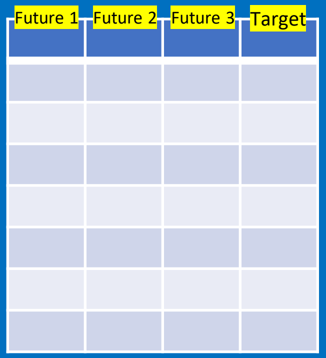
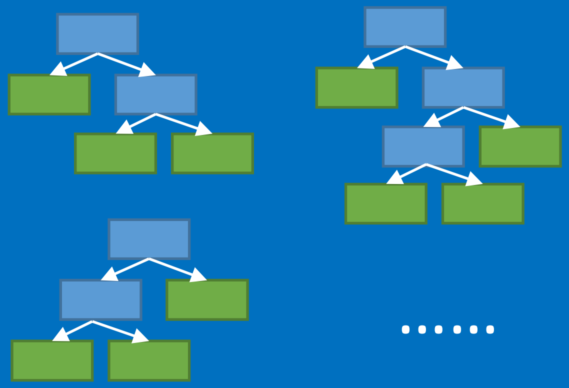
Feature importance – Gini based method

Sijin Zhang

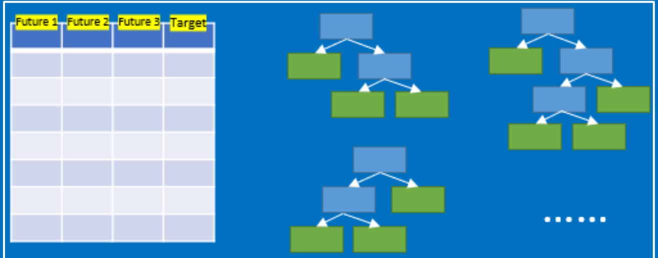
Impurity based feature importance





Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Impurity based feature importance

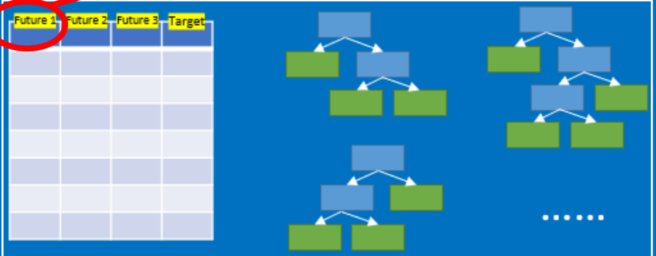


Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature
  - For each tree
    - Compute impurity decrease
    - Weighted by number of samples at node
  - Averaged over all trees
- Normalize the results if needed

Impurity based feature importance



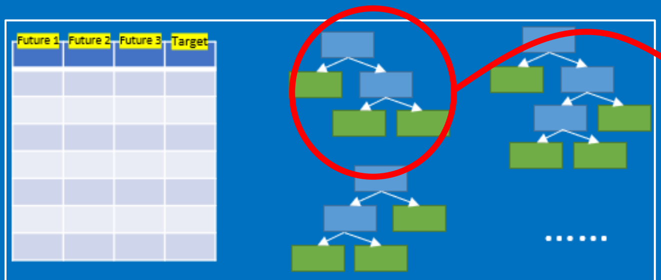
Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance

For example, we have 3 features in this example, let’s look at feature 1, “F1”

- For each feature
  - For each tree
    - Compute impurity decrease
    - Weighted by number of samples at node
  - Averaged over all trees
- Normalize the results if needed

Impurity based feature importance



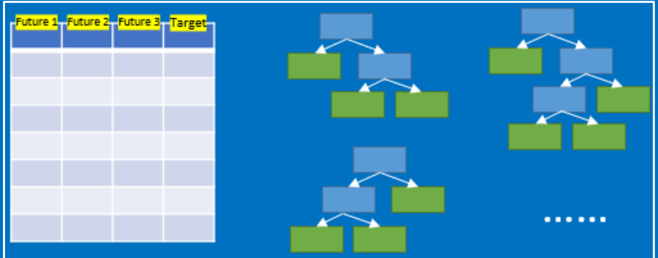
Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
  - For each tree

