

# 机器学习第一周作业

樊泽羲 2200010816

2023 年 9 月 14 日

**题目 1.** 简要论述交叉验证法与自助法的异同

**解答.**

交叉验证法：分为简单交叉验证法与  $k$  折交叉验证法。前者通过给定  $T'$  集与  $V$  集的数量或比例对原始的  $T$  集进行随机划分，再进行训练，通过多次训练的  $R$  取平均值得到模型的测评结果。后者通过在模型训练前把  $T$  随机分为  $k$  个互不相交、大小相似的子集，每次取其中的一个作为  $V$ ，其余用作此次训练的  $T'$ 。当  $V$  遍历  $k$  份数据中的每一个后，将这  $k$  次的经验误差取均值作为评估结果

自助法：固定训练集大小  $|T|$ ，对训练集进行有放回抽样  $|T|$  次，每次将结果拷贝到新的训练集  $T'$ ，并将  $T$  中未被抽中的部分分出作为  $V$ ，把  $T'$  上训练的网络在  $V$  上的  $R$  作为此次训练的误差，多次重复的误差取平均值作为测评结果

共同点：

1. 都是重抽样方法；
2. 都可以用于选择合适的模型参数或结构；
3. 都需要多次采样取结果的平均值；
4. 都通过划分来评估模型的性能

不同点：

1. 抽样方式不同：交叉验证法是无放回抽样，而自助法是有放回抽样；
2. 训练数据的结构不同：交叉验证法的训练集规模小于  $|T|$ ，且不会改变数据的分布，而自助法中的训练集可能会包含  $T$  中的同一个数据点多次，导致数据的分布改变；
3. 对  $h_T$  估计的准确性不同：自助法的  $T'$  规模达到了  $|T|$ ，对  $h_T$  的估计相较交叉验证法更加精准；
4. 交叉验证法更适用于数据集较大的情况下，能够更准确地评估模型的性能和泛化能力，自助法则更适用于数据集较小的情况下，能够更高效地利用数据进行模型评估