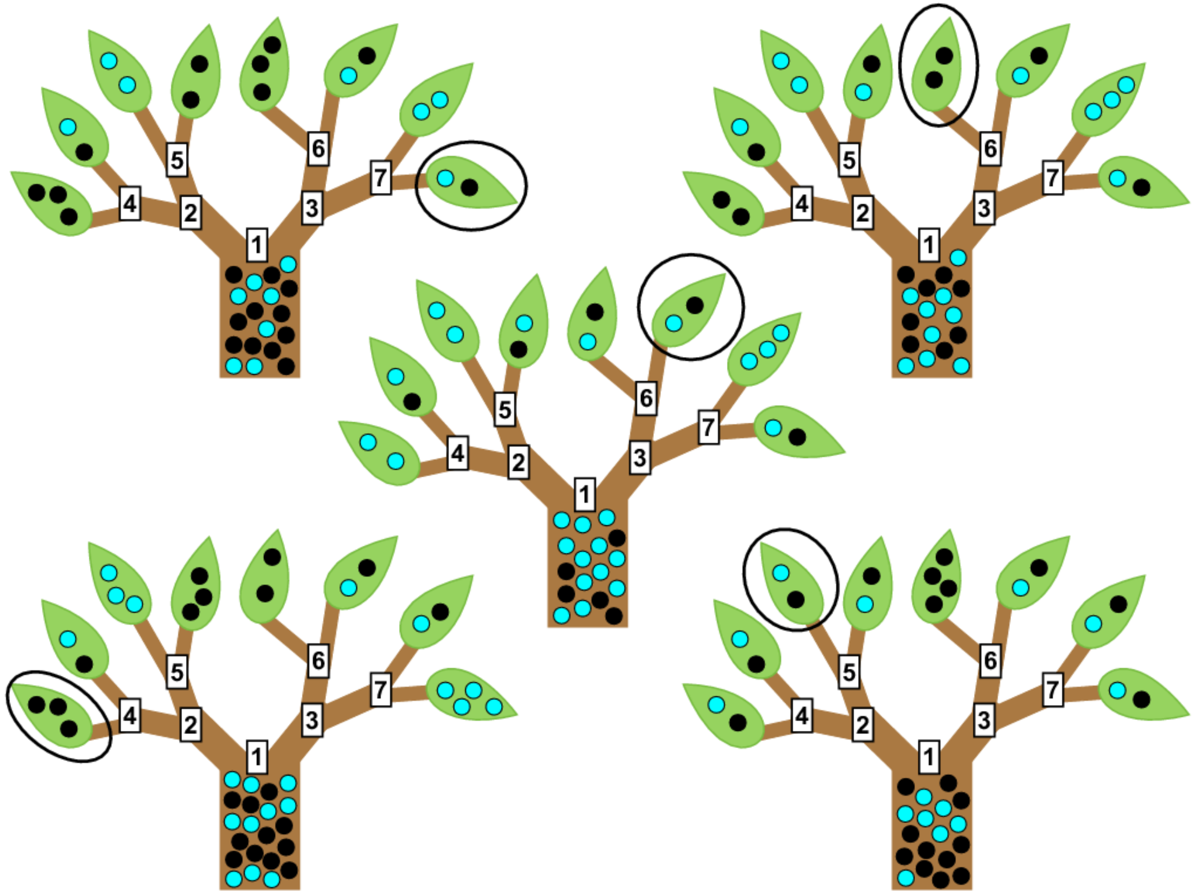
Random Forest



MAC 5739

Artificial Intelligence

Jun 2018

IME – USP

Author: Roberto Shimizu

Contents

- Random Forest at a Glance
- Decision Trees Recap
- Random Forest Model
- Examples of Applications
- Bibliography and Other Sources of Information

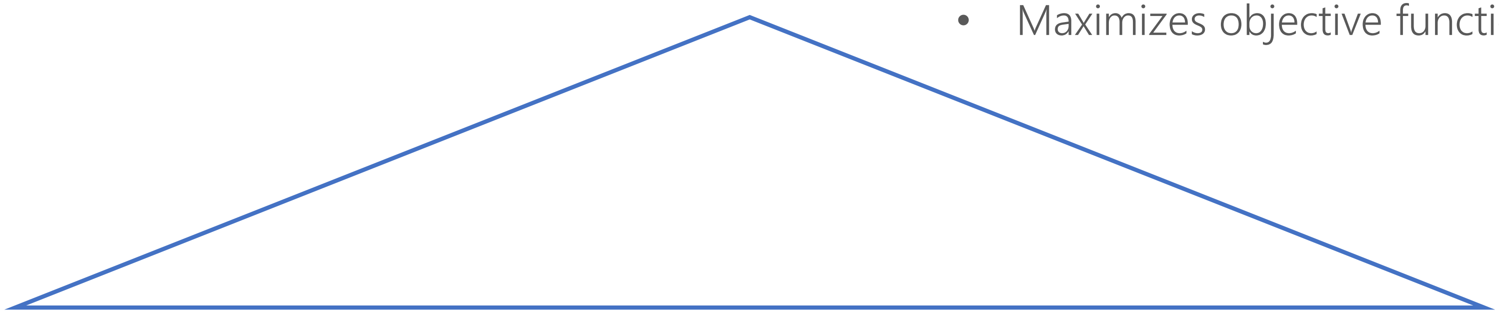
Random Forest at a Glance

Key facts

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the collective prediction of all the individual trees.