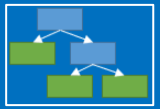
    Let's look at the first tree
    - Compute impurity decrease
    - Weighted by number of samples at node
  - Averaged over all trees
- Normalize the results if needed

Impurity based feature importance

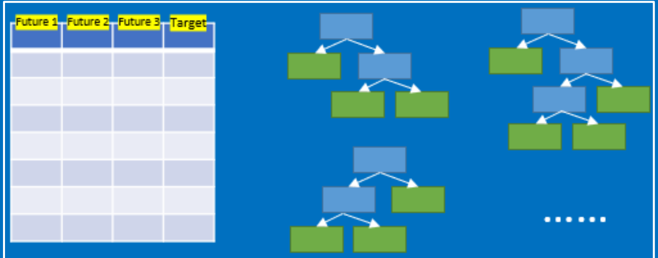


Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
    - For each tree 
        - Compute impurity decrease
        - Weighted by number of samples at node
    - Averaged over all trees
- Normalize the results if needed

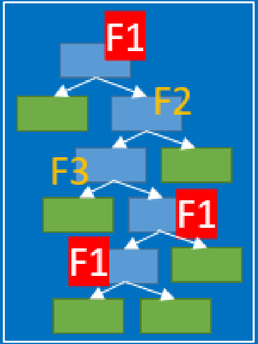
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

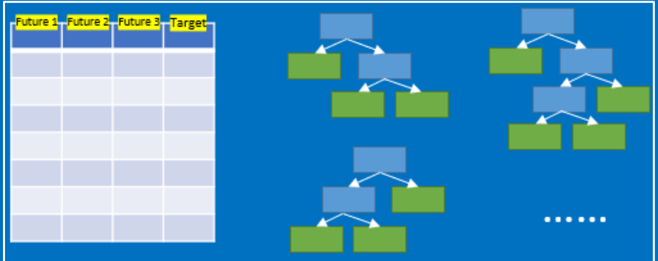
Usually the importance is estimated as below:

- For each feature F1
    - For each tree 
        - Compute impurity decrease
        - Weighted by number of samples at node
    - Averaged over all trees
- Normalize the results if needed



Assuming we have 5 levels of tree, and 3 “F1” features are included.

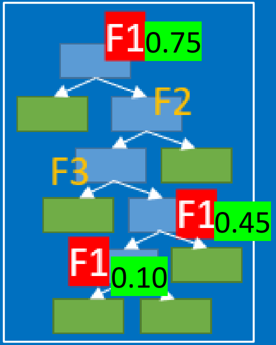
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

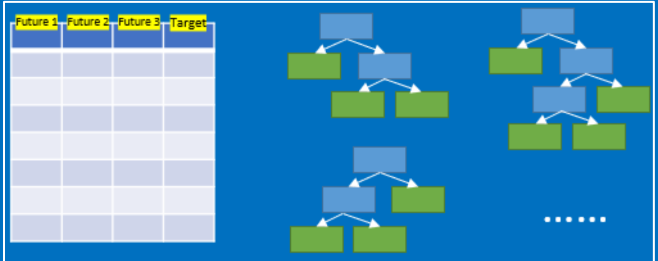
Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samples at node
  - Averaged over all trees
- Normalize the results if needed



Assuming we have the “Gini” values as left …

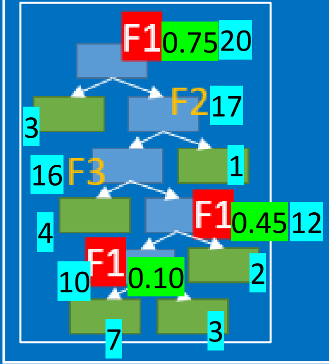
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

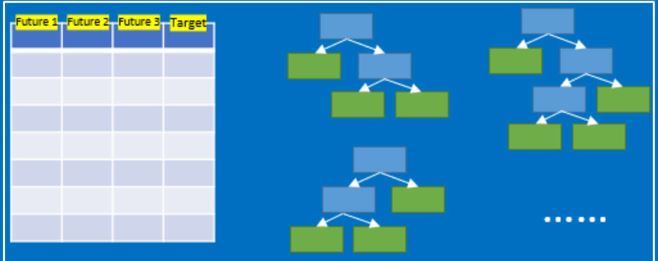
Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samples at node
  - Averaged over all trees
- Normalize the results if needed



