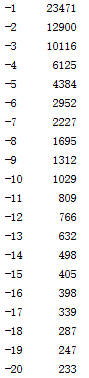
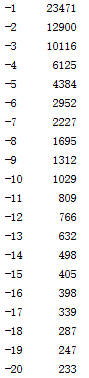
1. Train类用户的总用户量131209，平均每个用户的订单量15.7。test类75000
2. A类商品：购买两次以上；B类商品：只购买过一次；C类商品：没有买过
3. 注：merged\_prior.csv里的erordered
4. 只按两个feature：order\_num\_ratio、time\_ratio。来预测A类商品中，下次再购买，简单逻辑回归，准确率在0.839，但F1=0.322
5. 平均会有两件物品之前只购买过一次，但下次有购买
6. 把购买过的全预测中，下次买不买，F1也有0.74+
7. 在train里，但之前只购买过一次的商品，是在倒数第几次购物购买的

1. 线性回归预测用户下一次会买多少件商品
2. 创建features：总销量和再次购买销量。但放入feature来预测A类商品，准确率反而下降0.013，F1直接下降了0.158
3. 本来想把A类和B类分开预测的，但后来觉得一起预测可以筛掉一些没用的A类
4. feature1-5：order\_num\_ratio、time\_ratio、last\_order\_gap、total\_sale、rebuy\_total\_sale

|  |  |  |
| --- | --- | --- |
|  | Threshold | F1 |
| 只预测A类 | 0.5 | 0.322 |
| Feature前2 | 0.18 | 0.36306 |
| 前3 | 0.2 | 0.36623 |
| 前4 | 0.2 | 0.36498 |
| 前5 | 0.2 | 0.36601 |

1. feature总汇

是否在一个时间循环中？待定，因为在5天误差内预测中下次购买时间，预测率在0.38

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | 前n的F1 |
| 1 | order\_num\_ratio |  |  |
| 2 | time\_ratio |  |  |
| 3 | last\_order\_gap |  |  |
| 4 | rebuy\_product\_ratio | 商品再购率 | 0.37016 |
| 5 | rebuy\_depertment\_ratio | 所属department再购率 | 0.37088 |
| 6 | rebuy\_aisle\_ratio | 所属aisle再购率 | 0.37449 |
| 7 | max\_order\_num \_average | 用户总订单数 | 0.37485 |
| 8 | add\_to\_cart \_ratio \_average | 加入购物车的顺序占总购买数的比的均值 | 0.37475 |
| 9 | order\_hour\_dif\_sum | 购买时间与目标时间的差值总和 | 0.37847 |
| 10 | order\_dow\_dif\_sum | 购买星期日与目标星期日的差值总和 | 0.37911 |
| 11 |  |  |  |
| 12 |  |  |  |
| 13 |  |  |  |

1. 逻辑回归的C从1调到0.2；0.37475---🡪0.3749