Decision Trees

- Maximizes objective function (e.g. information gain)



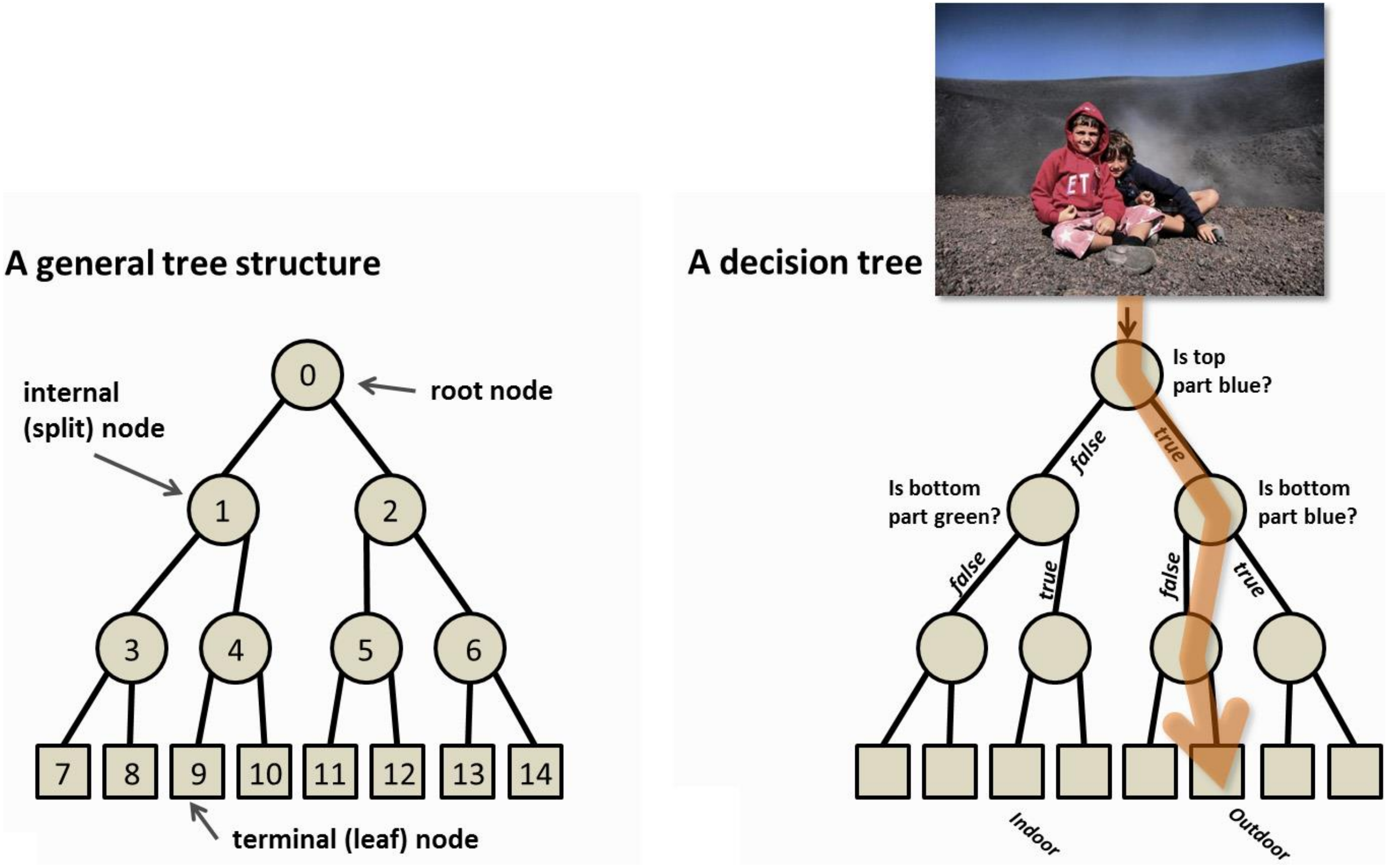
The Randomness Model

- random training set sampling (e.g. bootstrapping, bagging)
- randomized node optimization

Ensembles of Trees (Forest)

- committee of weak learners (e.g. boosting, aggregation)