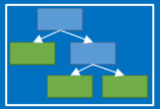
Assuming that in total we have 20 samples, and the samples that each node has is shown left

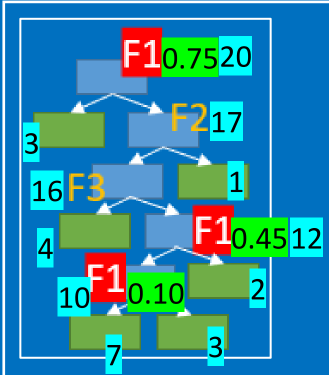
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
    - For each tree 
        - Compute impurity decrease
        - Weighted by number of samples at node
    - Averaged over all trees
- Normalize the results if needed



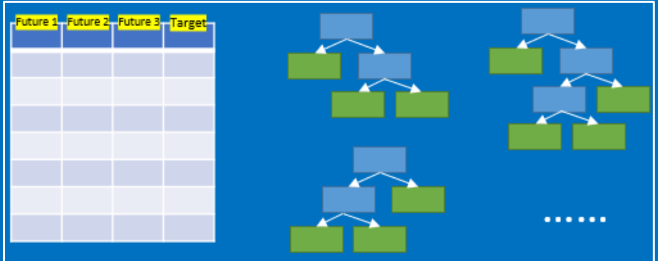
The impotence for F1 can be calculated as

$$I = \sum P \bigtriangleup Gini$$

weight

Gini decrease

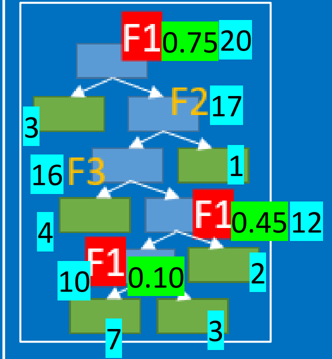
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
    - For each tree 
        - Compute impurity decrease
        - Weighted by number of samp
    - Averaged over all trees
- Normalize the results if needed



The impotence for F1 can be calculated as

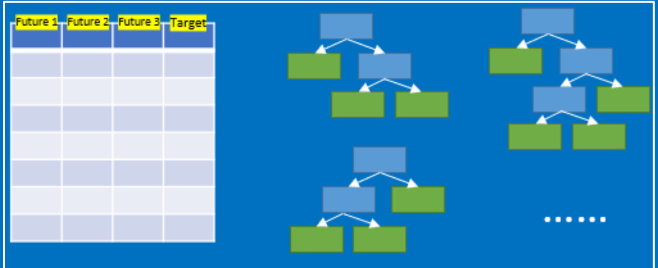
$$I = \sum P \triangle Gini$$

weight

Gini decrease

Step 1: calculate the percentage of samples for each level

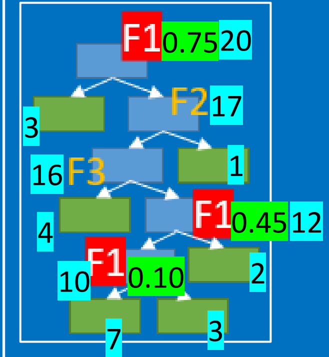
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samp
  - Averaged over all trees
- Normalize the results if needed



