

填空: 最大熵模型参数形式

简单: ①向前向后算法

②高斯混合模型的E-M算法

③最大熵原理

④C4.5原理

计算: ①.

例6 假设随机变量  $X$  有 5 个取值  $\{A, B, C, D, E\}$ , 要估计取各个值的概率  $P(A), P(B), P(C), P(D), P(E)$ .

解 这些概率值满足以下约束条件: 当有几个约束条件时可用

$$P(A) + P(B) + P(C) + P(D) + P(E) = 1$$

满足这个约束条件的概率分布有无穷多个. 如果没有任何其他信息, 仍要对概率分布进行估计, 一个办法就是认为这个分布中取各个值的概率是相等的:

$$P(A) = P(B) = P(C) = P(D) = P(E) = \frac{1}{5}$$

等概率表示了对事实的无知. 因为没有更多的信息, 这种判断是合理的.

有时, 能从一些先验知识中得到一些对概率值的约束条件, 例如:

$$P(A) + P(B) = \frac{3}{10}$$

$$P(A) + P(B) + P(C) + P(D) + P(E) = 1$$

满足这两个约束条件的概率分布仍然有无穷多个. 在缺少其他信息的情况下, 可以认为  $A$  与  $B$  是等概率的,  $C, D$  与  $E$  是等概率的, 于是,

$$P(A) = P(B) = \frac{3}{20}$$

$$P(C) = P(D) = P(E) = \frac{7}{30}$$

如果还有第 3 个约束条件:

$$P(A) + P(C) = \frac{1}{2}$$

$$P(A) + P(B) = \frac{3}{10}$$

$$P(A) + P(B) + P(C) + P(D) + P(E) = 1$$

⑦

例 5.2 对表 5.1 所给的训练数据集  $D$ , 根据信息增益准则选择最优特征.  
解 首先计算经验熵  $H(D)$ .

$$H(D) = -\frac{9}{15} \log_2 \frac{9}{15} - \frac{6}{15} \log_2 \frac{6}{15} = 0.971$$

然后计算各特征对数据集  $D$  的信息增益. 分别以  $A_1, A_2, A_3$  表示年龄、有工作、有自己的房子和信贷情况 4 个特征, 则

(1)

$$\begin{aligned} g(D, A_1) &= H(D) - \left[ \frac{5}{15} H(D_1) + \frac{5}{15} H(D_2) + \frac{5}{15} H(D_3) \right] \\ &= 0.971 - \left[ \frac{5}{15} \left( -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} \right) \right. \\ &\quad \left. + \frac{5}{15} \left( -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} \right) + \frac{5}{15} \left( -\frac{4}{5} \log_2 \frac{4}{5} - \frac{1}{5} \log_2 \frac{1}{5} \right) \right] \\ &= 0.971 - 0.888 = 0.083 \end{aligned}$$

这里  $D_1, D_2, D_3$  分别是  $D$  中  $A_1$  (年龄) 取值为青年、中年和老年的样本子集. 类似地,

(2)

$$\begin{aligned} g(D, A_2) &= H(D) - \left[ \frac{5}{15} H(D_1) + \frac{10}{15} H(D_2) \right] \\ &= 0.971 - \left[ \frac{5}{15} \times 0 + \frac{10}{15} \left( -\frac{4}{10} \log_2 \frac{4}{10} - \frac{6}{10} \log_2 \frac{6}{10} \right) \right] = 0.324 \end{aligned}$$

(3)

$$\begin{aligned} g(D, A_3) &= 0.971 - \left[ \frac{6}{15} \times 0 + \frac{9}{15} \left( -\frac{3}{9} \log_2 \frac{3}{9} - \frac{6}{9} \log_2 \frac{6}{9} \right) \right] \\ &= 0.971 - 0.551 = 0.420 \end{aligned}$$

5.3 决策树的生成

63

(4)

$$g(D, A_3) = 0.971 - 0.608 = 0.363$$

最后, 比较各特征的信息增益值. 由于特征  $A_3$  (有自己的房子) 的信息增益值最大, 所以选择特征  $A_3$  作为最优特征. ■

算法: 贝叶斯 (上课讲的)