Decision Tree

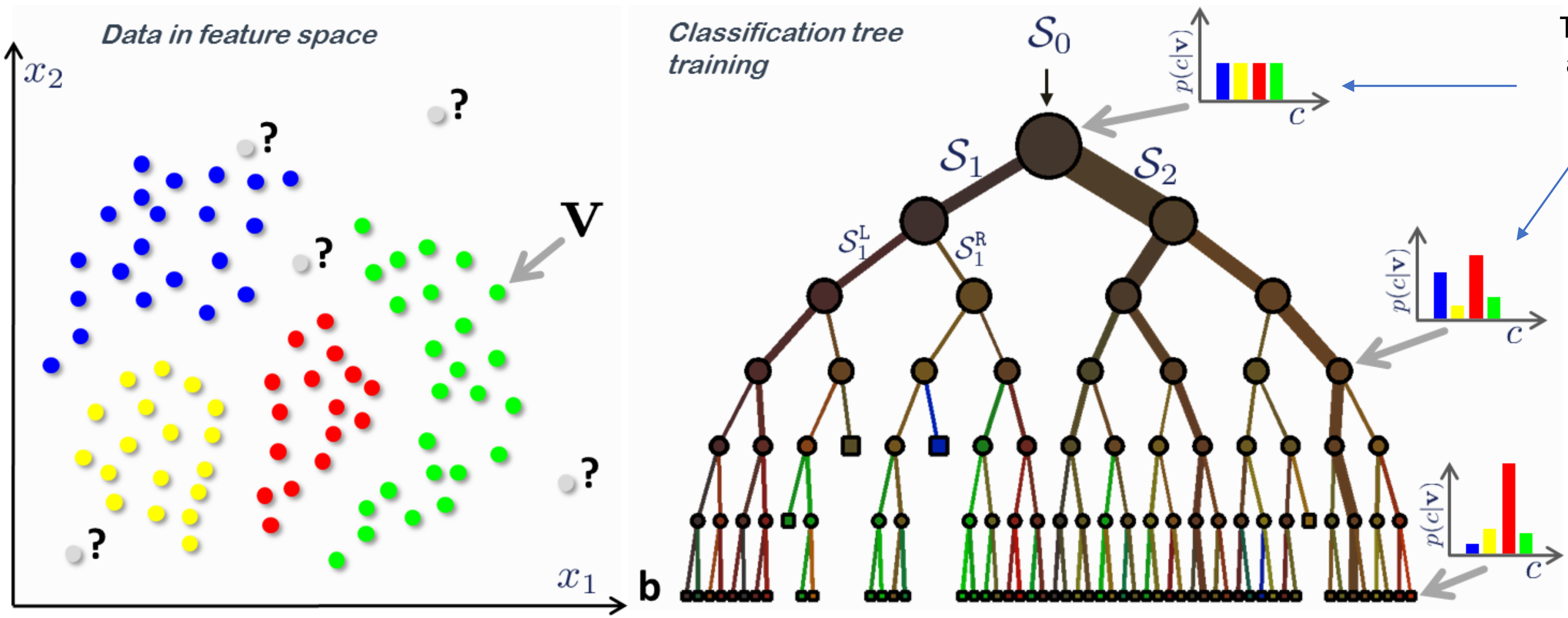


(a)

(b)

A. Criminisi, J. Shotton and E. Konukoglu (2012). *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*- pp 88.

Decision Tree Classifier



(a)

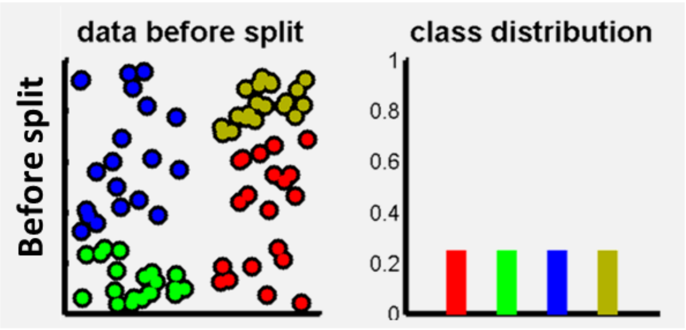
(b)

Test Functions/Split Functions also called **"Weak Learners"**

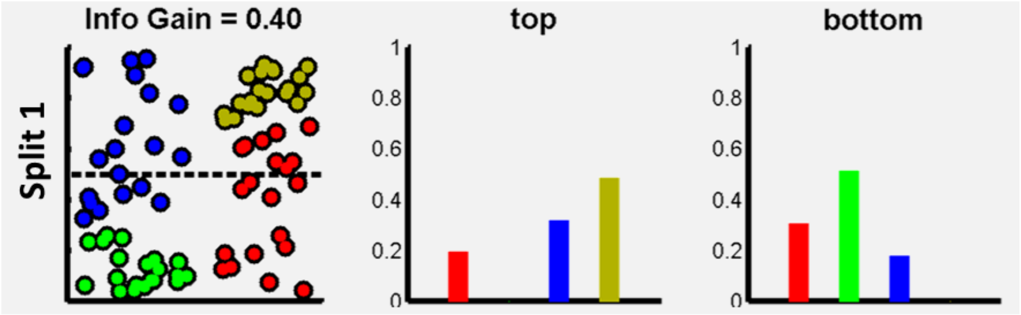
Leaf Statistics gathered in training to predict the label associated with an unseen point $p(c|\boldsymbol{v})$

A. Criminisi, J. Shotton and E. Konukoglu (2012). *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*- pp 107.

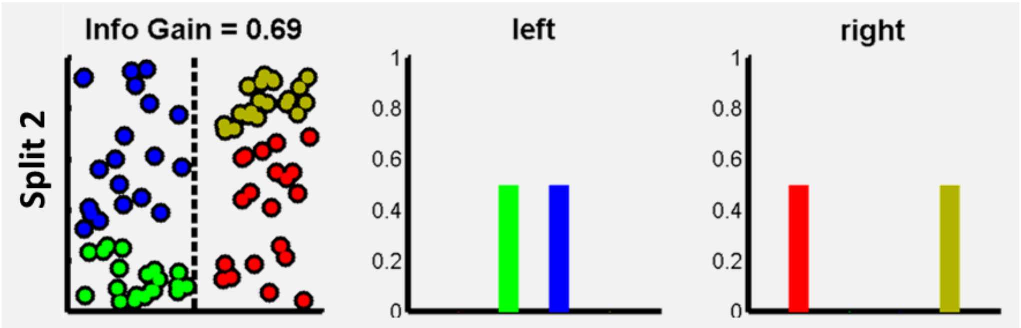
Objective Function



(a)



(b)



(c)

