

## Random forests (Breiman, 2001)

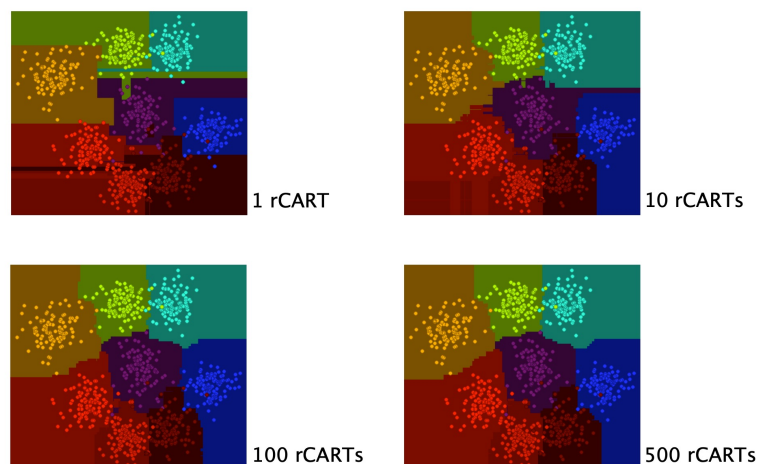
- từ tập học LS (learning set) có N phần tử
- xây dựng tập hợp T mô hình cơ sở độc lập nhau
- mô hình thứ i được xây dựng trên tập mẫu bootstrap, chú ý
  - **Tại nút trong, chọn ngẫu nhiên  $n'$  thuộc tính ( $n' \ll n$ ) và tính toán phân hoạch tốt nhất dựa trên  $n'$  thuộc tính này**
  - Cây được xây dựng đến độ sâu tối đa, không cắt nhánh
- 1 bootstrap : lấy mẫu N phần tử có hoàn lại từ tập LS
- khi phân loại : sử dụng luật bình chọn số đông (majority vote)
- hồi quy : tính giá trị trung bình của dự đoán của các mô hình

29

29

## Random forests (Breiman, 2001)

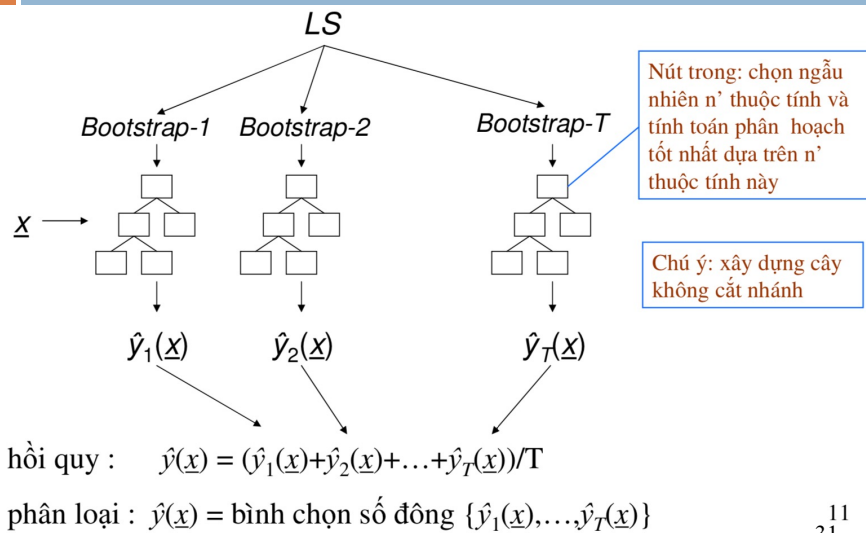
30



<https://perso.math.univ-toulouse.fr/motimo/files/2013/07/random-forest.pdf>

30

## Random forests (Breiman, 2001)



11  
31

31

## Random forests (Breiman, 2001)

| Diameter | Color  | Grows in summer | Shape  | Label  |
|----------|--------|-----------------|--------|--------|
| 3        | orange | yes             | circle | Orange |
| 1        | red    | yes             | circle | Cherry |
| ...      | ...    | ...             | ...    | ....   |



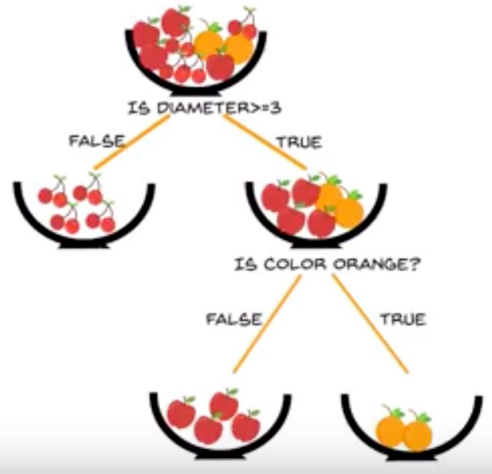
DIAMETER = 3  
COLOUR = ORANGE  
GROWS IN SUMMER = YES  
SHAPE = CIRCLE

