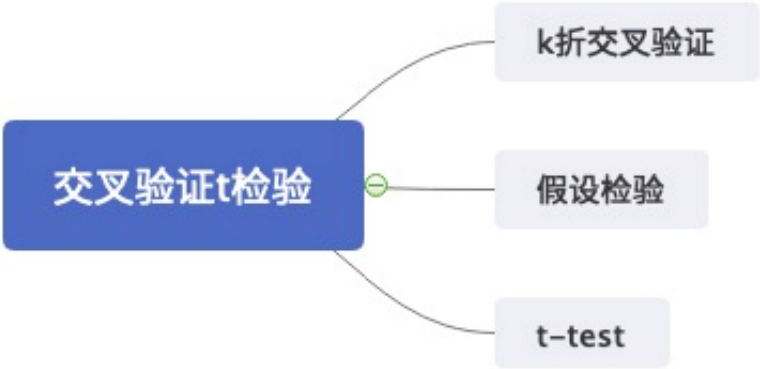
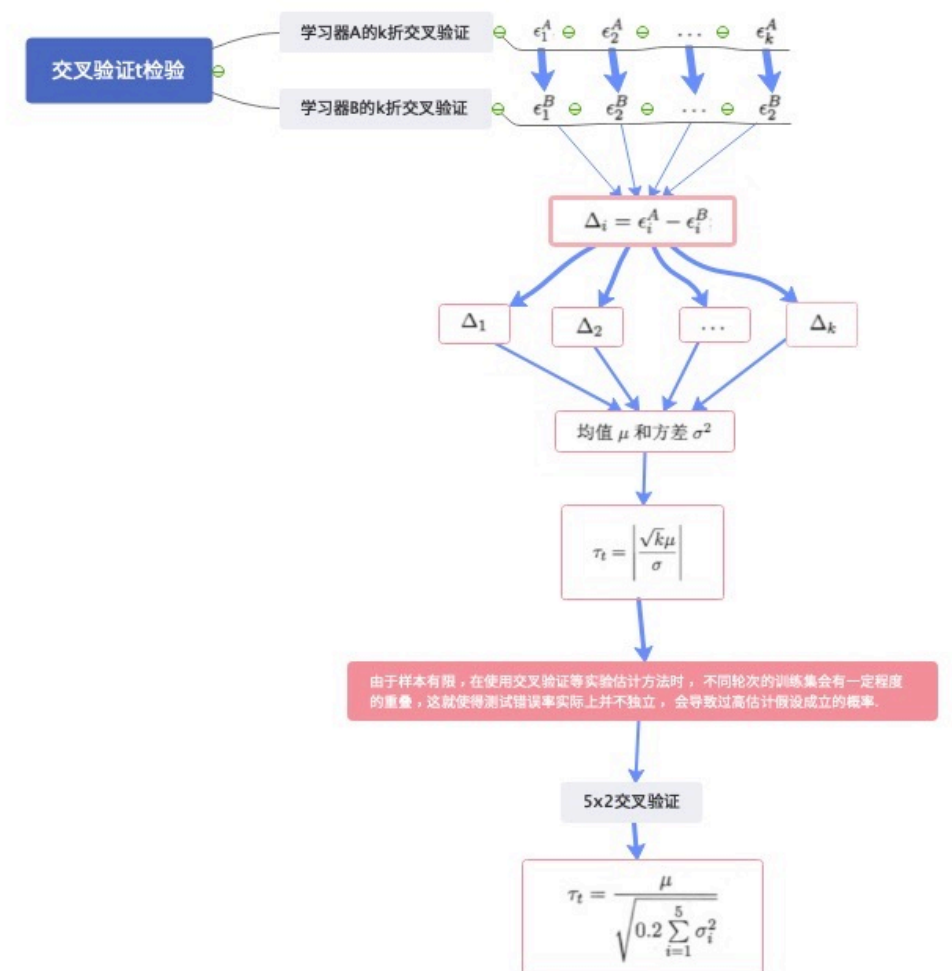


一、前言



二、步骤



1. k折交叉验证

对两个学习器A和B，若我们使用k折交叉验证法得到的测试错误率分别为 $\epsilon_1^A, \epsilon_2^A, \dots, \epsilon_k^A$ 和 $\epsilon_1^B, \epsilon_2^B, \dots, \epsilon_k^B$ ，其中 ϵ_i^A 和 ϵ_i^B 是在相同的第i折训练/测试集上得到的结果，则可用k折交叉验证"成对t检验"(paired t-tests)来进行比较检验. 这里的基本思想是若两个学习器的性能相同，则它们使用相同的训练/测试集得到的测试错误率应相同，即 $\epsilon_i^A = \epsilon_i^B$ 。

2. 假设检验

对 k 折交叉验证产生的 k 对测试错误率:先对每对结果求差， $\Delta_i = \epsilon_i^A - \epsilon_i^B$ ；若两个学习器性能相同，则差值均值班为零. 因此，可根据差值 $\Delta_1, \Delta_2, \dots, \Delta_k$ 来对"学习器 A 与 B 性能相同"(假设)

3.t-test

计算出差值的均值 μ 和方差 σ^2 ，在显著度 α 下，若变量 $\tau_t = \left| \frac{\sqrt{k}\mu}{\sigma} \right|$ 小于临界值 $t_{\alpha/2, k-1}$ 则假设不能被拒绝，即认为两个学习器的性能没有显著差别；否则可认为两个学习器的性能有显著差别，且平均错误率较小的那个学习器性能较优。这里 $t_{\alpha/2, k-1}$ 是自由度为 $k-1$ 的 t 分布上尾部累积分布为 $\alpha/2$ 的临界值。

4.问题

欲进行有效的假设检验，一个重要前提是测试错误率均为泛化错误率的独立采样；由于样本有限，在使用交叉验证等实验估计方法时，不同轮次的训练集会有一定程度的重叠，这就使得测试错误率实际上并不独立，会导致过高估计假设成立的概率。

5.缓解办法：5 x 2交叉验证

5x2交叉验证是做5次2折交叉验证，在每次2折交叉验证之前随机将数据打乱，使得5次交叉验证中的数据划分不重复。对两个学习器A和B，第*i*次2折交叉验证将产生两对测试错误率，我们对它们分别求差，得到第1折上的差值 Δ_i^1 和第2折上的差值 Δ_i^2 。为缓解测试错误率的非独立性，我们仅计算第1次2折交叉验证的两个结果的平均值 $\mu = 0.5(\Delta_i^1 + \Delta_i^2)$ ，但对每次2折实验的结果都计算出其方差 $\sigma_i^2 = \left(\Delta_i^1 - \frac{\Delta_i^1 + \Delta_i^2}{2} \right)^2 + \left(\Delta_i^2 - \frac{\Delta_i^1 + \Delta_i^2}{2} \right)^2$ 。变量 $\tau_t = \frac{\mu}{\sqrt{0.2 \sum_{i=1}^5 \sigma_i^2}}$ 服从自由度为5的 t 分布，其双边检验的临界值 $t_{\alpha/2, 5}$ 当 $\alpha = 0.05$ 时为 2.5706， $\alpha = 0.1$ 时为 2.0150。