

第10讲 HW

万晨阳 3210105327

1. 计算按照属性A和B划分时的信息增益

$S: [4+, 6-]$
 $E(S) = -0.4 \log_2 0.4 - 0.6 \log_2 0.6 = 0.971$

A

T F

$[4+, 3-]$ $[0+, 3-]$

$E_{A=T} = -\frac{4}{7} \log_2 \frac{4}{7} - \frac{3}{7} \log_2 \frac{3}{7} = 0.9852$ $E_{A=F} = 0$

$Gain(S, A) = E(S) - \frac{7}{10} E_{A=T} - \frac{3}{10} E_{A=F} = 0.2813.$

$S: [4+, 6-]$
 $E(S) = -0.4 \log_2 0.4 - 0.6 \log_2 0.6 = 0.971$

B

T F

$[3+, 1-]$ $[1+, 5-]$

$E_{B=T} = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} = 0.811$ $E_{B=F} = -\frac{1}{6} \log_2 \frac{1}{6} - \frac{5}{6} \log_2 \frac{5}{6} = 0.65$

$Gain(S, B) = E(S) - \frac{4}{10} E_{B=T} - \frac{6}{10} E_{B=F} = 0.2566.$

2. 决策树归纳算法将会选择哪个属性?

决策树归纳算法会计算每一个候选属性的信息增益，然后选择信息增益最高的一个。本题中

$Gain(S, A) = 0.2813 > Gain(S, B) = 0.2566$ ，因此决策树归纳算法会选择属性**A**。