1.用户特征

年龄，性别，在哪些店买东西

2.商家特征

有多少用户，有多少种类，有多少牌子，有多少物品，被购买物品多少次,点击数，喜欢，数。。。**时间特征，被重复购买用户数，重复购买用户的平均时间间隔**

3.商品特征

物品被购次数，物品点击数。。时间特征

4.牌子特征

牌子被购次数，。。牌子有多少个商品，有多少种类

5.种类特征

种类有多少个商品，有多少个牌子，喜欢数，购买数

6.商家与用户

用户在商家购买不同商品数，牌子数，种类数，商品总数，

7商品与用户关系

购买数，用户点击商品的次数，

8,用户与牌子关系

9用户与种类关系

10.商品与商家的关系

添加：

1、用户每个月的点击的次数，加入购物车的次数，

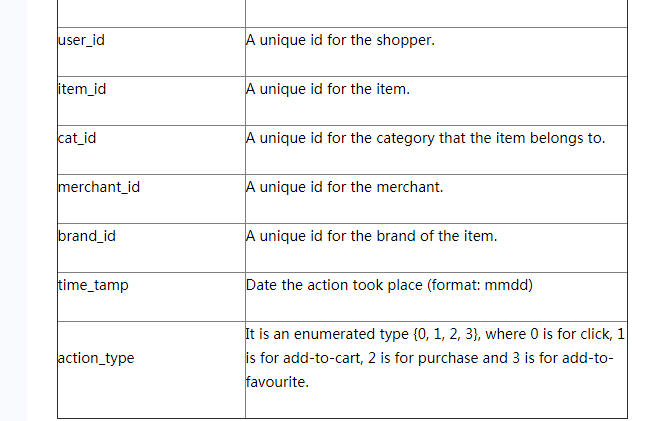
商人每个月被点击的次数

物品，种类和品牌

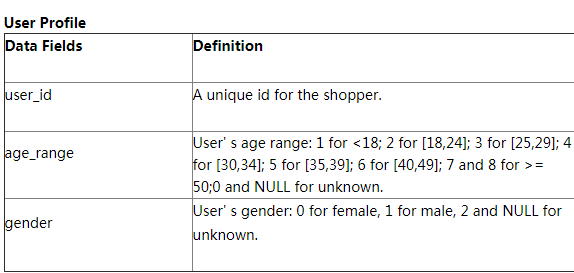
商人与物品

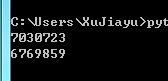
2、最近一个月的特征

3、趋势特征

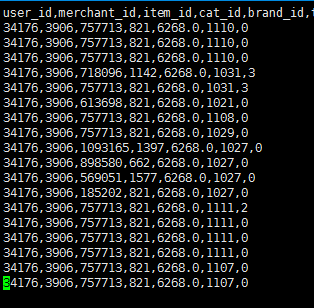


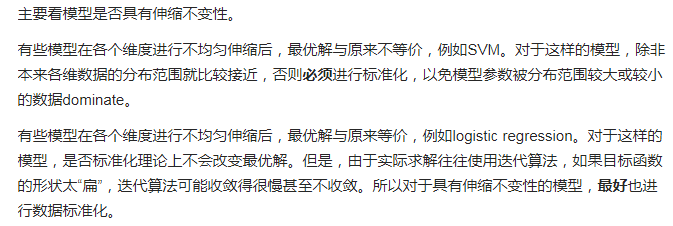


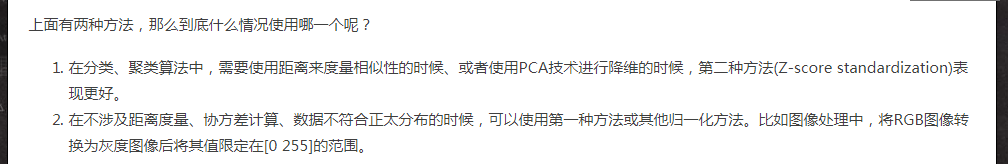




Label==-1的比例



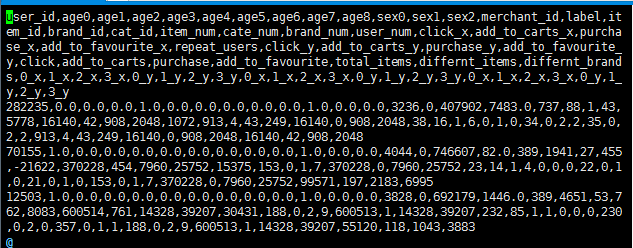




一个用户，一个商人，一个item可能有重复项，同样的actiontype或者不同的actiontype

Repeatbuyers在不同日期购买

用户，物品，商人，是不是会重购



Regression\_line

Slope:



