

HOMEWORK 5

(1) 使用训练数据学习贝叶斯分类器，分别采用极大似然估计和贝叶斯估计 ($\lambda = 1$)，请同学们写出给定输入 $x = (2, M)^T$ 时的类标记。注意写出推导过程。

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X^{(1)}$	1	1	1	1	2	2	2	2	2	3	3	3	3	3	3
$X^{(2)}$	S	M	L	S	M	M	L	L	S	L	S	M	M	L	L
Y	1	1	-1	1	-1	-1	-1	1	1	1	-1	1	1	1	1

(2) 根据《统计学习方法》中表 5.1 所给的训练集数据，利用信息增益比算法 (C4.5 算法) 生成决策树。

(3) 已知 Table1 所示的训练数据，试用平方损失准则生成一个二叉回归树。

Table 1: 训练数据表

x_i	1	2	3	4	5	6	7	8	9	10
y_i	4.50	4.75	4.91	5.34	5.80	7.05	7.90	8.23	8.70	9.00

(4) 在 CART 剪枝过程中，假设第 k 步，对每个内部节点 t 计算 $C(T_t)$ 、 $|T_t|$ 以及

$$g_k(t) = \frac{C(t) - C(T_t)}{|T_t| - 1}$$

记第 k 步所有内部节点的集合为 \mathcal{M}_k ，记 $\alpha_k = g_k(a) = \min_{t \in \mathcal{M}_k} g_k(t)$ ，即节点 a 是使函数 $g_k(t)$ 取值最小的内部节点（假设此内部节点唯一），则将 a 剪枝。记剪枝后内部节点的集合是 \mathcal{M}_{k+1} ，定义 $\alpha_{k+1} = g_{k+1}(b) = \min_{t \in \mathcal{M}_{k+1}} g_{k+1}(t)$ 。请证明 $\alpha_{k+1} > \alpha_k$ 。

以上证明题请以 PDF 格式提交。

提交时间：11 月 4 日，晚 20:00 之前。请预留一定的时间，迟交作业扣 3 分，作业抄袭 0 分。