Entropy

Higher

Lower

Entropy

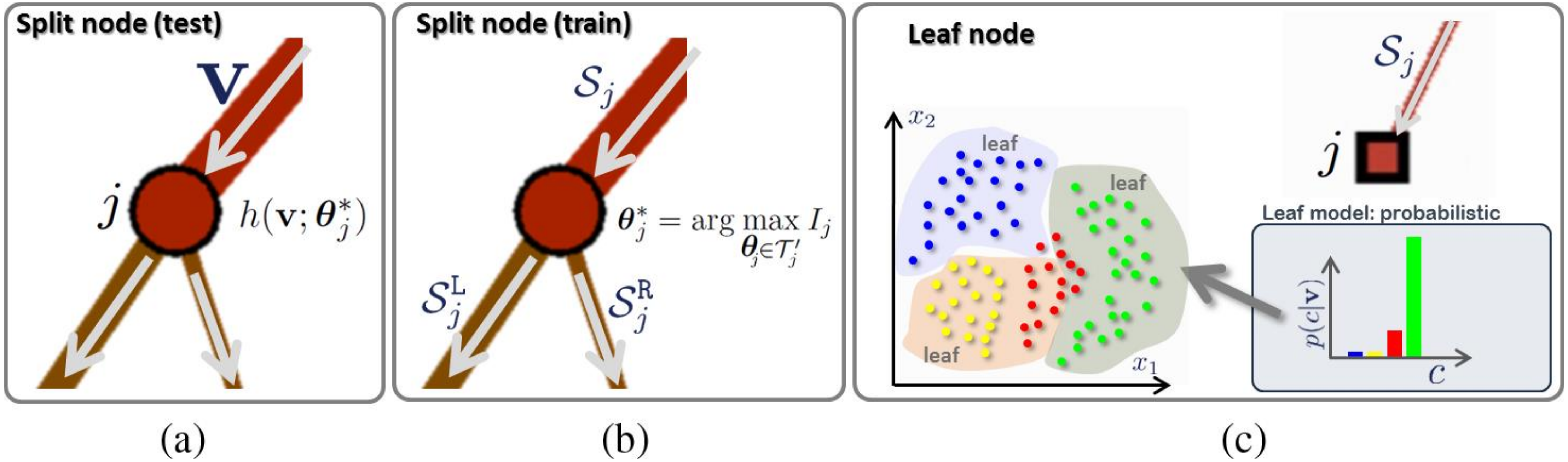
$$H(S) = -\sum_{c \in C} p(c) \log(p(c))$$

Information Gain

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

A. Criminisi, J. Shotton and E. Konukoglu (2012). *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*- pp 97.

Decision Tree Optimization Algorithm


Split node (test)
$\mathbf{v}$
j
$h(\mathbf{v};\boldsymbol{\theta}_j^*)$
Split node (train)
$\mathcal{S}_j$
$\boldsymbol{\theta}_j^* = \arg\max_{\boldsymbol{\theta}_j \in \mathcal{T}_j'} I_j$
$\mathcal{S}_j^{\mathrm{L}}$
$\mathcal{S}_j^{\mathrm{R}}$
Leaf node
x_2
leaf
leaf
leaf
x_1
$\mathcal{S}_j$
j
Leaf model: probabilistic
$p(c|\mathbf{v})$
c
(a)
(b)
(c)

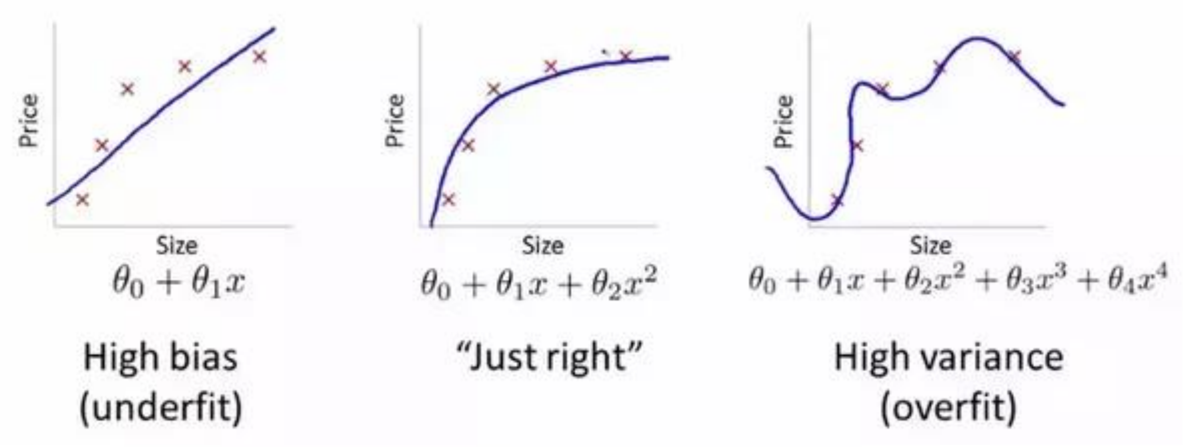
(b) Split node (training). Training the parameters θ_j of node j involves optimizing a chosen objective function (maximizing the information gain I_j)

Here, the optimization is done with respect to the entire parameter space τ.

A. Criminisi, J. Shotton and E. Konukoglu (2012). *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*- pp 91.

Main Characteristics of Decision Tree Classifier

- Model tend to be easily understood by humans, as opposed to “black box” models such as Neural Networks (NN) and Support Vector Machines (SVM)
- Application is limited to relatively low dimensional data, whereas most of Machine Learning application are devoted to high dimensional problems.
- When trees are grown very deep, it tends to learn highly irregular patterns, i.e., it overfits the training data.
- Deep decision trees have therefore **Low Bias** and **High Variance**.