归一化最近一个月的：

Merchant的brand market share 与用户的brand够买次数，向量相似度

两个商家之间，共同的用户，pca将维选取前10维

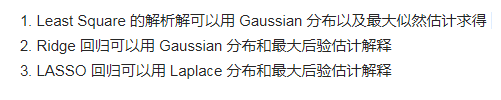
Lda:users,merchants

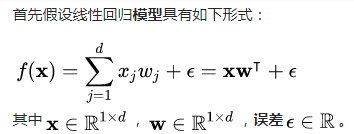
Users为doc,merchants为words

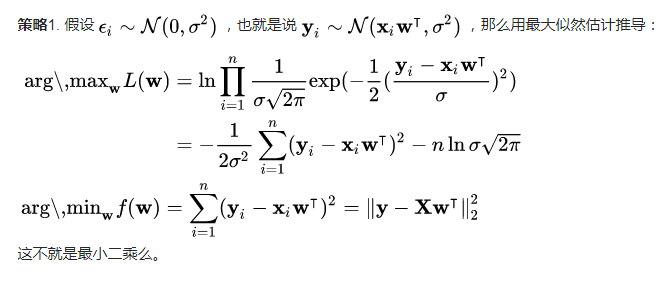
Merchants为doc,users为words

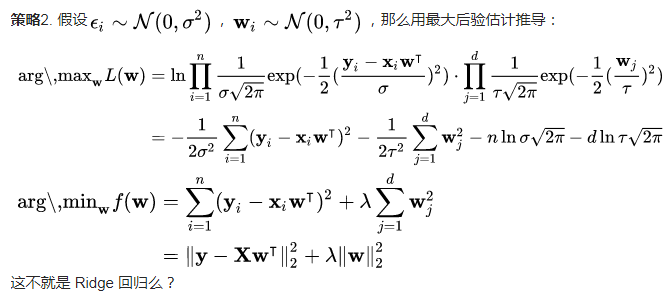
Penetration features 商品：购买的客户数，品牌，类别，商人

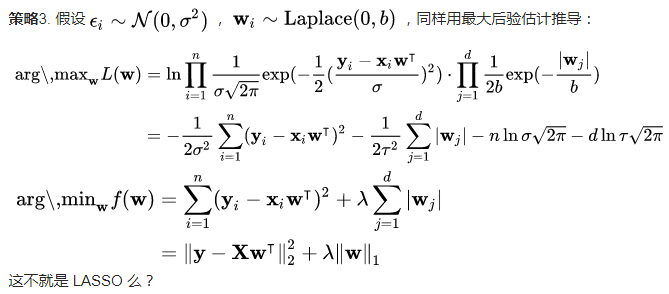
通过xgboost排序特征

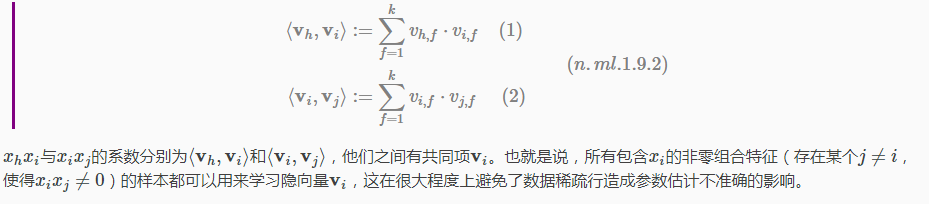






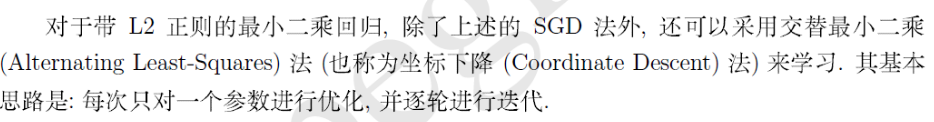
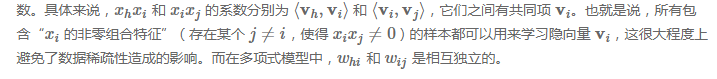






FM方程





FM回归适用于稀疏矩阵