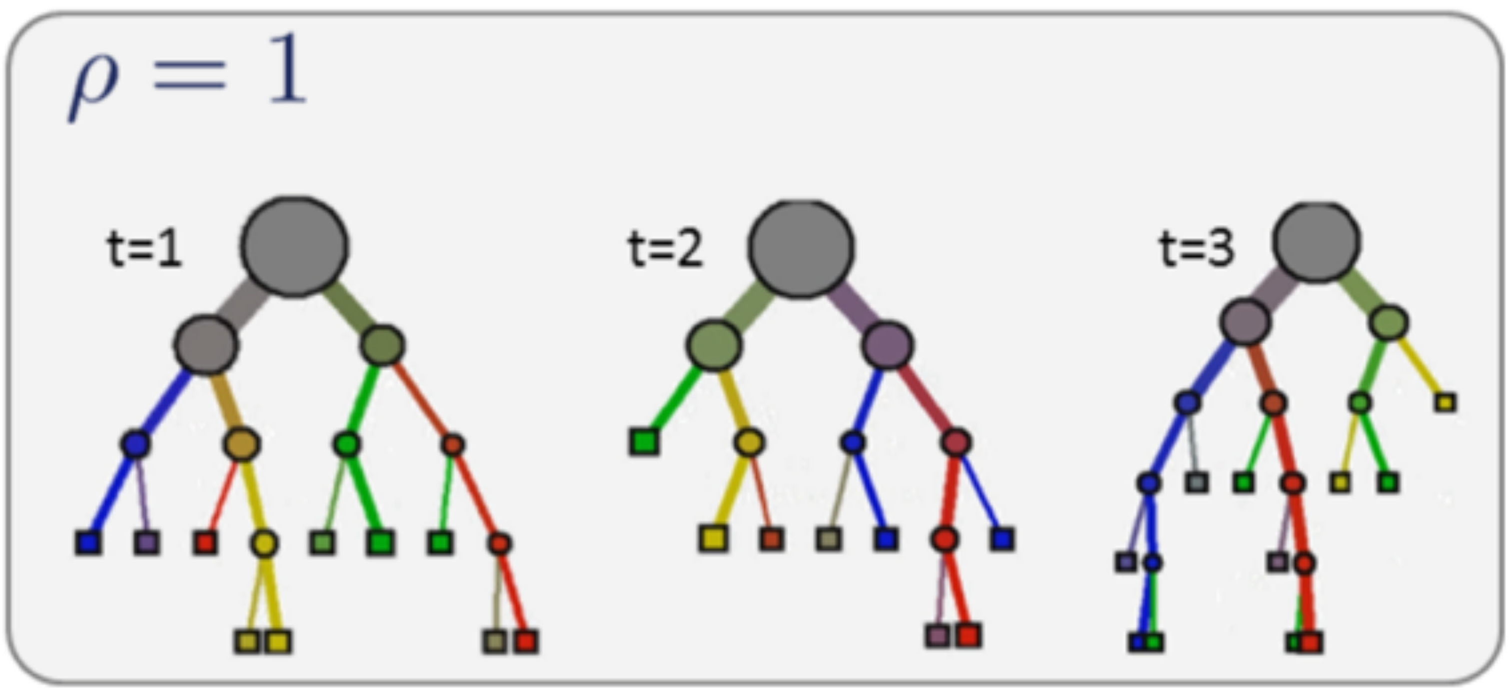
Option 2 : Choose one feature for splitting every node

Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Random Tree

First Source of Randomness : Training Features

Option 2 : Choose one feature for splitting every node



High randomness, low tree correlation

Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Random Tree

Second Source of Randomness : Training Data

Option 1 : Choose all observations to train each tree

Option 2 : Choose “some” observations for training

Independent subsets ensure independent training
Bootstrapping — Drawing random subsets of data

Central Idea: **B**ootstrap **agg**regat**ing** (Bagging)