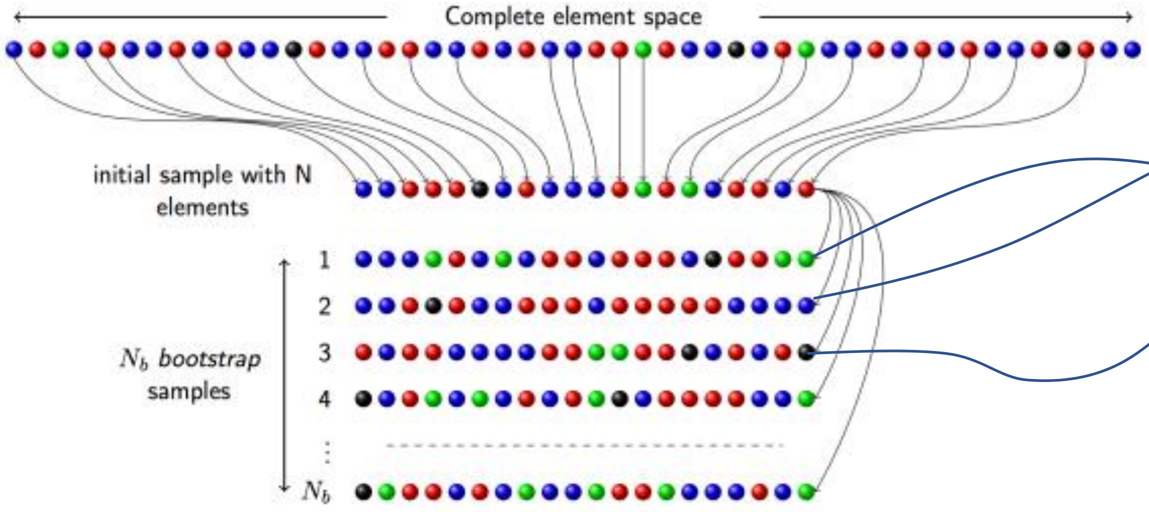


The **Bias** is an error from erroneous assumptions in the learning algorithm. **High bias** can cause an algorithm to miss the relevant relations between features and target outputs (**underfitting**).

The **Variance** is an error from sensitivity to small fluctuations in the training set. **High variance** can cause an algorithm to model the random noise in the training data, rather than the intended outputs (overfitting).

Random Forest: Ensemble of Random Trees

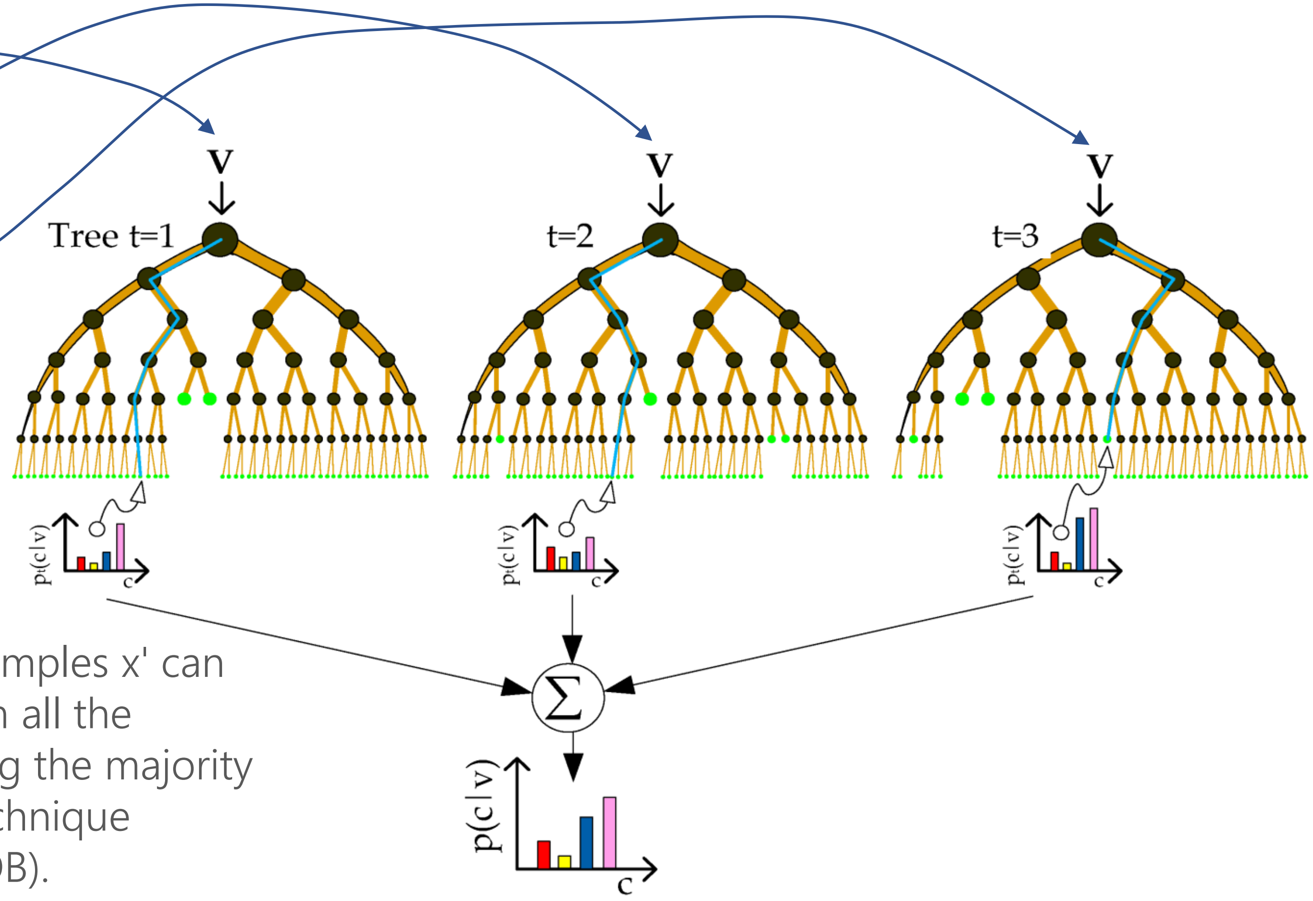
1 - random training set sampling: bootstrapping or bagging



