第十一章 集成学习
——Stacking

卿来云

Outline

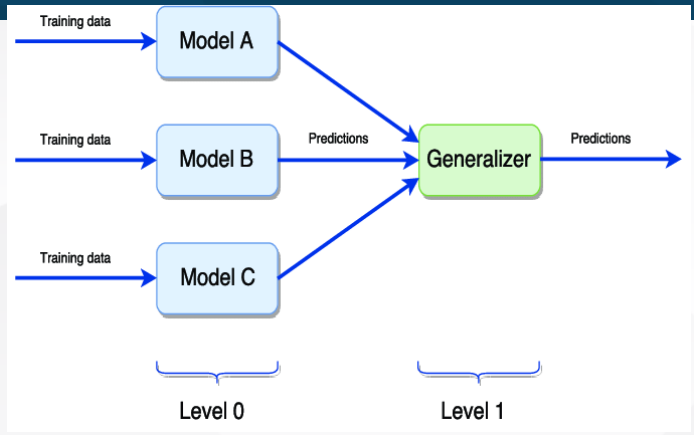
- 模型性能评价
  - No Free Lunch Theorems
  - Occam剃刀原理
  - 偏差-方差折中
- Bagging
  - 随机森林
- Boosting
  - AdaBoost
  - Gradient Boosting Decision Tree (GBDT)
  - XGBoost
  - LightGBM
- Stacking