Ref. — “Decision Forests” — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

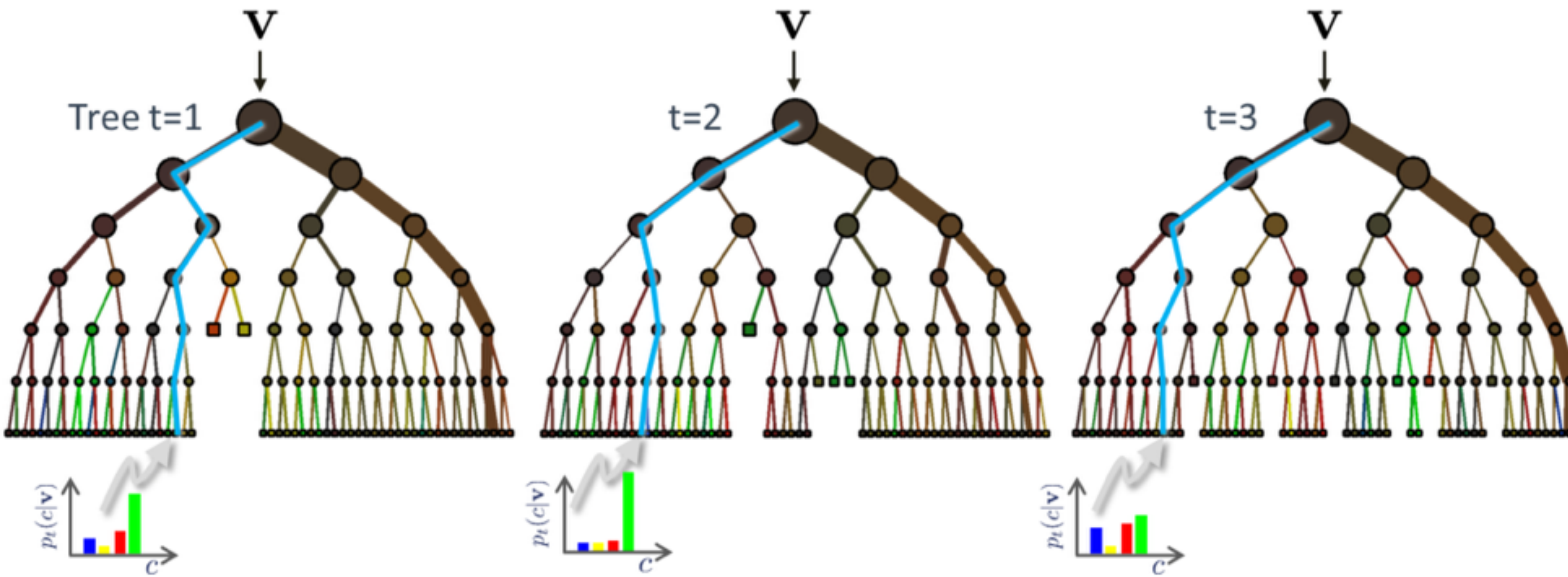
Random Forest

Sources of Randomness : Training Data and Features

1. For $b = 1$ to B:
   (a) Draw a bootstrap sample $\mathbf{Z}^*$ of size N from the training data.
   (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
      i. Select m variables at random from the p variables.
      ii. Pick the best variable/split-point among the m.
      iii. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$.

