

机器学习业务实践之路

课程4：推荐系统-基于用户和商品画像的商品推荐

阿里云 李博（傲海）

目录



1. 用户画像以及商品画像特征挖掘

2. 归一化介绍

3. 逻辑回归算法介绍

4. 基于画像的推荐系统实现

用户画像以及商品属性画像特征提取

user_id ▲	item_id ▲	active_type ▲	active_date ▲
10944750	13451	0	6月4日
10944750	13451	2	6月4日
10944750	13451	2	6月4日
10944750	13451	0	6月4日

数据源

字段名	含义	类型	描述
user_id	用户编号	string	购物的用户ID
item_id	物品编号	string	被购买物品的编号
active_type	购物行为	string	0表示点击，1表示购买，2表示收藏，3表示购物车
active_date	购物时间	string	购物发生的时间

字段信息

用户维度特征

- 总的购买量：用户的购物热情
- 总的点击量：表示用户的活跃度
- 总购买/总点击：用户的购物果断程度

商品维度特征

- 总的被购买量：商品热度
- 总的被点击量：产品是否吸引人
- 总被购买/总被点击：产品质量

归一化介绍

作用：去除量纲对于数据计算的影响

$y = (x - \text{MinValue}) / (\text{MaxValue} - \text{MinValue})$

_c2 ▲	pm10 ▼	so2 ▲	co ▲	no2 ▲
1	363	41	4.2	119
1	333	41	4.2	117
1	287	41	4	113
1	249	39	3.8	107
0	222	37	3.4	104
0	207	35	3.2	101

