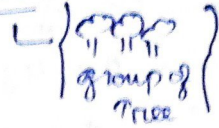


Random Forest

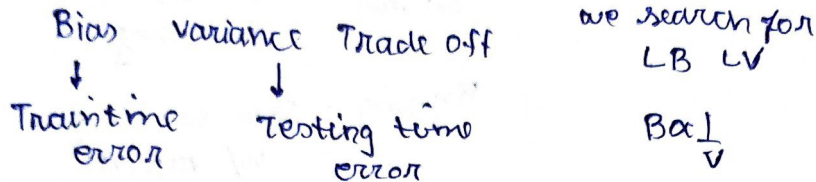
Sun Jan 15
10:17 AM



P-1

P-2

How RF performs so well?



P-3

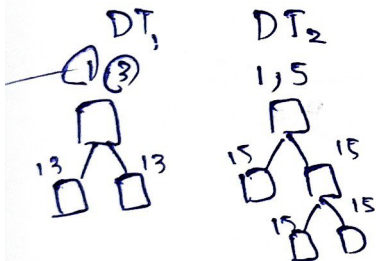
Bagging v/s Random Forest

Q. if I take base_estimator = DT then it is Random Forest

No!
↓
why

Bagging

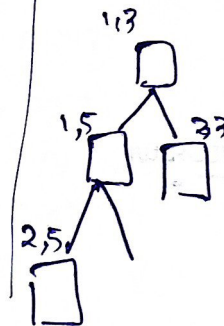
feature sampling = 2



Tree based
col. sampling

Before Tree
formed already
col. selected

Random Forest



Node based
sampling

Hyper-parameters

part

- Random Forest tune
- Decision Tree Train
- common

P-4

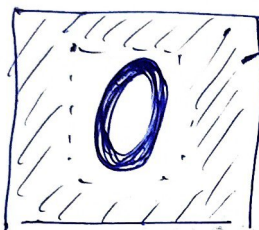
OOB Evaluation

Out of Bagging

- Some are rows taken multiple time
- Some other rows may be not even once

Feature Importance

MNIST



This pixel are not so imp.

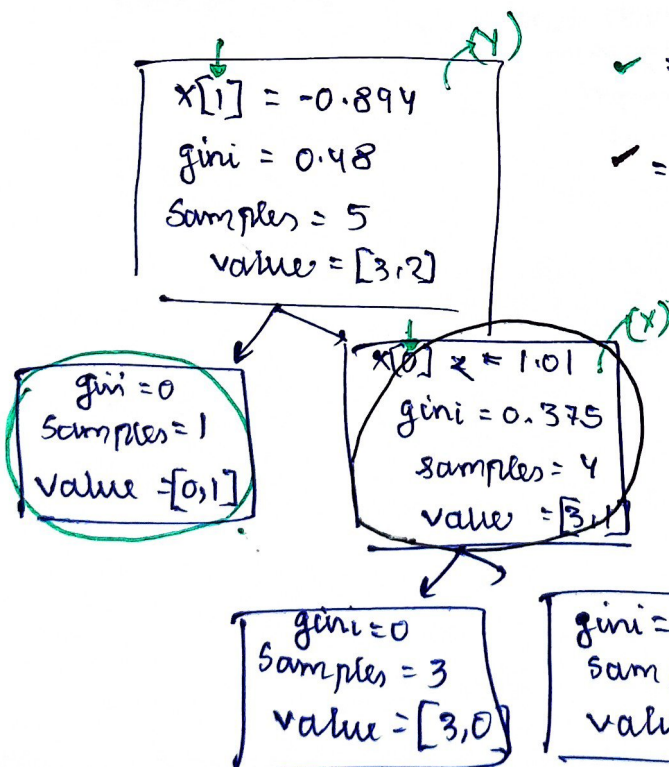
feature imp.

$$f_{ik} = \frac{\sum_{i \in \text{node split feature } k} n_i}{\sum_{i \in \text{all nodes}} n_i}$$

node
no of rows

$$n_i = \frac{N-t}{N} \left[\text{impurity} - \left(\frac{N-t_r}{N-t} \times \text{right impurity} \right) - \left(\frac{N-t_l}{N-t} \times \text{left impurity} \right) \right]$$