The impotence for F1 can be calculated as

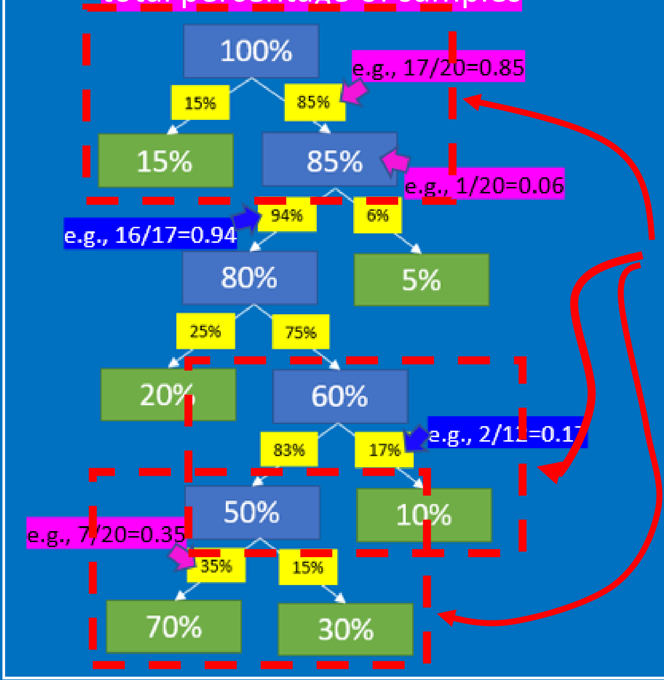
$$I = \sum P \triangle Gini$$

weight

Gini decrease

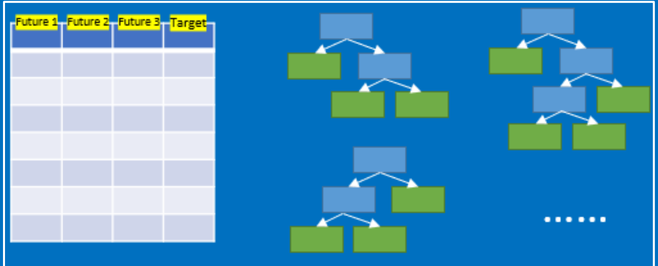
Step 1: calculate the

- level dependant percentage of samples, and
- total percentage of samples



We have 3 “F1” feature appear in the tree

Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

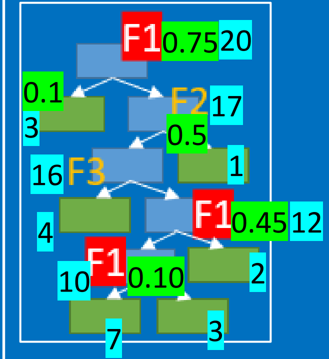
Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samp

We can calculate $\triangle\, Gini$ as

$$\triangle\, Gini_1 = 0.75 - (0.85 \times 0.5 + 0.15 \times 0.1) = 0.31$$

Node Gini
Total percentage of samples
Next node Gini
Total percentage of samples
Next leaf Gini



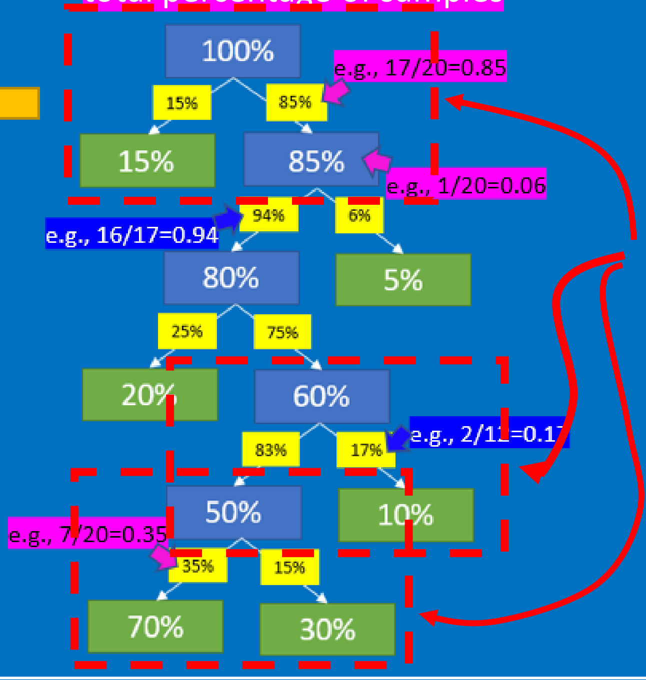
The impotence for F1 can be calculated as

$$I = \sum P \triangle\, Gini$$

weight

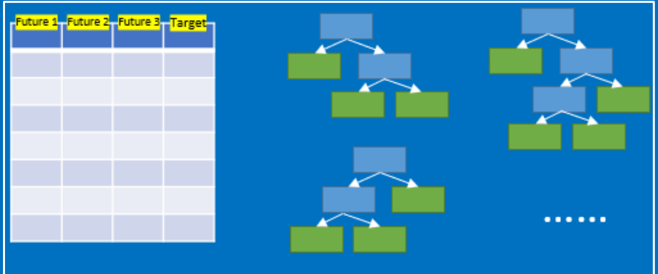
Gini decrease

Step 1: calculate the
- level dependant percentage of samples, and
- total percentage of samples