2 - randomized node optimization: select randomly a subset of features (parameters) in each node:

$\theta_j^* = \arg\max I_{j\theta\in\tau_j}$

Ammount of Randomness is controlled by ratio $\frac{\rho}{|\tau|}$ where $\rho = 1, \dots, |\tau|$



3 - After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x' or by taking the majority vote in the case of classification trees. A technique commonly used is Out-of-Bag Sample (OOB).

Random Forest Algorithm

Algorithm 15.1 *Random Forest for Regression or Classification.*

```
1. For $b = 1$ to $B$:
    (a) Draw a bootstrap sample $Z^*$ 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 $\boldsymbol{m}$ variables at random from the $\boldsymbol{p}$ variables.
            ii. Pick the best variable/split-point among the $\boldsymbol{m}$.
            iii. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$.
To make a prediction at a new point $x$:
*Regression*: $f_{rf}^B(x) = \frac{1}{B}\sum_{b=1}^B T_b(x)$.
*Classification*: Let $C_b(x)$ be the class prediction of the $b_{th}$ random-forest
    tree. Then $C_{rf}^B(x) = majority\ vote\ \{C_b(x)\}_1^B$.
```

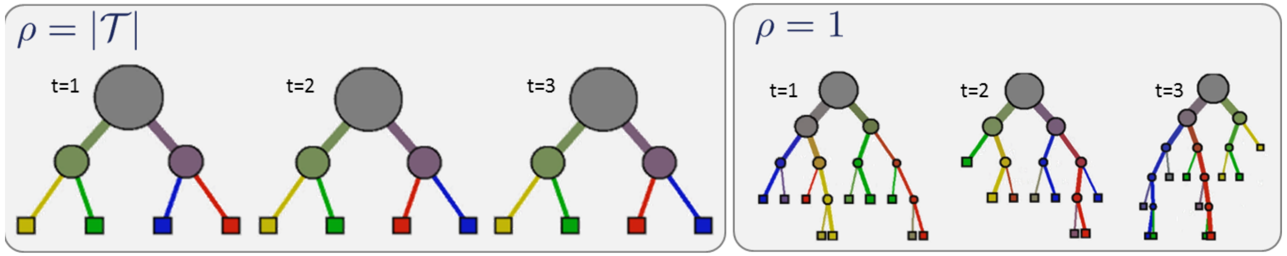
Hastie, T., Tibshirani, R. and Friedman, J. (2008). *The Elements of Statistical Learning* - 2^{nd} Edition – Springer – Chapter 15: 588.

Key Model Parameters

- The size N of bootstrap sample Z^*
- The maximum allowed tree Depth D
- The amount of randomness (controlled by ρ)
    - Typically values for $\boldsymbol{m}$ are $\sqrt{\boldsymbol{p}}$ or even as low as 1
- The forest size (number of trees $\boldsymbol{B}$)
    - Typical size: it depends.

Main Characteristics of Random Forest Classifier

- Decorrelation between the individual tree predictions



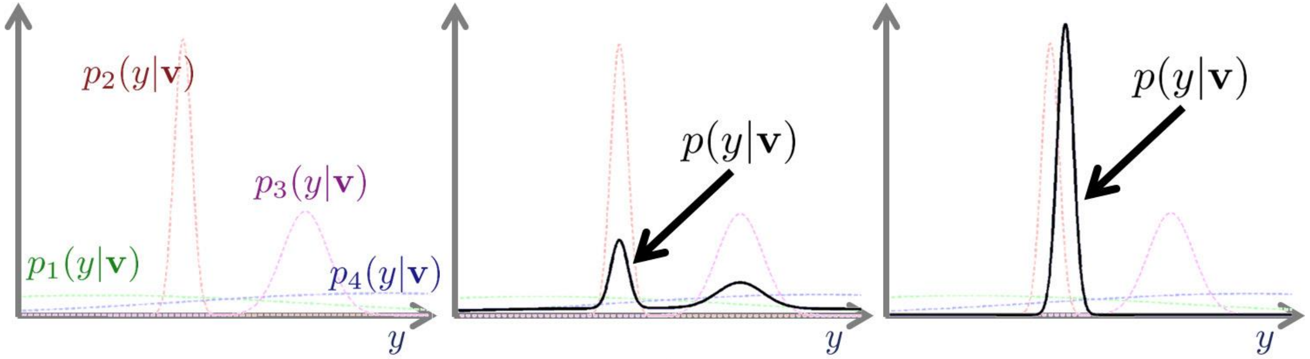
Low randomness, high tree correlation

(a)

