



集成学习 (Ensemble Learning)

作者: Calvin

QQ: 179209347

Mail: 179209347@qq.com

介绍

笔记简介:

- 面向对象: 深度学习初学者
- 依赖课程: **线性代数, 统计概率**, 优化理论, 图论, 离散数学, 微积分, 信息论

知乎专栏:

<https://zhuanlan.zhihu.com/p/693738275>

Github & Gitee 地址:

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

* 版权声明:

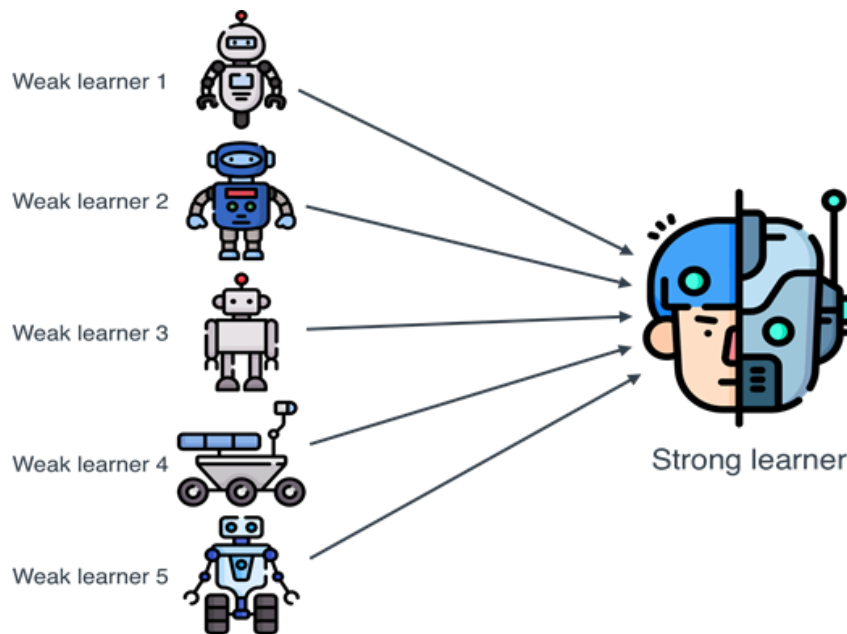
- 仅限用于个人学习
- 禁止用于任何商业用途

集成学习 (Ensemble Learning)

集成学习 (Ensemble Learning) 是一种机器学习方法，通过结合多个基本学习器（弱学习器）的预测结果，来构建一个更强大、更稳健的学习器（强学习器）。集成学习的核心思想是“三个臭皮匠顶个诸葛亮”，即通过组合多个模型的预测结果，来提高整体的泛化能力和预测性能。

定理 10.1: 对于 M 个不同的模型 $f_1(\mathbf{x}), \dots, f_M(\mathbf{x})$ ，平均期望错误为 $\bar{\mathcal{R}}(f)$ 。基于简单投票机制的集成模型 $f^{(c)}(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M f_m(\mathbf{x})$ ，其期望错误在 $\frac{1}{M} \bar{\mathcal{R}}(f)$ 和 $\bar{\mathcal{R}}(f)$ 之间。