Total no of row in sample dataset

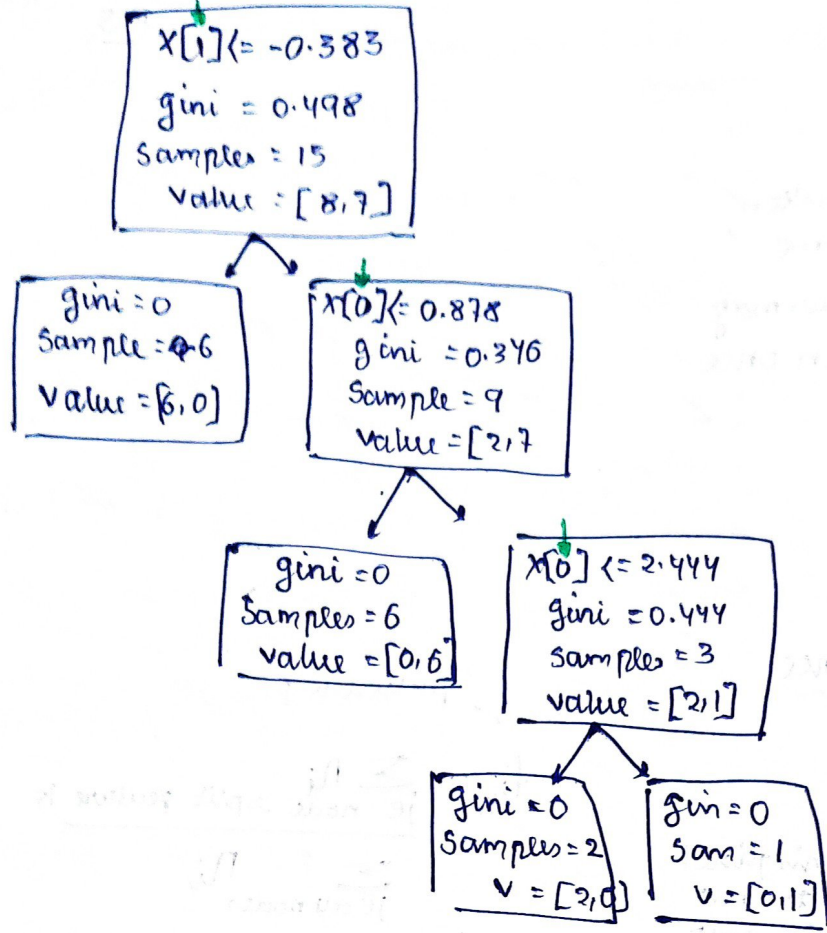


$$\checkmark = \frac{5}{5} \left[0.48 - \frac{4}{5} \times 0.375 - \frac{1}{5} \times 0 \right] = 0.18$$

$$\checkmark = \frac{4}{5} \left[0.375 - 0 - 0 \right] = 0.30$$

$$0^{th} = \frac{x}{x+y} = \frac{0.3}{0.3+0.18} = 0.625$$

$$1^{th} = \frac{y}{y+x} = \frac{0.18}{0.3+0.18} = 0.375$$



first node

$$\frac{15}{15} [0.49 - \frac{9}{15} \times 0.346 - \frac{6}{15} \times 0] = 0.290$$

second node

$$(i) \frac{9}{15} [0.346 - \frac{3}{15} \times 0.444 - \frac{6}{15} \times 0] = 0.118$$

$$(ii) \frac{3}{15} [0.444 - \frac{1}{15} \times 0 - \frac{2}{15} \times 0] = 0.088$$

$$f_{i[0]} = \frac{0.118 \times 0.088}{0.118 + 0.088 + 0.290} = \frac{0.0103}{0.496} = 0.0207$$

$$f_{i[1]} = \frac{0.29}{0.48} = 0.60 = 0.583$$

