

Assessing feature importance with random forests

Nhóm 4

DamSanX

7/2020

- 1 Random forests
 - Definition
 - Random forest evaluation
- 2 Feature evaluation
 - Impurity feature importance
 - Permutation feature importance

- 1 Random forests
 - Definition
 - Random forest evaluation
- 2 Feature evaluation
 - Impurity feature importance
 - Permutation feature importance

Random forests

Definition

Random forest là một tập hợp gồm nhiều Decision trees (cây quyết định)



Random forests

Random forest evaluation

Có hai cách cơ bản để đánh giá độ chính xác của một random forest

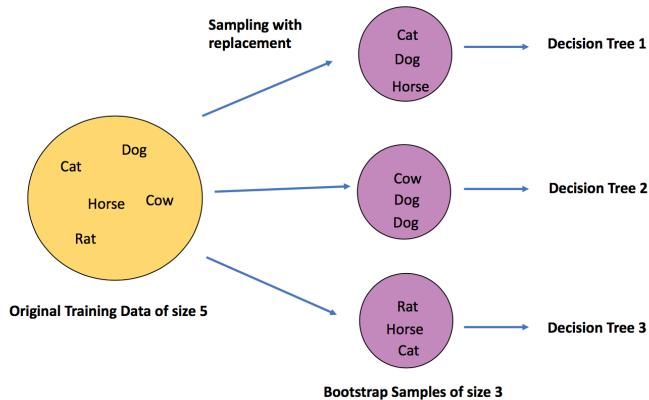
- kiểu "k-fold": chia tập dữ liệu ra làm 2 tập con, đó là tập train và tập test, tập train dùng để luyện mô hình (random forests), còn tập test là để đánh giá độ chính xác của mô hình vừa luyện.
- dựa vào OOB: sử dụng toàn bộ dữ liệu để luyện mô hình, sau đó sử dụng tập OOB để đánh giá độ chính xác của mô hình vừa luyện.

OOB là gì?

Random forests

Random forest evaluation

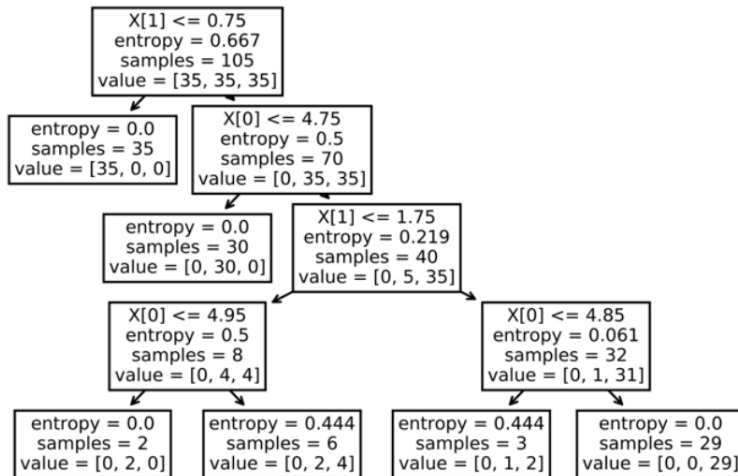
OOB là viết tắt của Out Of Bag.



- 1 Random forests
 - Definition
 - Random forest evaluation
- 2 Feature evaluation
 - Impurity feature importance
 - Permutation feature importance

Feature evaluation

Impurity feature importance



Feature evaluation

Permutation feature importance



color	battery	zin	new	label
blue	80	1	85	B
red	90	0	99	A
yellow	95	1	95	A
white	70	1	99	B
green	88	0	90	B
yellow	85	1	96	A
red	97	0	89	B