**task2**

**1、特征**

(1)特征选择

(2)特征构造

点击特征排序：每行数据app\_start\_day的rank值

背景：app点击特征受限于时间，可能会出现train 数据中点击高峰在前几天，test 数据点击高峰在中间几天，这样可能会导致模型误算结果。

处理：将单条app点击特征，按照点击次数排序，这样可以避免点击特征点击高峰不一致的情况。

**2、Regression**

Model 1：筛选特征 => lightgbm

Model 2：筛选特征+构造特征 => lightgbm

Model 3：筛选特征+构造特征 => lbt

Combination: (regression1+regression2+regression3)/3

Postprocess: if regression<1, regression = 0.

多回归器的结果融合，可以非常有效的减小预测偏差！

**3、Classification**

Model 1：筛选特征 => gbdt

Model 2：全特征 => lightgbm

Model 3：全特征 + sample weights(square of log(amount)) => lightgbm

并集或交集：Model 1 + Model 2 + Model 3

**4、Output**

Classification + Regression => result