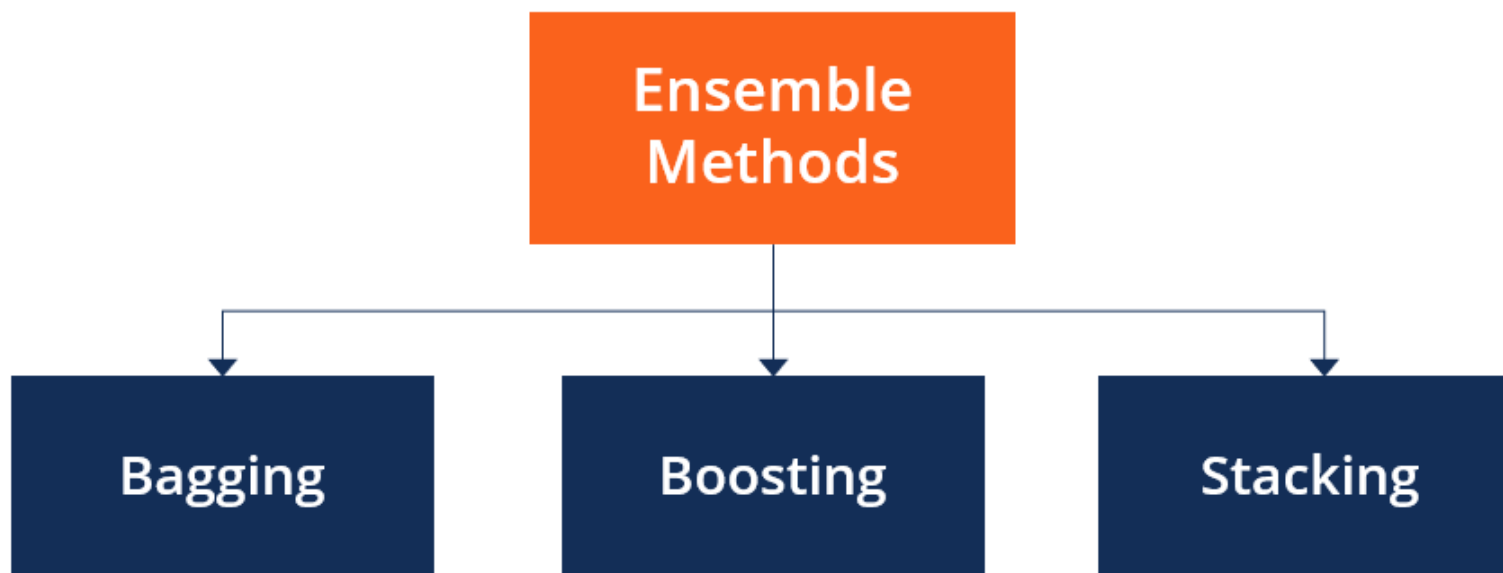
邱锡鹏 <<神经网络与深度学习>>



集成方式

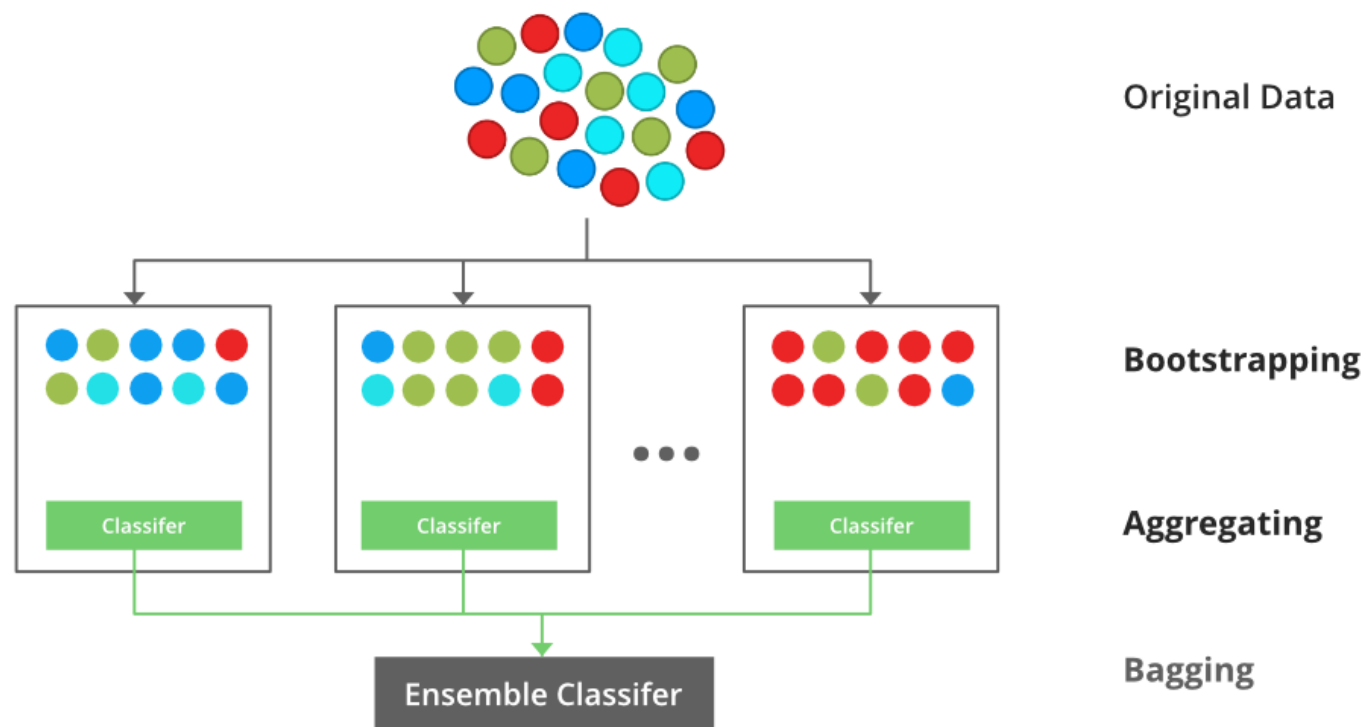
在集成学习中，常见的方法包括：

- Bagging (Bootstrap Aggregating) : 如: 随机森林 (Random Forest)
- Boosting: 常见的Boosting算法包括AdaBoost、Gradient Boosting和XGBoost等。
- Stacking