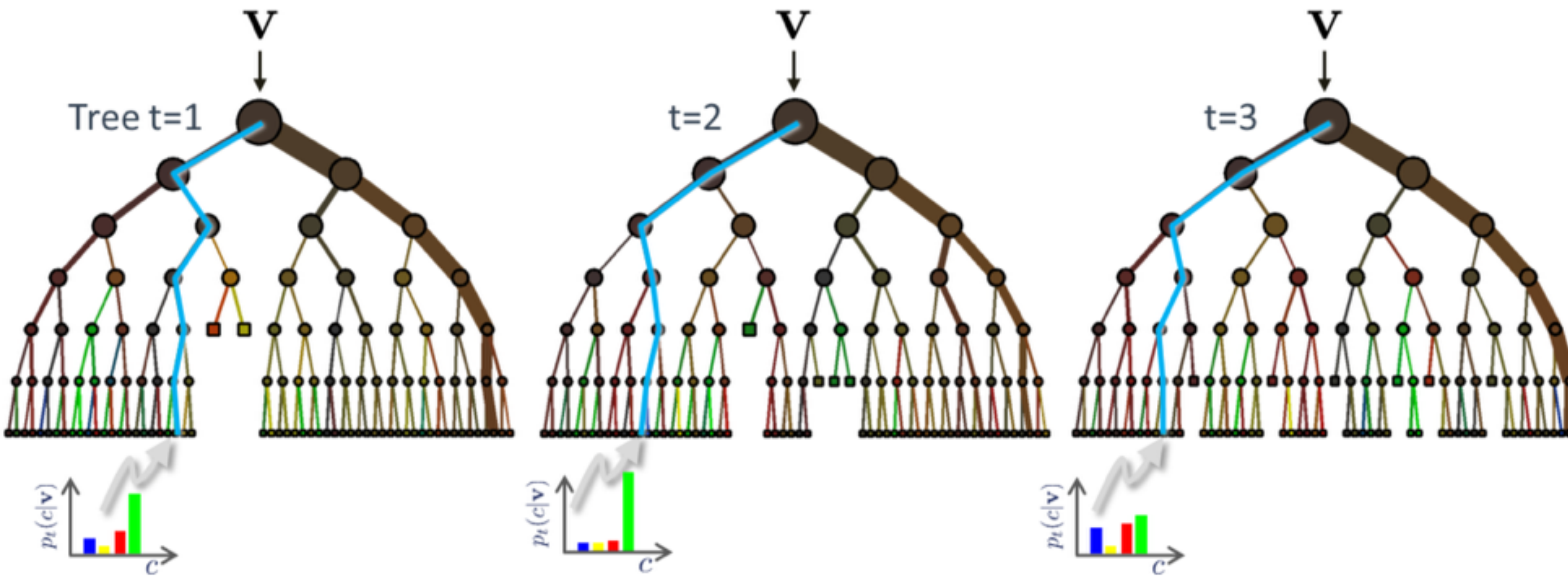
Ref. — “The Elements of Statistical Learning” — Trevor Hastie, Robert Tibshirani, Jerome Friedman

Random Forest



Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Random Forest



Aggregate the outputs of individual decision trees:

$$p(c|\mathbf{v}) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|\mathbf{v})$$

Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Important Parameters

The size of the forest, i.e., number of trees

The maximum allowed depth for each tree

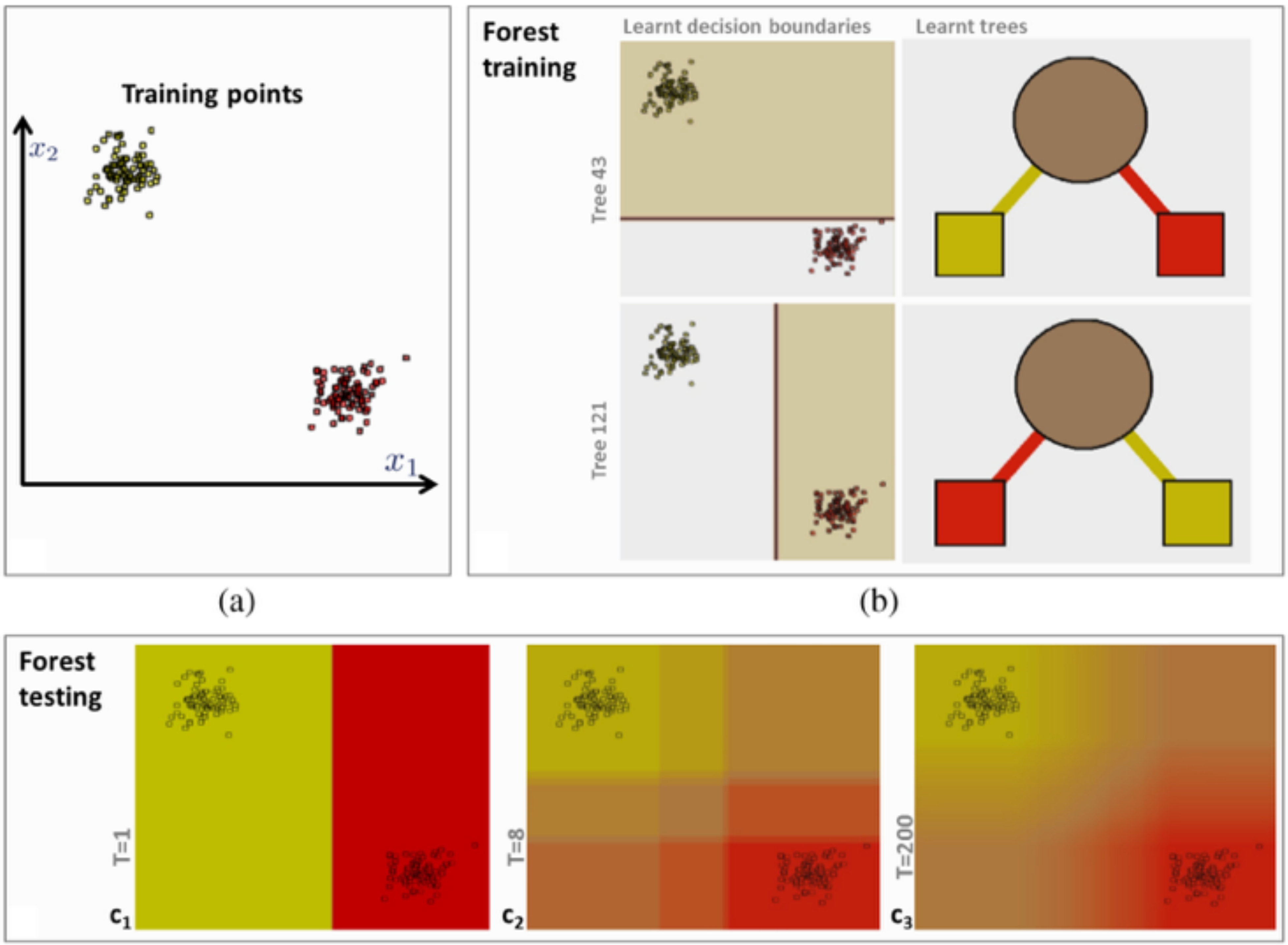