We have 3 “F1” feature appear in the tree

Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samp

We can calculate $\triangle\, Gini$ as

$$\triangle\, Gini_1 = 0.75 - (0.85 \times 0.5 + 0.15 \times 0.1) = 0.31$$

Node Gini

Total percentage of samples

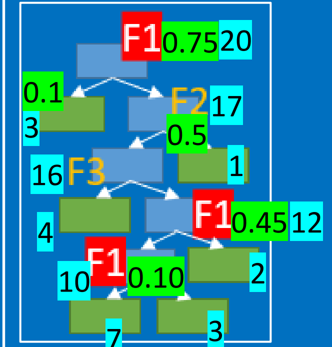
Next node Gini

Total percentage of samples

Next leaf Gini

$$\triangle\, Gini_2 = 0.25$$

$$\triangle\, Gini_3 = 0.1$$



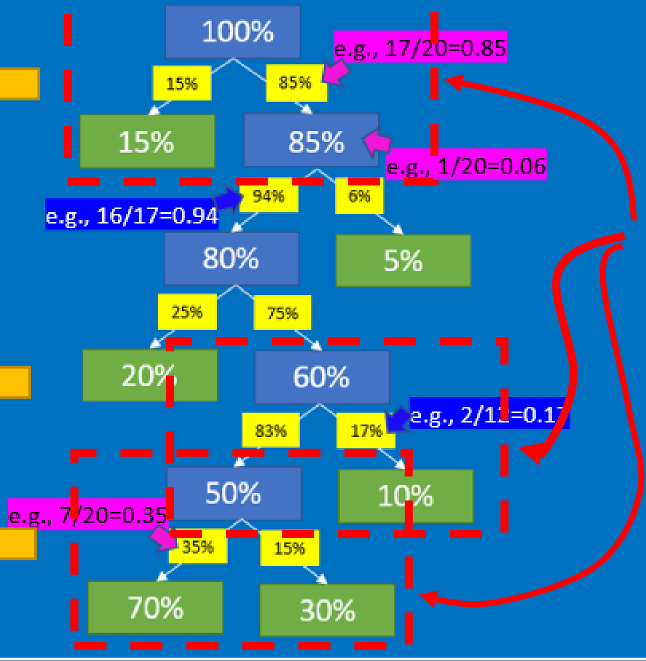
The impotence for F1 can be calculated as

$$I = \sum P \triangle\, Gini$$

weight

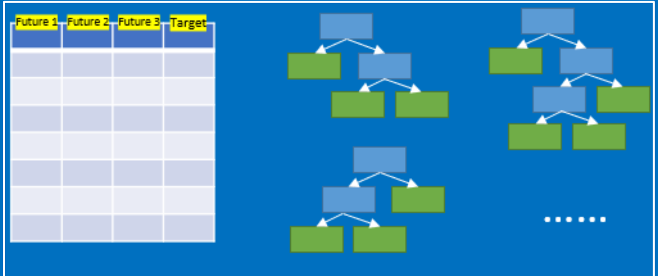
Gini decrease

Step 1: calculate the
- level dependant percentage of samples, and
- total percentage of samples



We have 3 “F1” feature appear in the tree

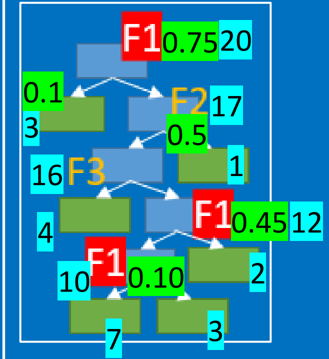
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samp