32

32

## Random forests (Breiman, 2001)

LET THIS BE TREE 1

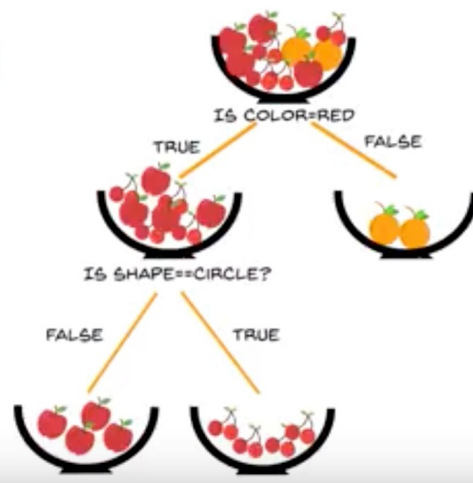


33

33

## Random forests (Breiman, 2001)

LET THIS BE TREE 2

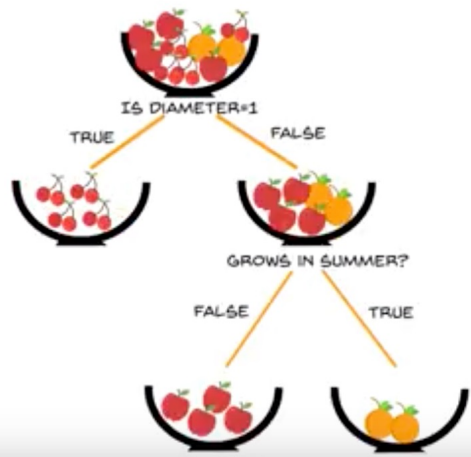


34

34

## Random forests (Breiman, 2001)

LET THIS BE TREE 3



35

## Random forests (Breiman, 2001)



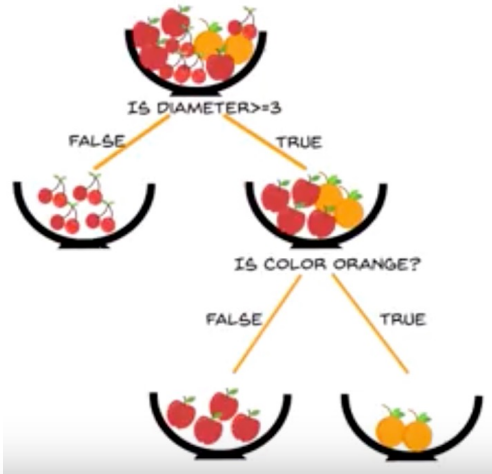
DIAMETER = 3  
COLOUR = ORANGE  
GROWS IN SUMMER = YES  
SHAPE = CIRCLE

