Standard Classification Trees – Split impurity measures

From "Elements of Statistical Learning" Section 9.2.3: "*If the target is a classification outcome taking values 1, 2, . . . ,K, the only changes needed in the tree algorithm pertain to the criteria for splitting nodes and pruning the tree.*"

(i) **Misclassification error** – For any region $\Omega \in \mathcal{T}$ let

$$p_{\Omega,l} := \frac{\#\{y_i \in C_l \ : x_i \in \Omega\}}{\#\{x_i \in \Omega\}}.$$

Let $l_\Omega := \max_{l'} p_{\Omega,l}$. Then we look for a split $\Omega' \bigcup \Omega'' = \Omega$, that minimizes

$$1 - p_{\Omega',l(\Omega')} + 1 - p_{\Omega'',l(\Omega'')}.$$

With normalization

$$\frac{\#\{x_i \in \Omega'\}}{\#\{x_i \in \Omega\}}\left(1 - p_{\Omega',l(\Omega')}\right) + \frac{\#\{x_i \in \Omega''\}}{\#\{x_i \in \Omega\}}\left(1 - p_{\Omega'',l(\Omega'')}\right) \Leftrightarrow \#\{x_i \in \Omega' : y_i \notin l_{\Omega'}\} + \#\{x_i \in \Omega'' : y_i \notin l_{\Omega''}\}.$$

(ii) **Gini index -** $\sum_{l=1}^{L} p_{\Omega,l}\left(1 - p_{\Omega,l}\right)$ promotes the probabilities to be zero or one. So we are minimizing a split $\Omega' \bigcup \Omega'' = \Omega$, for

$$\sum_{l=1}^{L} p_{\Omega',l}\left(1 - p_{\Omega',l}\right) + \sum_{l=1}^{L} p_{\Omega'',l}\left(1 - p_{\Omega'',l}\right).$$

With normalization,

$$\frac{\#\{x_i \in \Omega'\}}{\#\{x_i \in \Omega\}}\sum_{l=1}^{L} p_{\Omega',l}\left(1 - p_{\Omega',l}\right) + \frac{\#\{x_i \in \Omega''\}}{\#\{x_i \in \Omega\}}\sum_{l=1}^{L} p_{\Omega'',l}\left(1 - p_{\Omega'',l}\right).$$

(iii) Cross entropy - $\text{Entropy}(\Omega) := -\sum_{l=1}^{L} p_{\Omega,l} \log\left(p_{\Omega,l}\right)$: we are looking for a "compact representation of the classes". We don't need to choose one as in (i).

With normalizations

$$\frac{\#\{x_i \in \Omega'\}}{\#\{x_i \in \Omega\}}\text{Entropy}(\Omega') + \frac{\#\{x_i \in \Omega''\}}{\#\{x_i \in \Omega\}}\text{Entropy}(\Omega'')$$

Information gain

$$\text{Entropy}(\Omega) - \left(\frac{\#\{x_i \in \Omega'\}}{\#\{x_i \in \Omega\}}\text{Entropy}(\Omega') + \frac{\#\{x_i \in \Omega''\}}{\#\{x_i \in \Omega\}}\text{Entropy}(\Omega'')\right)$$