# 概述

## 定义

集成学习(Ensemble Learning)是一种机器学习的框架，它会创建多个基模型，每个基模型被训练出来解决同一个问题，然后通过集成这些基模型的预测结果来提升整体表现。

因此对于构建集成学习算法需要解决的两个问题如下：

(1) 基模型如何构建，如何训练？

(2) 以何种方式组合不同类型的基模型，以获得准确且准确的模型？

## 分类

根据解决上述两问题的思路不同，集成学习可以分为三种类型：

(1) Bagging：选用相同的强学习器作为基模型，每个基模型的训练数据不是全部训练数据，而是通过对全部训练数据由放回采样产生的随机子集，预测时各个基模型等权重投票，是一种并行训练结构。

(2) Boosting：选用相同的弱学习器作为基模型，依次训练基模型，每个基模型的训练集根据前一次模型的预测结果进行调整，重点关注被前面模型错误预测的样本，以逐步休整前面基模型的误差。最终的预测结果通过基模型的线性组合产生。是一种串行的训练结构。

(3) Stacking：对不同类型模型的融合。对每个基模型进行训练，并将预测结果作为新的特征，对新特征构成的训练集进行一次训练，最终的预测结果由其产生。

# Bagging

## 定义

## 算法实现

## 随机森林

# Boosting

## 定义

## 算法实现

# 偏差(Bias)、方差(Variance)

## 偏差与方差

## Bagging与Boosting直观理解