

Decision Trees, Random Forests and Extreme Trees

Marcin Kuta

Decision Trees

- Decision Trees
- Random Forests
- Extra Trees

Splitting and pruning criteria

Classification

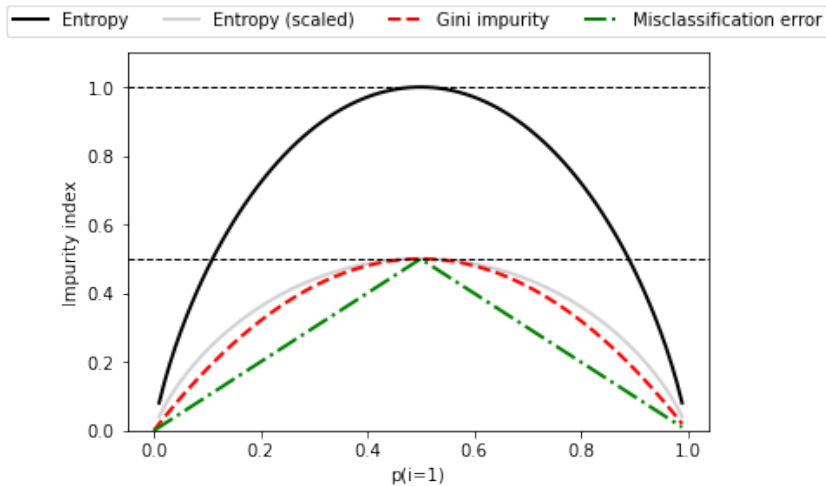
- Entropy

$$-\sum_{i=1}^K p_i \log_2 p_i \quad (1)$$

- Information gain
- Gini impurity

$$\sum_{i=1}^K p_i(1 - p_i)$$

Attribute selection criteria



Splitting and pruning criteria

Regression

- Mean Squared Error

$$\frac{1}{n} \sum_{i=1} (y_i - \bar{y})^2$$

- Half Poisson deviance

$$\frac{1}{n} \sum_{i=1}^K (y_i \log \frac{y_i}{\bar{y}} - y_i + \bar{y})^2$$

- [1] <https://github.com/pietroventurini/machine-learning-notes/blob/main/3%20-%20Decision%20Trees.ipynb>
- [2] <https://jakevdp.github.io/PythonDataScienceHandbook/05.08-random-forests.html>
- [3] <https://github.com/rasbt/machine-learning-book/blob/main/ch03/ch03.ipynb>
Decision tree learning
- [4] https://inria.github.io/scikit-learn-mooc/trees/trees_module_intro.html