

练习题与思考

1. k-近邻算法的性能可以通过集成学习策略加以改善，请设计一种策略用于提升 k-近邻分类器的性能，并选择 3 个数据集加以实验验证。
2. k-means 聚类算法的性能也可以通过集成学习策略加以改善，请设计两种策略用于提升 k-means 算法的性能，并选择 3 个数据集给出实验验证。
3. 设定专家数目为 5，请选用一种简单的分类器作为专家(基分类器)，实现 AdaBoost 算法。请完成：
 - (1) 使用 scatter 函数显示 TwoMoon.mat 数据集，不同类别使用不同颜色显示；
 - (2) 使用 TwoMoon 数据集检验你所设计的 AdaBoost 算法；
 - (3) 在检查每个专家（基分类器）的性能和决策面，以及整体的性能和决策面。
4. 假设训练集上有 5% 的标签被打错，请重复问题 3 中的几个问题，并分析你所得到的实验结果，思考训练集上的错误标签可能会对 AdaBoost 产生的影响。
5. 思考增加训练样本和引入待求解问题的先验知识对解决学习过程的偏倚/方差困境的影响。