集成学习

- 我们已经开发了很多机器学习算法/代码。
- 单个模型的性能已经调到最优，很难再有改进。

- 集成学习：用很少量的工作，组合多个基模型，使得系统总的性能提高
  - 基模型最好变化多样，这样不同的基模型集成后形成互补。

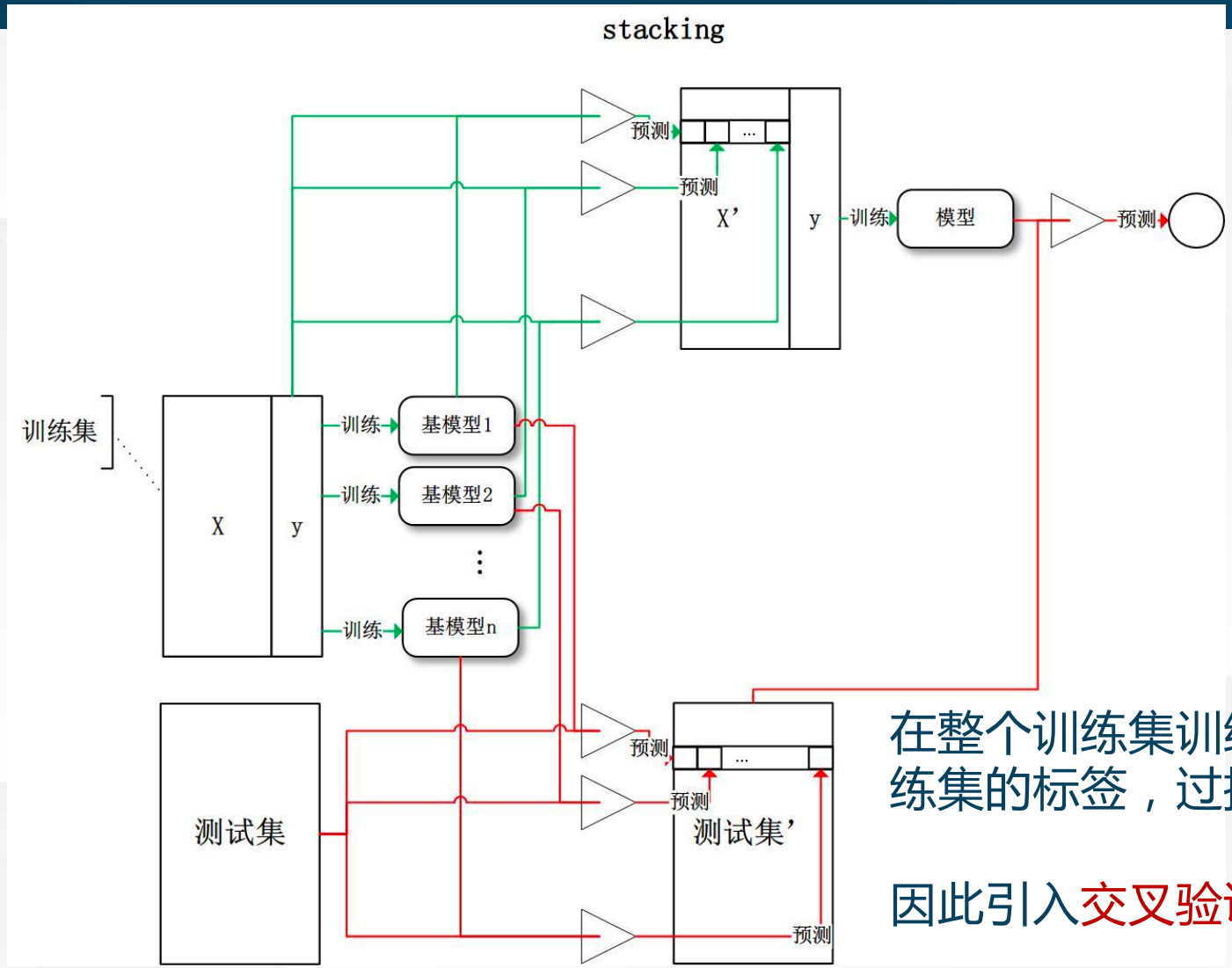
Stacking



- Stacking是一种分层的结构
- 二层Stacking：
  - 将训练好的基模型对训练集进行预测
    - 新的训练集：第j个基模型对第i个训练样本的预测值将作为新的训练集中第i个样本的第j个特征值
    - 新的测试集：所有基模型的对测试集的预测
  - 在新的训练集上训练模型，在新的测试集上进行预测

Stacked Gereralization, David H. Wolpert, *Neural Networks*, Volume 5, Issue 2, Pages 241-259.

Stacking

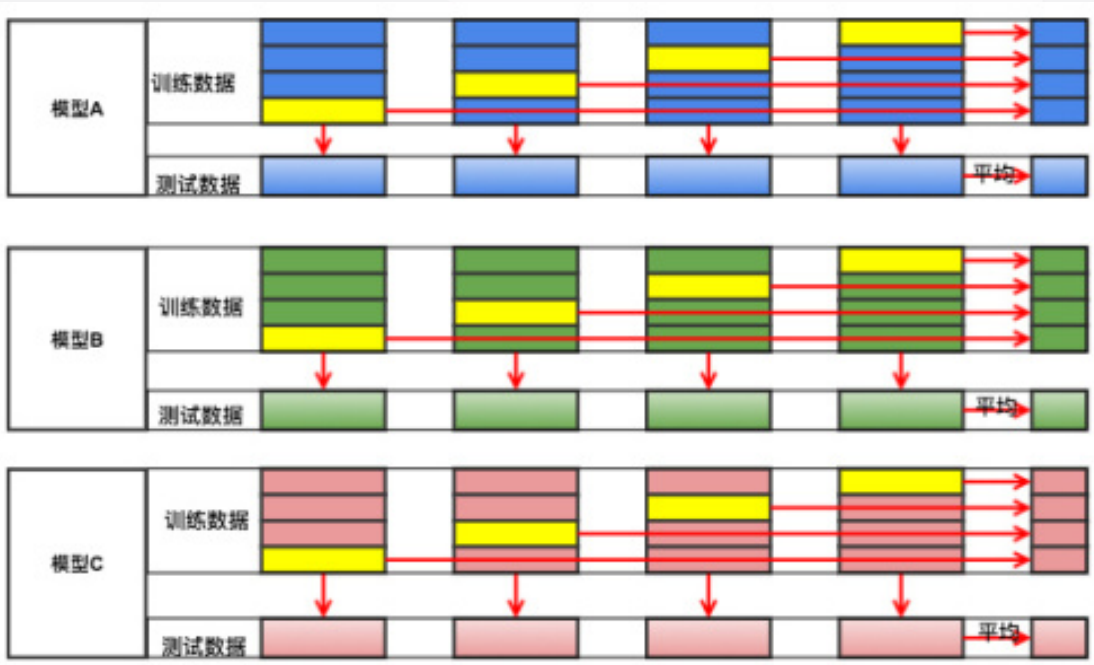


在整个训练集训练的模型反过来去预测训练集的标签，过拟合会非常非常严重 X

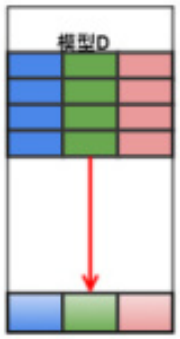
因此引入交叉验证

带交叉验证的Stacking

- 以3个基模型、4折交叉验证为例：



基模型

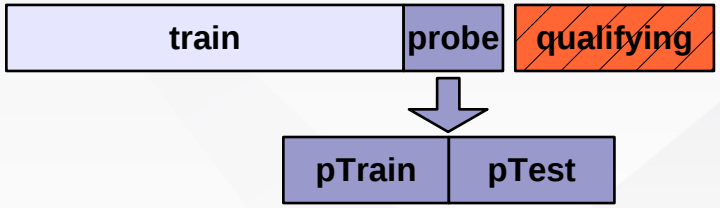


融合后的模型

例：blending.py

交叉融合（ Blending ）

- Blending是由Netflix获胜者提出来的一个词，与Stacking类似
  - 不需要对训练集创建折外预测
  - 但需创建一个小的留出集（probe），用于训练stacker模型



- Train：训练基模型
- Probe：训练融合模型
- Qualifying：测试集

- Blending只使用了整体中数据一部分，最终的模型有可能对留出集过拟合
- Stacking使用交叉验证比使用单一留出集更加稳健 (但需要在更多折上进行计算)

Combining Predictions for Accurate Recommender Systems, Michael Jahrer, Andreas Töscher, Robert Legenstein, KDD2010

The End