High randomness, low tree correlation

(b)

- Reduction of the effect of possibly noisy tree contributions



(a)

Posteriors of all four decision trees

(b)

Average of all tree posteriors

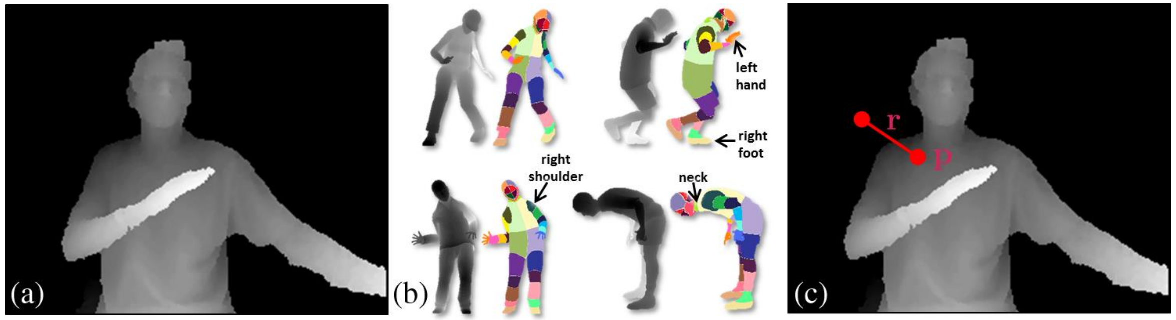
(c)

Product of all tree posteriors

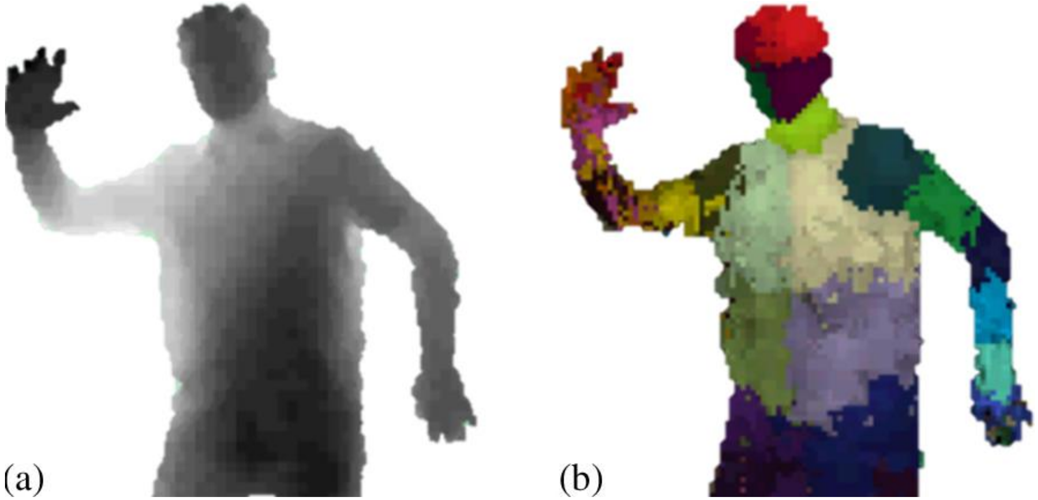
- Improvement variance reduction of bagging by reducing the correlation between the trees, without increasing the variation too much.
- Since trees are notoriously noisy, they benefit greatly from the averaging.
- They generalize well to previously unseen data.
- They can be used to select and rank those variables with the greatest ability to discriminate between classes.
- Speed of processing.

A. Criminisi, J. Shotton and E. Konukoglu (2012). *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*- pp 102-103.

Example of Application: Microsoft Kinect for Xbox 360



Given a depth image such as the one shown in Figure above (a) we wish to say which body part each pixel belongs to. This is a typical job for a classification forest. In this application there are 31 different body part classes: c ∈ {left hand, right hand, head, l. shoulder, r. shoulder, . . .}. The unit of computation is a single pixel in position **p** ∈ R2 and with associated feature vector v(p) ∈ Rd.



During testing, given a pixel **p** in a previously unseen test image we wish to estimate the posterior p(c|v). Visual features are simple depth comparisons between pairs of pixel locations

A. Criminisi, J. Shotton and E. Konukoglu (2012). *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*- pp 128.

Example of Application: RF in Remote Sensing

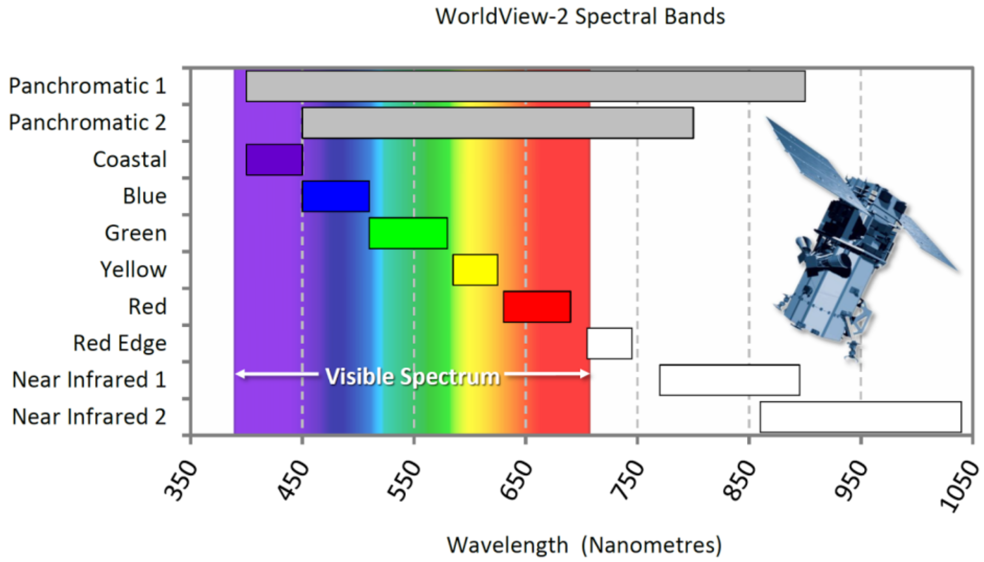


Image: false colour image generated using bands 532 (Near IR, Red, Green)