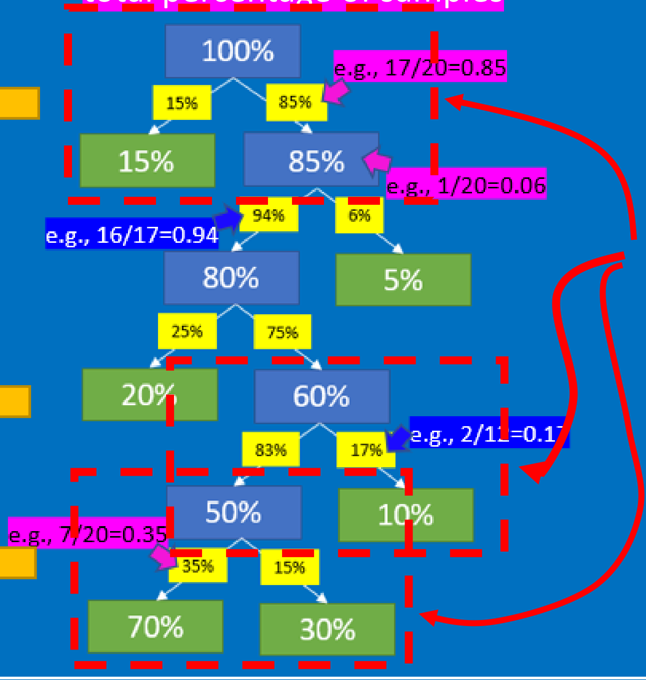
The impotence for F1 can be calculated as

$$I = \sum P \triangle Gini$$

weight

Gini decrease

Step 1: calculate the
- level dependant percentage of samples, and
- total percentage of samples



We have 3 "F1" feature appear in the tree

Level dependant percentage of samples

100%

We can calculate $\triangle\, Gini$ as

$$\triangle\, Gini_1 = 0.75 - (0.85 \times 0.5 + 0.15 \times 0.1) = 0.31$$

Node Gini

Total percentage of samples

Next node Gini

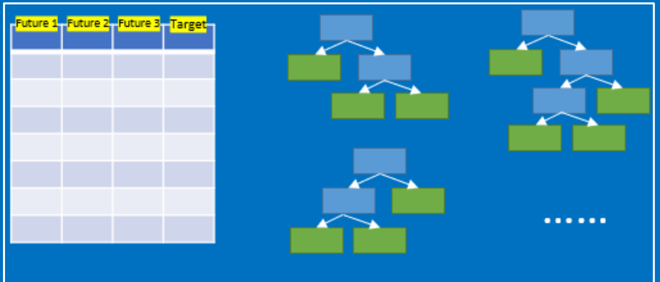
Total percentage of samples

Next leaf Gini

60% $\triangle\, Gini_2 = 0.25$

50% $\triangle\, Gini_3 = 0.1$

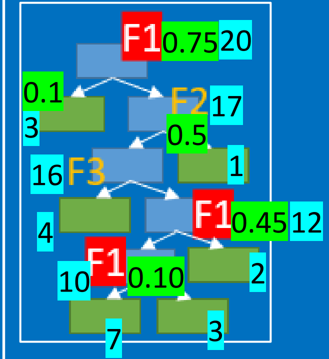
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samp