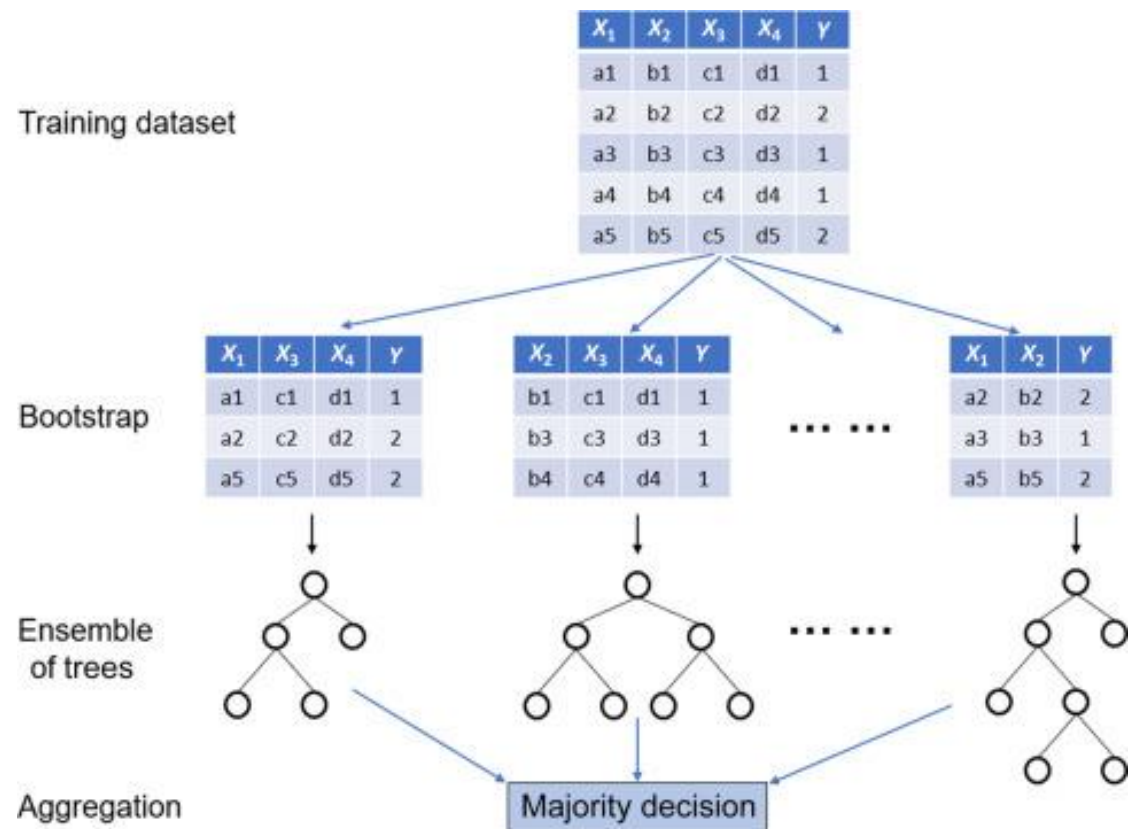
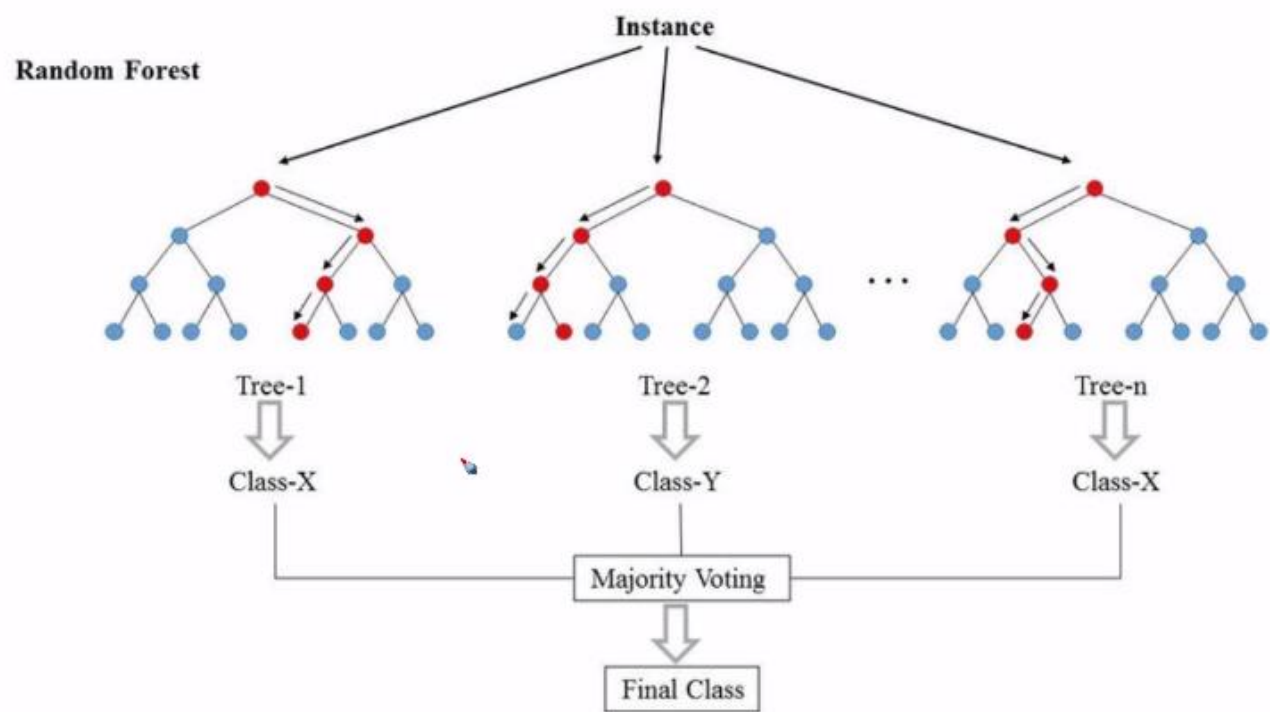
集成方式 - Bagging

