Decision Tree Algorithm

Linh Nguyễn

November 2021

1 EX1:

Tự biến đổi lại gini score, gini index và cách xây dựng cây decision tree

$$Gini = 1 - \sum_{i=1}^{C} \left(\frac{n_i}{N}\right)^2 = 1 - \sum_{i=1}^{C} \frac{n_i^2}{N^2} = 1 - \sum_{i=1}^{N} {a_i}^2$$

with

$$a_i = \frac{n_i}{N}$$

$=> \sum a_i = 1$
we have:
S=$\sum_{i=1}^{N} a_i^2$
with: 0$\leq a_i \leq 1 and \sum a_i = 1$
S_{min} :
Bunhia:
$(a_1^2 + a_2^2 + ... + a_n^2)(b_1^2 + b_2^2 + ... + b_n^2) \geq (a_1b_1 + a_2b_2 + ... + a_nb_n)$

$$=> NS \geq (\sum a_i)^2 = 1 (because b_i = 1)$$

$$=> S \geq \frac{1}{N}$$

Dấu bằng xảy ra <=> $a_1 = a_2 = ... = a_N = \frac{1}{N}$
S_{max} :

$$\sum a_i^2 \leq (\sum a_i)^2 = 1$$

$$\left(because \left(\sum a_i\right)^2 = \sum a_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j\right)$$

Dấu "=" <=> $a_j = 1 and a_{i \neq j} = 0$
=> S_{max} => $a_j = 1, a_{i \neq j} = 0$ => $gini_{min}$ <=> $n_j = N and n_{i \neq j} = 0$
=> S_{min} => $a_i = a_j$ <=> $gini_{max}$ <=> $n_i = n_j$
Gini index:

$$g_p - \sum_{i=1}^{N} \left(\frac{n_i}{N}\right) g(c_i)$$