

1. 以下说法中正确的是 () ?
 - A. SVM 对噪声(如来自其他分布的噪声样本)鲁棒
 - B. 在 AdaBoost 算法中,所有被分错的样本的权重更新比例相同
 - C. Boosting 和 Bagging 都是组合多个分类器投票的方法,二者都是根据单个分类器的正确率决定其权重
 - D. 给定 n 个数据点,如果其中一半用于训练,一般用于测试,则训练误差和测试误差之间的差别会随着 n 的增加而减少
2. 对于随机森林和 GradientBoosting Trees, 下面说法正确的是: ()
 - 1.在随机森林的单个树中, 树和树之间是有依赖的, 而 GradientBoosting Trees 中的单个树之间是没有依赖的
 - 2.这两个模型都使用随机特征子集, 来生成许多单个的树
 - 3.我们可以并行地生成 GradientBoosting Trees 单个树, 因为它们之间是没有依赖的, GradientBoosting Trees 训练模型的表现总是比随机森林好
 - A. 2
 - B. 1 and 2
 - C. 1 and 3
 - D. 2 and 3
3. 对应 GradientBoosting tree 算法, 以下说法正确的是 : ()
 - 1.当增加最小样本分裂个数, 我们可以抵制过拟合
 - 2.当增加最小样本分裂个数, 会导致过拟合
 - 3.当我们减少训练单个学习器的样本个数, 我们可以降低 variance
 - 4.当我们减少训练单个学习器的样本个数, 我们可以降低 bias
 - A. 2 和 4
 - B. 2 和 3
 - C. 1 和 3
 - D. 1 和 4
4. 简单介绍一下 bagging?
5. 简单介绍一下 boosting, 常用 boosting 算法有哪些?
6. 判断——从偏差-方差分解的角度分析, Boosting 主要关注降低偏差, 因此 Boosting 能基于泛化性能相当弱的学习器构建出很强的集成。()
7. 判断——从偏差-方差分解的角度看, Bagging 主要关注降低方差, 因此它在不剪枝决策树、神经网络等易受扰动的学习器上效用更加明显。()
8. 简述 Bagging 和 Boosting 的区别。
9. GBDT 和随机森林哪个容易过拟合?
10. Gradient Boosting 是一种常用的 Boosting 算法, 是分析其与 AdaBoost 的异同。

11. 试述随机森林为什么比决策树 Bagging 集成的训练速度快。
12. 试设计一种能提升 k 近邻分类器性能的集成学习算法。
13. GBDT 的优点和局限性有哪些？
14. XGBoost 与 GBDT 的联系和区别有哪些？
15. 什么是偏差和方差？
16. 如何从减小方差和偏差的角度解释 Boosting 和 Bagging 的原理？

开课吧
Kaikaba