

# Feature Importance

Friday 3 May 2024 11:14 PM

→ `feature_importances_`: ndarray of shape (n\_features,)  
The impurity-based feature importances.  
The higher, the more important the feature.  
The importance of a feature is computed as the (normalized) total reduction of the criterion brought by that feature. It is also known as the Gini importance [4].  
  
Warning: impurity-based feature importances can be misleading for high cardinality features (many unique values). See :func:`sklearn.inspection.permutation\_importance` as an alternative.

$$\rightarrow f_{i,k} = \frac{\sum_{j \in \text{node split on feature } k} n_i}{\sum_{j \in \text{all nodes}} n_i}$$

$$\rightarrow n_i = \frac{N_t}{N} \left[ \text{impurity} - \left( \frac{N_{tr}}{N_t} \times \text{right\_impurity} \right) - \left( \frac{N_{tl}}{N_t} \times \text{left\_impurity} \right) \right]$$

Annotations for the equation above:  
-  $N_t$ : samples at current node  
-  $N_{tr}$ : samples at right child node  
-  $N_{tl}$ : samples at left child node  
-  $N$ : # no. of samples