

机器学习第二章作业

任齐轩

重庆大学-计卓 2 班-20204154

日期: September 11, 2022

1 第一题

根据分层采样可知, 划分方式有 $C_{500}^{350} * C_{500}^{350}$ 种

2 第三题

由题意假设 $BEP = x$, 则有:

$$BEP = R = P = x \quad (1)$$

$$F1 = \frac{2 * P * R}{P + R} = x \quad (2)$$

因此若 $F1_A > F1_B$, 则 $BEP_A > BEP_B$

3 第五题

证明. 由题意可知, $AUC = \frac{1}{m_-} \sum_{i=1}^{m_-} \sum_{x^+ \in D^+} II(f(x^+) > f(x_i)) * \frac{1}{m_+}$

则存在:

$$\begin{aligned} AUC &= \frac{1}{m_- m_+} \sum_{x^- \in D^-} \sum_{x^+ \in D^+} II(f(x^+) > f(x_i)) \\ &= \frac{1}{m_- m_+} (m_- m_+ - \sum_{x^- \in D^-} \sum_{x^+ \in D^+} II(f(x^+) < f(x_i))) \\ &= 1 - \frac{1}{m_- m_+} \sum_{x^- \in D^-} \sum_{x^+ \in D^+} II(f(x^+) < f(x_i)) \\ &= 1 - \ell_{rank}. \end{aligned} \quad (3)$$

综上所述, $AUC = 1 - \ell_{rank}$

□

4 第七题

证明. ROC 曲线上点的坐标为 (FPR, TPR) , 由此可以计算出 FNR ; 而代价曲线为从 $(0, 0)$ 到 $(FNR, 1)$ 的直线。因此, 在得知 ROC 曲线后, 可以通过计算 FNR 得到代价曲线, 反之亦然。因此任意一条 ROC 曲线都有一条代价曲线与之对应, 反之亦然。

□

5 第九题

χ^2 检验过程:

Step1. 建立无关性假设, 通过数据构建四格表

Step2. 根据假设生存新的理论四格表

Step3. 计算 χ^2 值, $\chi^2 = \sum \frac{(A-T)^2}{T}$

Step4. 比较 χ^2 值与临界值, 若小于临界值, 则接受无关性假设, 否则拒绝无关性假设