

解答:

Example	A_1	A_2	A_3	Output y
x_1	1	0	0	0
x_2	1	0	1	0
x_3	0	1	0	0
x_4	1	1	1	1
x_5	1	1	0	1

如果我们要计算每一次 split, 我们需要计算每一个属性 A_i 的 attribute 的 $Remainder(A_i)$, 我们选择每一次 split 中 $Remainder(A_i)$ 最小的那个值, 选择那个属性去 split。

所以, 第一次 split 的情况是:

$$Remainder(A_1) = -\left(\frac{4}{5}\right) * \left(\frac{2}{4} * \log\left(\frac{2}{4}\right) + \frac{2}{4} * \log\left(\frac{2}{4}\right)\right) - \left(\frac{1}{5}\right) \left(0 + \frac{1}{1} \log\left(\frac{1}{1}\right)\right) = 0.8$$

$$Remainder(A_2) = -\left(\frac{3}{5}\right) * \left(\frac{2}{3} \log\left(\frac{2}{3}\right) + \frac{1}{3} \log\left(\frac{1}{3}\right)\right) - \left(\frac{2}{5}\right) * \left(0 + \frac{2}{2} \log\left(\frac{2}{2}\right)\right) = 0.55$$

$$Remainder(A_3) = -\left(\frac{2}{5}\right) * \left(\frac{1}{2} \log\left(\frac{1}{2}\right) + \frac{1}{2} \log\left(\frac{1}{2}\right)\right) - \left(\frac{3}{5}\right) * \left(\frac{1}{3} \log\left(\frac{1}{3}\right) + \frac{2}{3} \log\left(\frac{2}{3}\right)\right) = 0.95$$

所以我们可以看出, 最小的是 $Remainder(A_2)$, 第一次 split 的属性是 A_2 。我们发现当 A_2 等于 0 的时候, y 的输出都是 0。所以我们接下来只用考虑, A_1 和 A_3 属性有关与 (x_3, x_4, x_5) 的 $Remainder(A_i)$, 我们通过此计算第二次 split。

Example	A_1	A_2	A_3	Output y
x_1	1	0	0	0
x_2	1	0	1	0
x_3	0	1	0	0
x_4	1	1	1	1
x_5	1	1	0	1

所以第二次 split 的情况是:

$$Remainder(A_1) = -\left(\frac{2}{3}\right) * \left(\frac{2}{2} \log\left(\frac{2}{2}\right) + 0\right) - \left(\frac{1}{3}\right) * \left(0 + \frac{1}{1} \log\left(\frac{1}{1}\right)\right) = 0$$

$$Remainder(A_3) = -\left(\frac{1}{3}\right) * \left(\frac{1}{1} \log\left(\frac{1}{1}\right) + 0\right) - \left(\frac{2}{3}\right) * \left(\frac{1}{2} \log\left(\frac{1}{2}\right) + \frac{1}{2} \log\left(\frac{1}{2}\right)\right) = 0.67$$

所以我们可以看出, 最小的是 $Remainder(A_1)$, 第二次 split 的是属性 A_1 。至此, (x_1, x_2, x_3, x_4, x_5) 都被正确分类了。