

## 6. 模型融合

### 6.1 常见方法

- 平均
  - 简单平均法
  - 加权平均法
- 投票
  - 简单投票法
  - 加权投票法
- 综合
  - 排序融合
  - Log 融合
- stacking: 构建多层模型, 利用预测结果再拟合预测
- blending: 选取部分数据预测得到结果作为新特征带入剩下的数据中预测
- boosting/bagging

```
lgb_score_mean: 0.7311197790922761
lgb_score_std: 0.001507802995682687
xgb_score_mean: 0.7328694823569253
xgb_score_std: 0.00159958725807308
cat_score_mean: 0.7318428632974363
cat_score_std: 0.0018918585561348224

rh_test = 0.3*lgb_test+0.4*xgb_test+0.3*cat_test
```

图 6.1 使用加权平均进行模型融合调参后模型于验证集上结果

### 6.2 stacking & blending

Stacking 是把若干基学习器的预测结果作为新的训练集来训练一个学习器。Blending 是将预测值作为新的特征与原特征合并, 构成新的特征用于预测。两者的差异如下:

- stacking
  - stacking 的两层使用数据不同, 可以避免信息泄露;
  - 无需共享 seed;
- blending
  - blending 将数据分为两部分, 在最后预测时部分信息将被忽略;
  - 第二层数据较少, 易产生过拟合。