The amount of randomness between trees

The weak learner model used for each tree

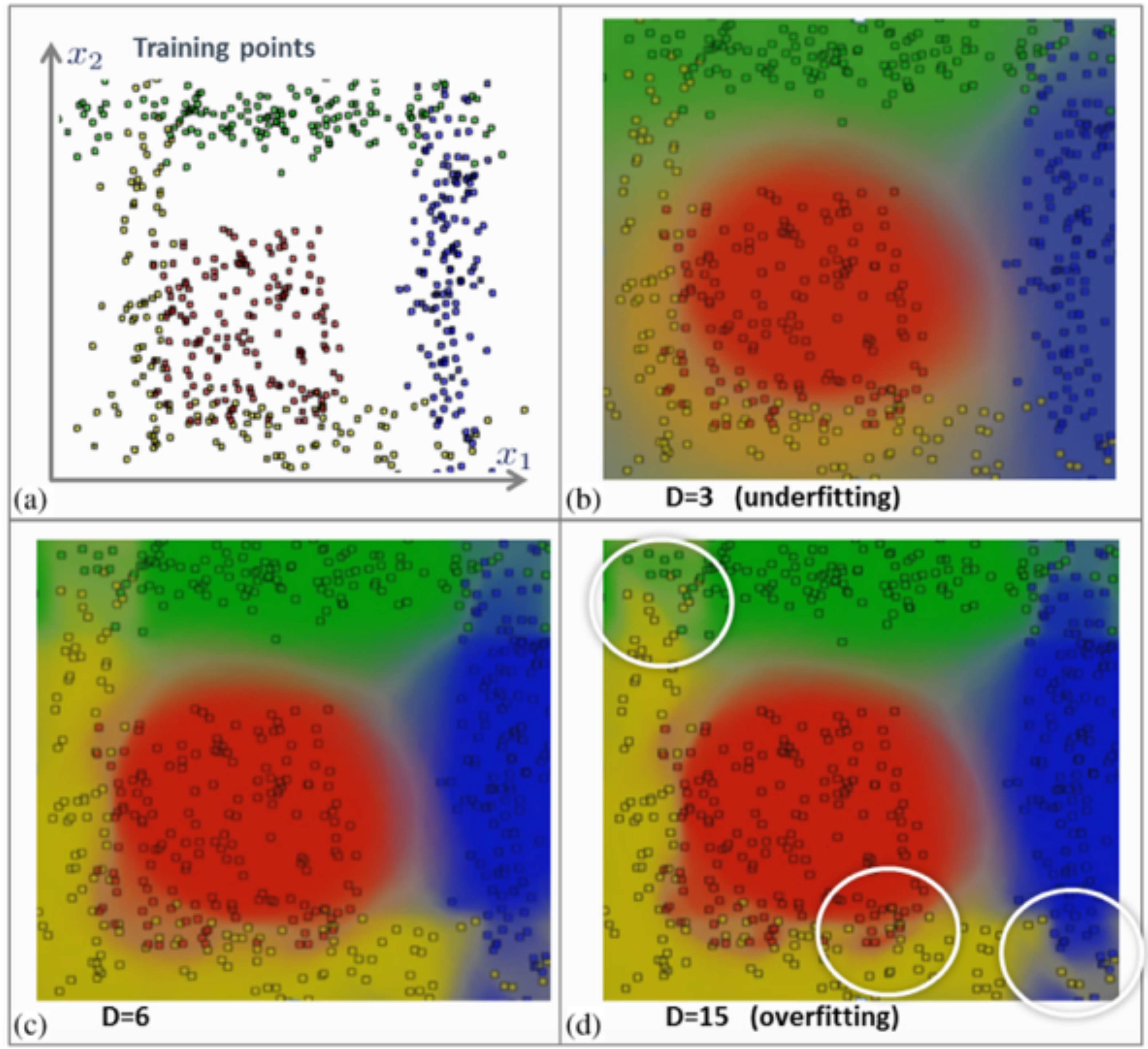
The training objective function for each tree

Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

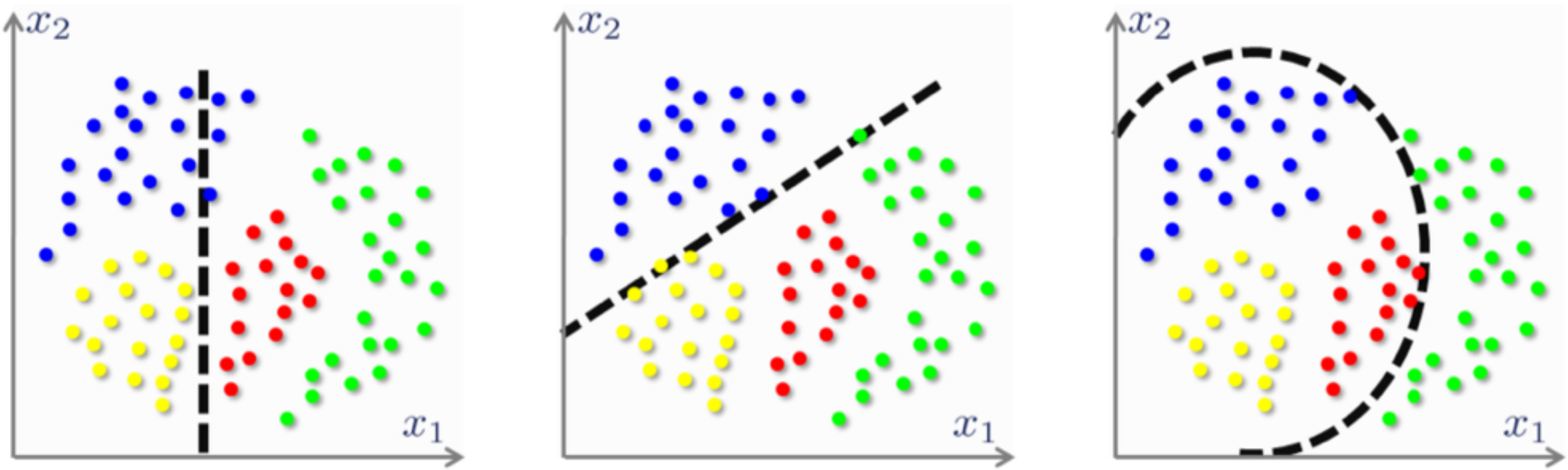
Effect of Forest Size



Effect of Tree Depth



Weak Learner Model



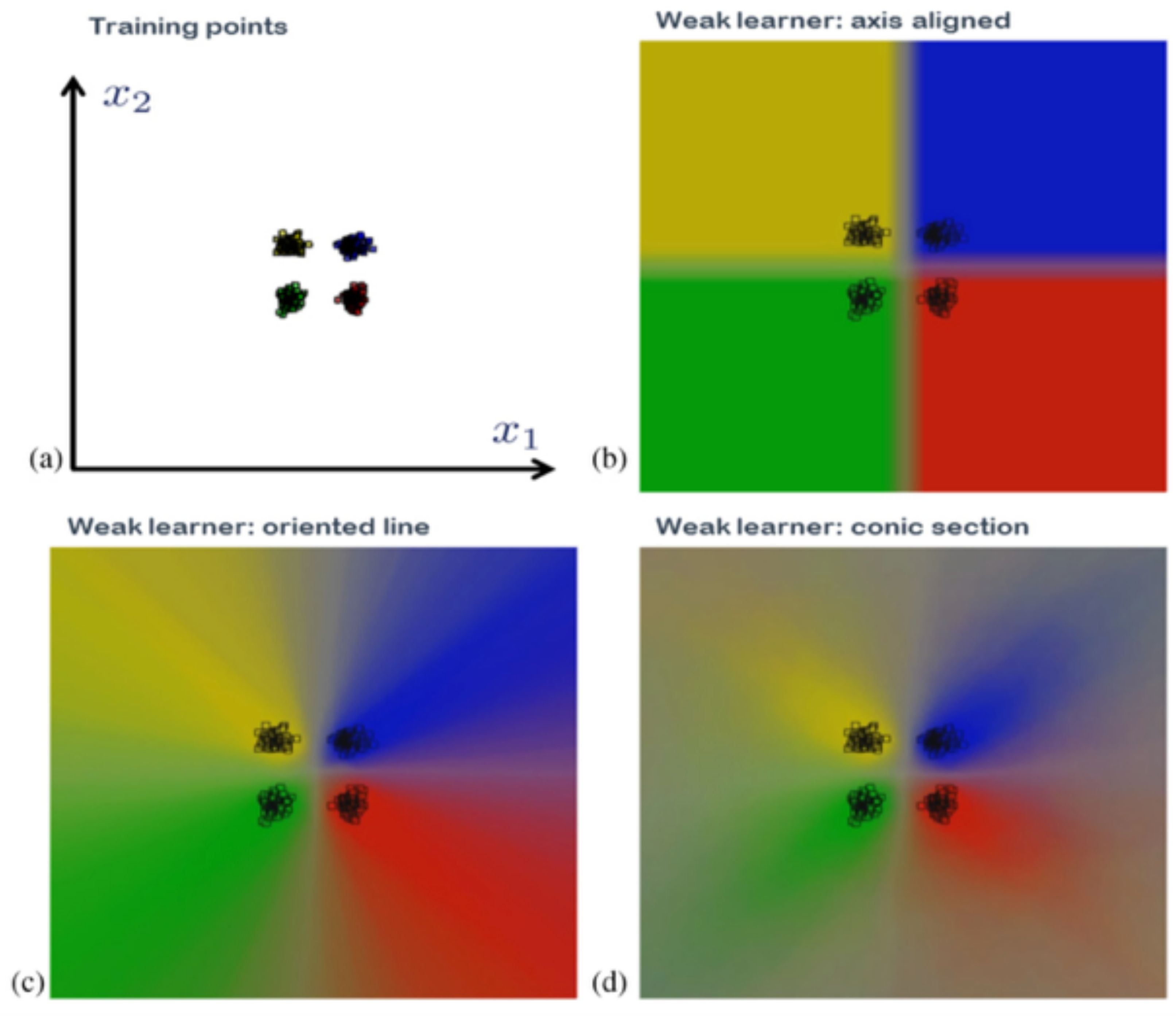
(a) Axis-aligned Hyperplane

(b) General oriented Hyperplane

(c) Quadratic/Conic in 2D

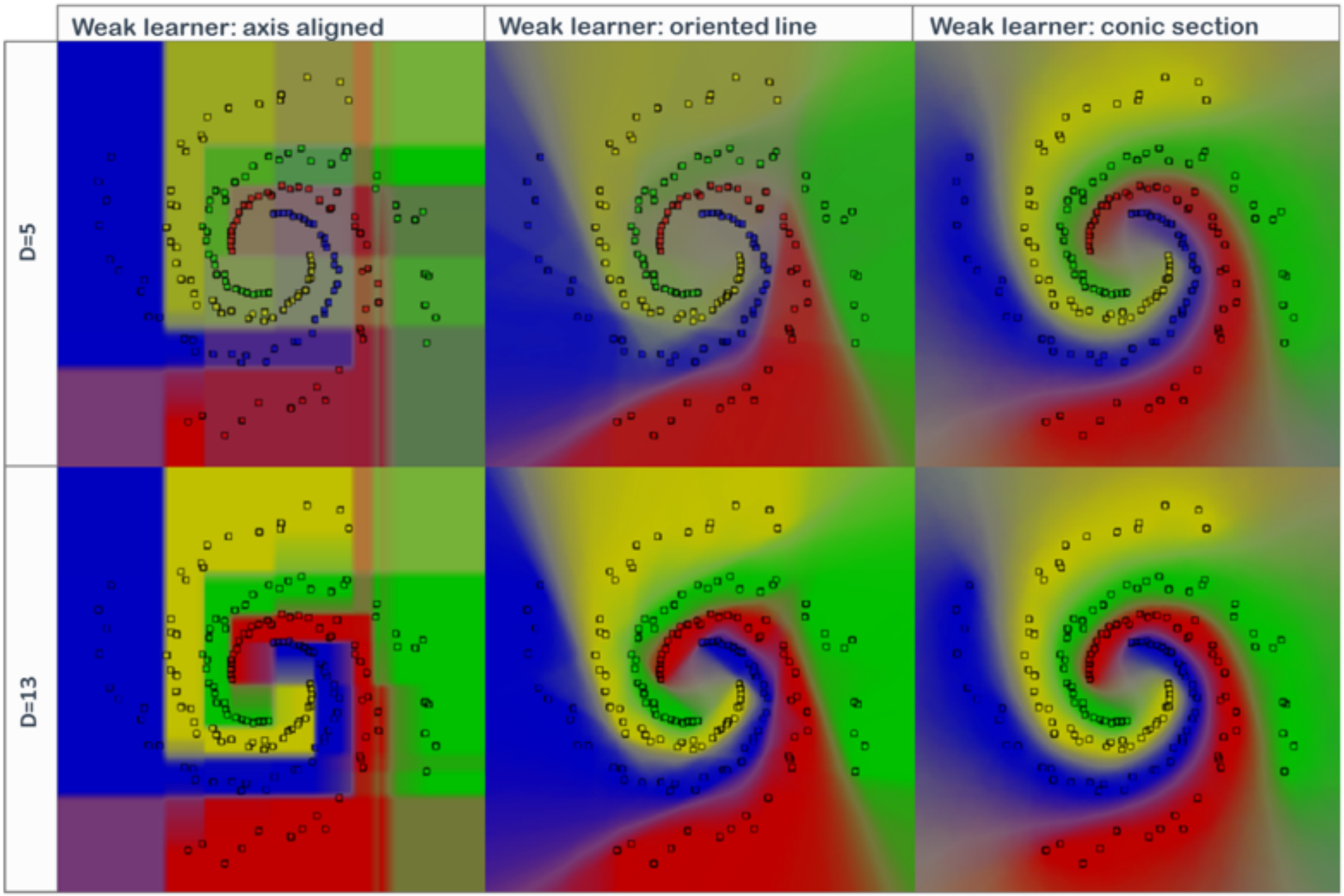
Ref. — “Decision Forests” — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Effect of Weak Learner



Ref. — “Decision Forests” — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Effect of Weak Learner



Ref. — "Decision Forests" — Antonio Criminisi, Jamie Shotton, and Ender Konukoglu

Important Parameters

The size of the forest, i.e., number of trees
The maximum allowed depth for each tree
The amount of randomness between trees
The weak learner model used for each tree
The training objective function for each tree

Typical choices for randomness:

Bootstrapping for random sampling of observations
Choosing m = [sqrt(p)] number of random features
Minimum allowed node size is one for classification

Ref. — "The Elements of Statistical Learning" — Trevor Hastie, Robert Tibshirani, Jerome Friedman