The impotence for F1 can be calculated as

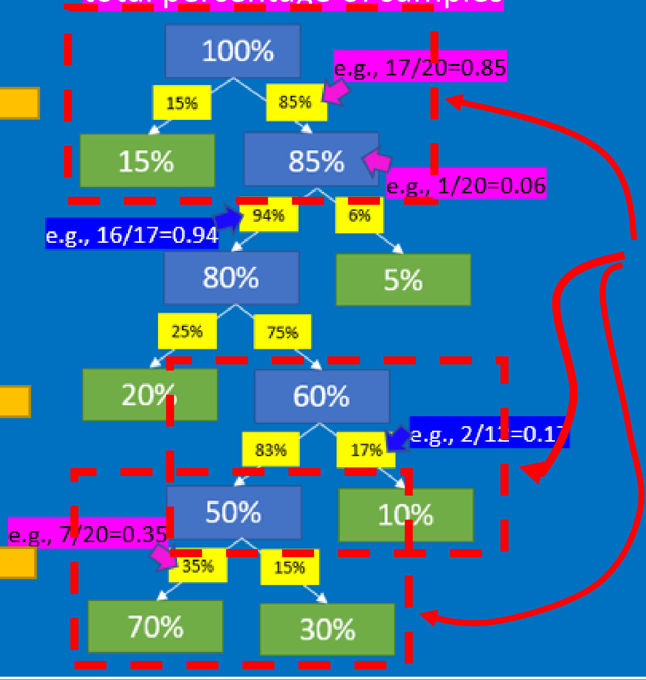
$$I = \sum P \triangle Gini$$

weight

Gini decrease

Step 1: calculate the

- level dependant percentage of samples, and
- total percentage of samples



We have 3 "F1" feature appear in the tree

We can calculate $\triangle Gini$ as

$$\triangle Gini_1 = 0.75 - (0.85 \times 0.5 + 0.15 \times 0.1) = 0.31$$

Node Gini

Total percentage of samples

Next node Gini

Total percentage of samples

Next leaf Gini

100%

60%

$$\triangle Gini_2 = 0.25$$

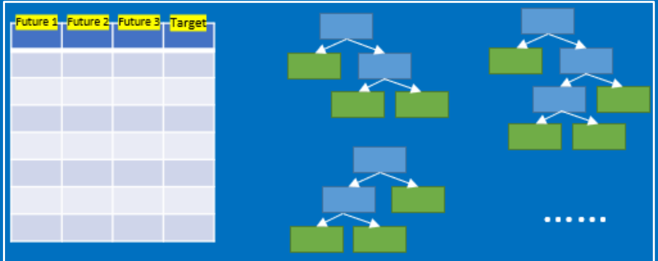
50%

$$\triangle Gini_3 = 0.1$$

$$I_{F_1} = 1 \times 0.31 + 0.6 \times 0.25 + 0.5 \times 0.1 = 0.51$$

The importance for F1

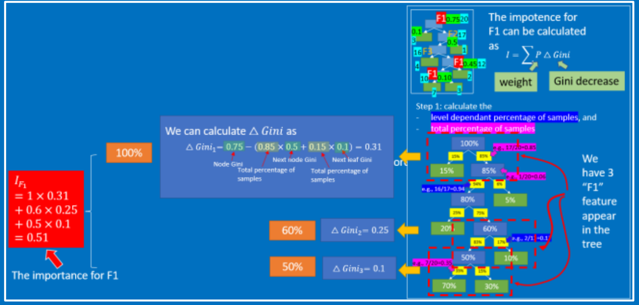
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samples at node



  - Averaged over all trees

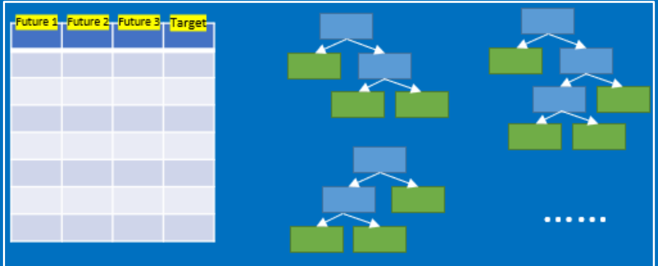
We go through all the trees with the last step, and average the results

- Normalize the results if needed

We through all the features for the above process, and normalize the results.
For example, we can divide the averaged results by the number of trees being used …

The high the value, the more important that the feature is

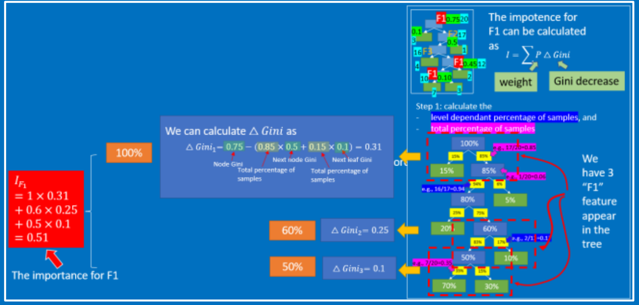
Impurity based feature importance



Assuming that we have a bunch of samples, and through Bagging we created many decision trees

Usually the importance is estimated as below:

- For each feature F1
  - For each tree 
    - Compute impurity decrease
    - Weighted by number of samples at node



  - Averaged over all trees

We go through all the trees with the last step, and average the results

- Normalize the results if needed

We through all the features for the above process, and normalize the results.
For example, we can divide the averaged results by the number of trees being used …

Eventually, the higher the value, the more important the feature is