- RapidEye Image (Google Earth)
- WorldView-2
- LandSat
- MODIS
- IKONOS
- SAR (oil spills)

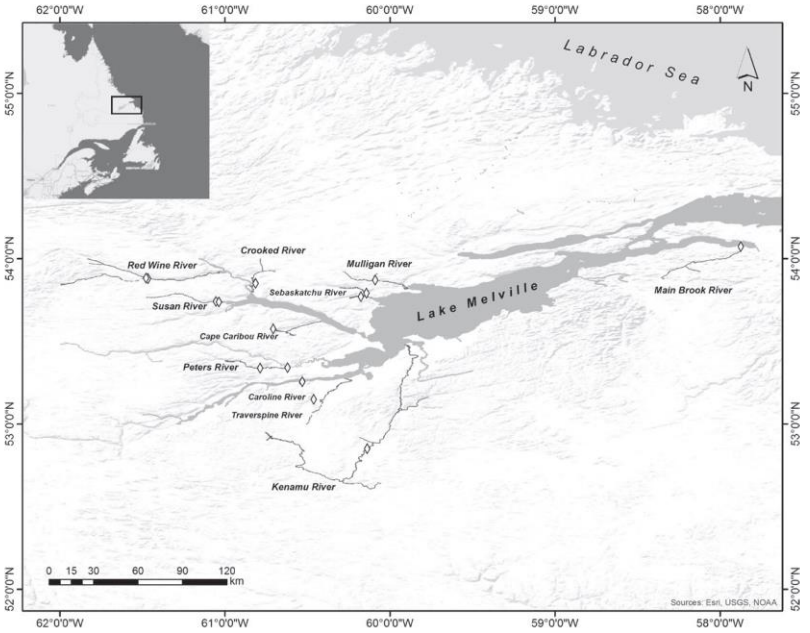
Map Classes:

- Urban Buildings
- Insect defoliaton levels
- Boreal forest habitats
- Biomass
- Tree health and canopy cover
- Oil Spills

M. Belgiu, L. Dragut (2015). *Random forest in remote sensing: A review of applications and future directions*- ISPRS Journal of Photogrammetry and Remote Sensing.

Example of Application: RF for genetic population Assignment

Atlantic salmon (Salmo salar) and Alaskan Chinook salmon (Oncorhynchus tshawytscha)



Genetic assignment of individuals to their source populations is useful for uncovering the spatial distribution of populations and migration patterns relevant to wildlife management and conservation. For exploited species, assignment tests may be used to monitor population-specific exploitation, ensuring the maintenance of genetic diversity and improving management practices through the identification of over-exploited stocks.

- Genotyping to Phenotyping.
- Filtering of SNP, (Single-nucleotide polymorphism) which are the most common type of genetic variation among people. Each SNP represents a difference in a single DNA building block, called a nucleotide.
- RF demonstrated >= 90% Self- Assignment Accuracy. A improvement in 8-11% over traditional F_{ST} method.

E. Sylvester, P. Bentzen. I. Bradbury, M. Clement, J. Pearce, J. Horne, R. Beiko (2017).
Applications of random forest feature selection for fine-scale genetic population assignment - Evolutionary Applications published by John Wiley & Sons Ltd

Bibliography and Other Sources of Information

- A. Criminisi, J. Shotton and E. Konukoglu (2012). Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning
- Hastie, T., Tibshirani, R. and Friedman, J. (2008). The Elements of Statistical Learning - 2^nd Edition – Springer – Chapter 15: 588.
- M. Belgiu, L. Dragut (2015). Random forest in remote sensing: A review of applications and future directions- ISPRS Journal of Photogrammetry and Remote Sensing.
- E. Sylvester, P. Bentzen. I. Bradbury, M. Clement, J. Pearce, J. Horne, R. Beiko (2017). Applications of random forest feature selection for fine-scale genetic population assignment - Evolutionary Applications published by John Wiley & Sons Ltd

Seminal Articles

- L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, Classification and Regression Trees. Chapman and Hall/CRC, 1984.
- R. E. Schapire, "The strength of weak learnability," Machine Learning, vol. 5, no. 2, pp. 197–227, 1990.
- Y. Amit and D. Geman, "Randomized inquiries about shape; an application to handwritten digit recognition," Technical Report 401, Department of Statistics, University of Chicago, IL, 1994.
- Y. Amit and D. Geman., "Shape quantization and recognition with randomized trees," Neural Computation, vol. 9, pp. 1545–1588, 1997.
- T. K. Ho, "Random decision forests," in International Conference on Document Analysis and Recognition, pp. 278–282, 1995.
- L. Breiman, "Random forests," Technical Report TR567, UC Berkeley, 1999.
- L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.