Bagging (Bootstrap Aggregating) 通过对训练数据集进行有放回抽样 (bootstrap sampling), 生成多个子数据集, 然后在每个子数据集上训练一个基学习器 (base learner)。最终的预测结果是这些基学习器的平均值 (回归问题) 或投票结果 (分类问题)。



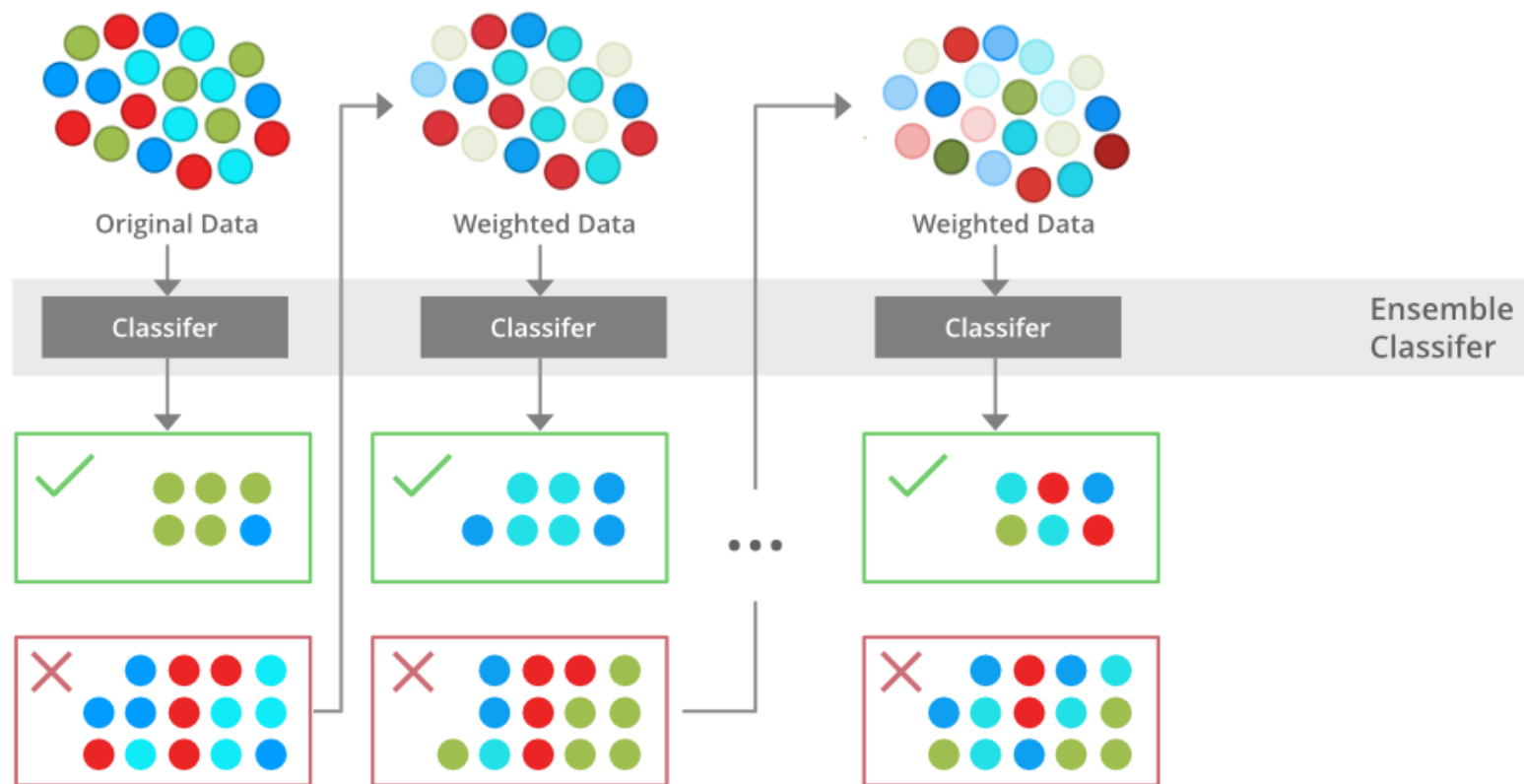
集成方式 – Bagging - 随机森林

随机森林 (Random Forest) 是在Bagging的基础上再引入了随机特征，进一步提高每个基模型之间的独立性。在随机森林中，每个基模型都是一棵决策树。