36

## Random forests (Breiman, 2001)

LET THIS BE TREE 1

DIAMETER = 3  
COLOUR = ORANGE  
GROWS IN SUMMER = YES  
SHAPE = CIRCLE



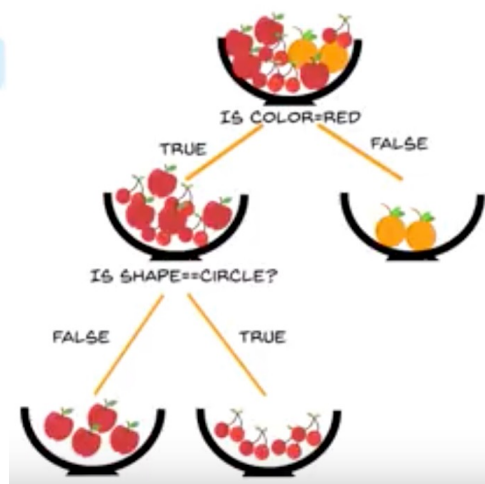
37

37

## Random forests (Breiman, 2001)

LET THIS BE TREE 2

DIAMETER = 3  
COLOUR = ORANGE  
GROWS IN SUMMER = YES  
SHAPE = CIRCLE



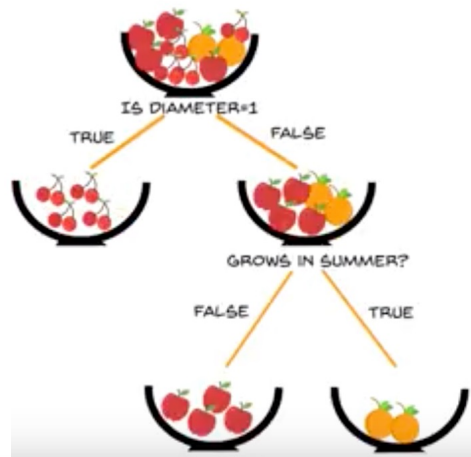
38

38

## Random forests (Breiman, 2001)

LET THIS BE TREE 3

DIAMETER = 3  
COLOUR = ORANGE  
GROWS IN SUMMER = YES  
SHAPE = CIRCLE



39

■ boosting technique

40

## Boosting (Freund & Schapire, 1995)

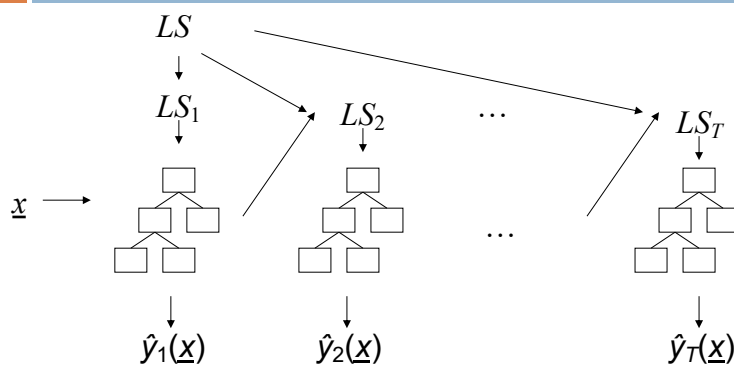
### ■ Boosting

- từ tập học LS có N phần tử
- xây dựng tập hợp T mô hình cơ sở tuần tự
- mô hình thứ i được xây dựng trên tập mẫu lấy từ LS, **tập trung vào các phần tử bị phân loại sai bởi mô hình thứ i-1 trước đó**
- khi phân loại : sử dụng majority vote có trọng số
- hồi quy : tính giá trị trung bình của dự đoán của các mô hình có sử dụng trọng số

41

41

## Boosting (Freund & Schapire, 1995)



hồi quy :  $\hat{y}(\underline{x}) = b_1 \cdot \hat{y}_1(\underline{x}) + b_2 \cdot \hat{y}_2(\underline{x}) + \dots + b_T \cdot \hat{y}_T(\underline{x})$

phân loại :  $\hat{y}(\underline{x}) =$  bình chọn số đông  $\{\hat{y}_1(\underline{x}), \dots, \hat{y}_T(\underline{x})\}$   
với các trọng số tương ứng  $\{b_1, b_2, \dots, b_T\}$

42

42

Giới thiệu về Ensemble-based

**Boosting, Boosting**

kết luận và hướng phát triển

Boosting (Freund & Schapire, 1995)

3

43

Giới thiệu về Ensemble-based

**Boosting, Boosting**

kết luận và hướng phát triển

Boosting (Freund & Schapire, 1995)

44

44

■ 8



Giới thiệu về Ensemble-based

**Boosting, Boosting**

kết luận và hướng phát triển

## Boosting (Freund & Schapire, 1995)

45

45

Giới thiệu về Ensemble-based

**Boosting, Boosting**

kết luận và hướng phát triển

## Boosting (Freund & Schapire, 1995)

46

46

## Nội dung

- Giới thiệu về Ensemble-based
- Bagging, Boosting
- Kết luận và hướng phát triển

47

47

## Phương pháp ensemble-based

- cải thiện rất tốt hiệu quả các phương pháp học thông thường như cây quyết định, naïve Bayes, SVM, etc.
  - dựa trên cơ sở bias/variance
  - xây dựng tập hợp các mô hình cơ sở dựa trên tập học
  - kết hợp các mô hình khi phân loại cho độ chính xác cao
  - kết quả rất khó diễn dịch, ví dụ như 1 rừng gồm hàng trăm cây quyết định

48

48

## Ensemble-based

### ■ phương pháp ensemble-based

- giải quyết các vấn đề về phân loại, hồi quy, gom nhóm, etc.
- cho kết quả tốt, tuy nhiên không thể dịch được kết quả sinh ra
- được ứng dụng thành công trong hầu hết các lĩnh vực tìm kiếm thông tin, nhận dạng, phân tích dữ liệu, etc.

49

49

## Hướng phát triển

- học trên dữ liệu không cân bằng
- diễn dịch kết quả sinh ra
- kiểm chứng sự hợp lệ của phương pháp

50

50

# Python

51

- `sklearn.ensemble`.BaggingRegressor
- `sklearn.ensemble`. BaggingClassifier
- `sklearn.ensemble`.RandomForestRegressor
- `from sklearn.ensemble`.RandomForestClassifier
- `sklearn.ensemble`. AdaBoostClassifier
- `sklearn.ensemble`.AdaBoostRegressor

51

*The End*

52

52