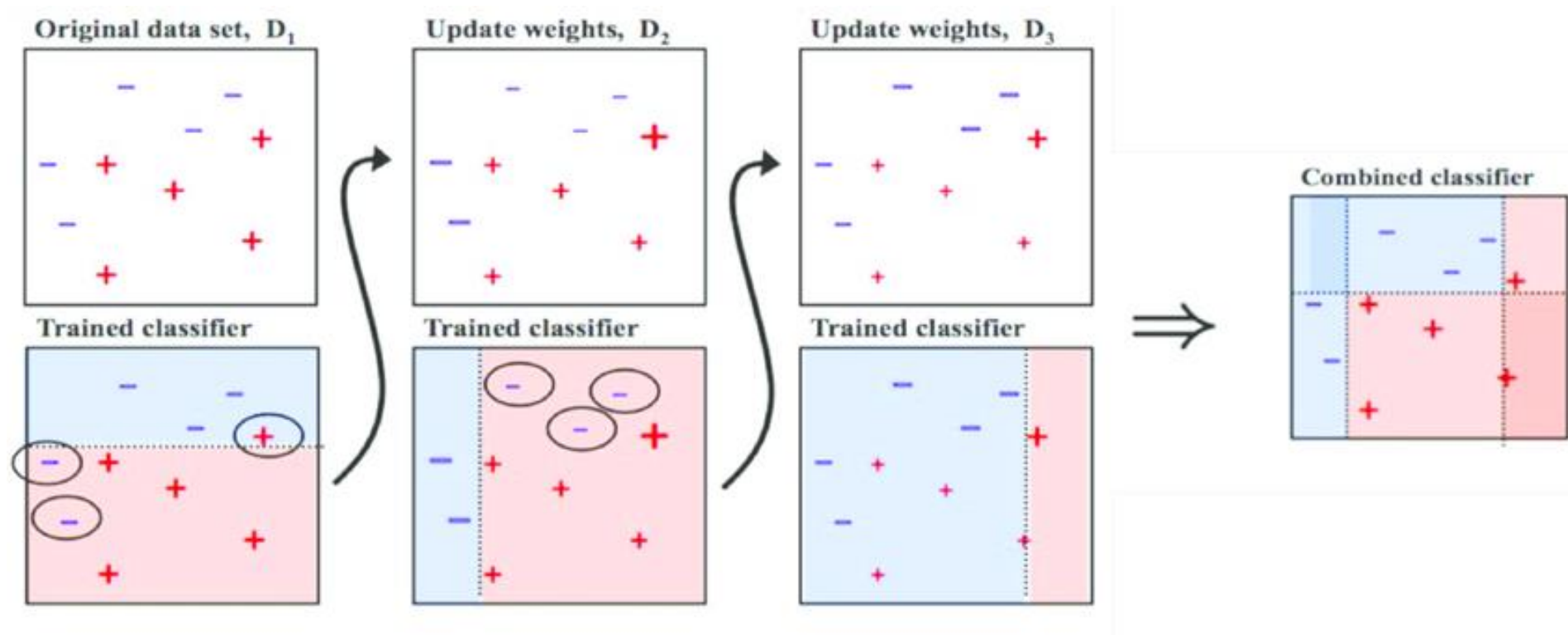
集成方式 - Boosting

Boosting 通过结合多个弱学习器（通常是决策树）来构建一个强学习器。Boosting的主要思想是迭代地训练一系列的弱分类器，每个分类器都试图修正前一个分类器的错误。最终，所有分类器的预测结果通过加权投票或加权求和来得出最终的预测结果。



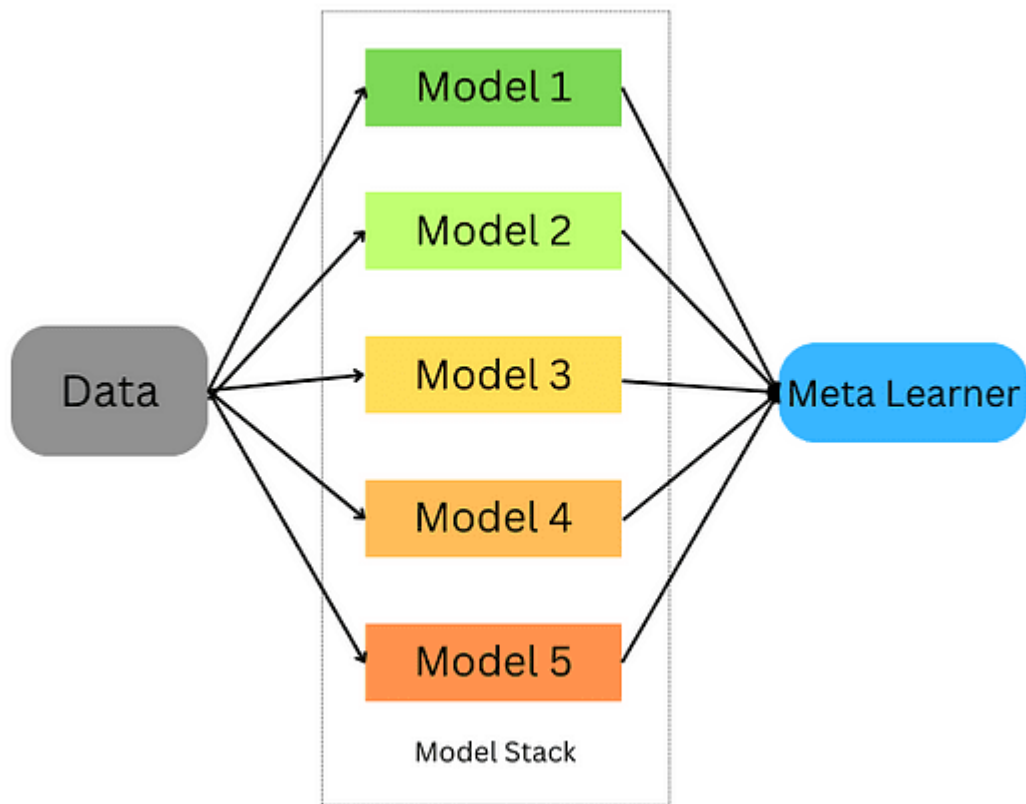
集成方式 - Boosting - AdaBoost

AdaBoost (Adaptive Boosting) 是一种集成学习方法，旨在提高分类器的性能。它通过串行训练一系列弱分类器（通常是决策树），每个分类器都试图修正前一个分类器的错误，从而构建一个强大的集成分类器。



集成方式 - Stacking

Stacking 通过结合多个基本模型的预测结果来生成最终的预测。在Stacking中，首先将训练数据集分成几个子集，然后在每个子集上训练多个基本模型。接着，使用这些基本模型对测试集进行预测，并将它们的预测结果作为新的特征输入到一个元模型中，从而得到最终的预测结果。





Thank

You