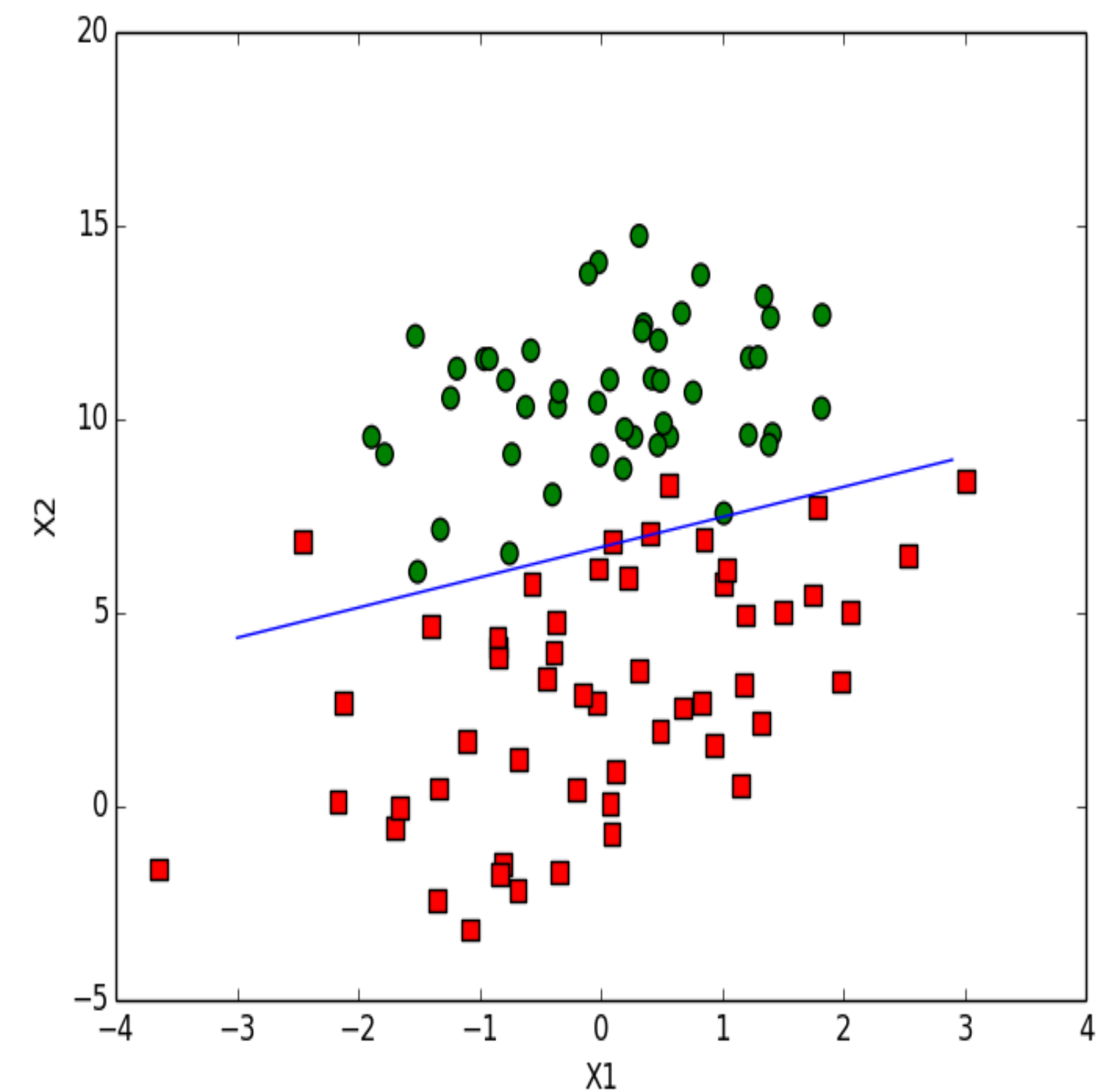
归一化



_c2 ▲	pm10 ▼	so2 ▲	co ▲	no2 ▲
1	0.7442827...	0.534246...	0.89130...	0.732484076433121
1	0.6819126...	0.534246...	0.89130...	0.7197452229299363
1	0.5862785...	0.534246...	0.84782...	0.6942675159235668
1	0.5072765...	0.506849...	0.80434...	0.6560509554140127
0	0.4511434...	0.479452...	0.71739...	0.6369426751592356
0	0.4199584...	0.452054...	0.67391...	0.6178343949044586

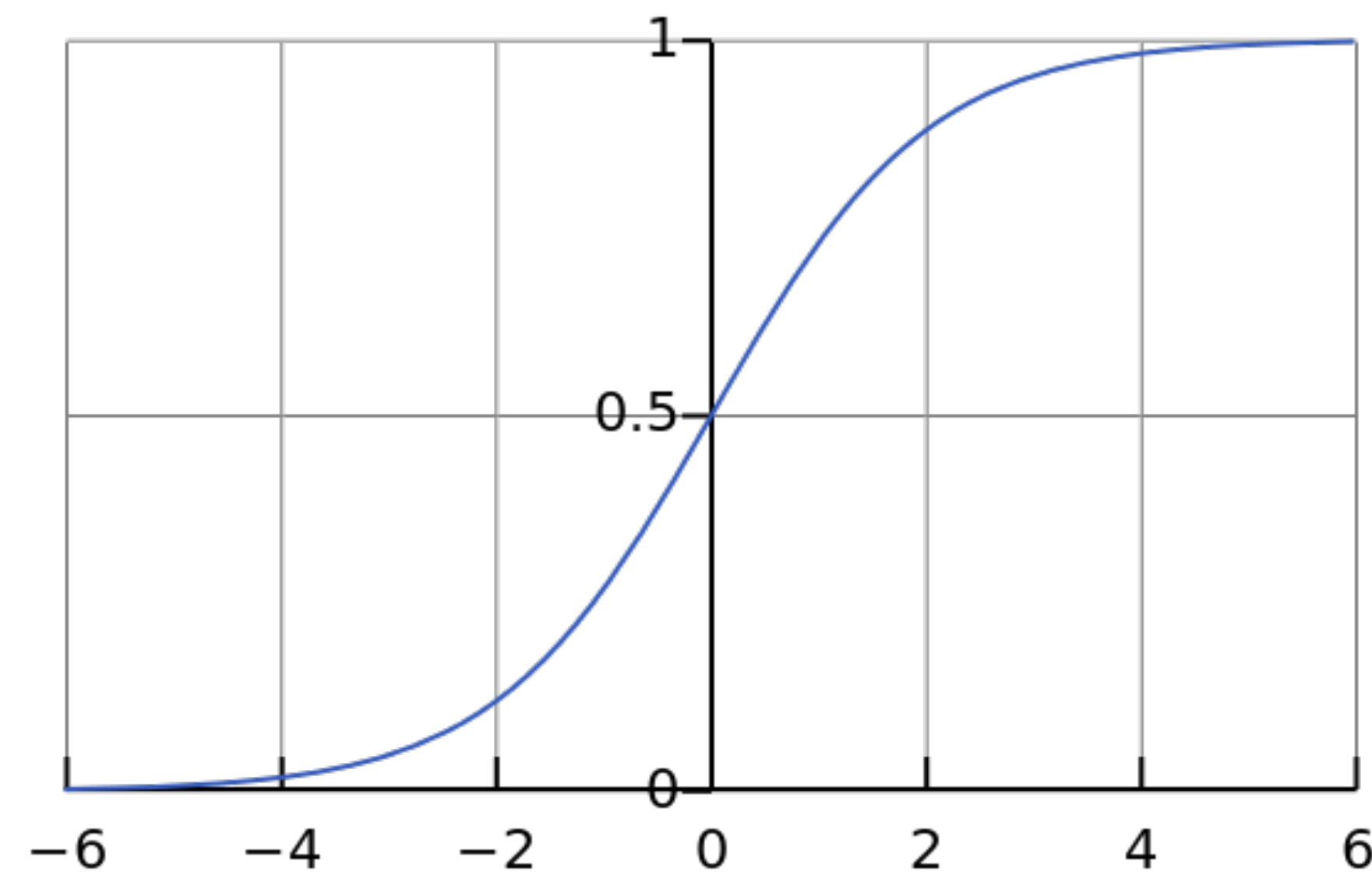
逻辑回归算法介绍1

逻辑回归（Logistic Regression, LR）是一种广义的线性回归分析模型，属于监督学习算法。逻辑回归可以用在回归、二分类、多分类等问题上，但是最常用的还是二分类。



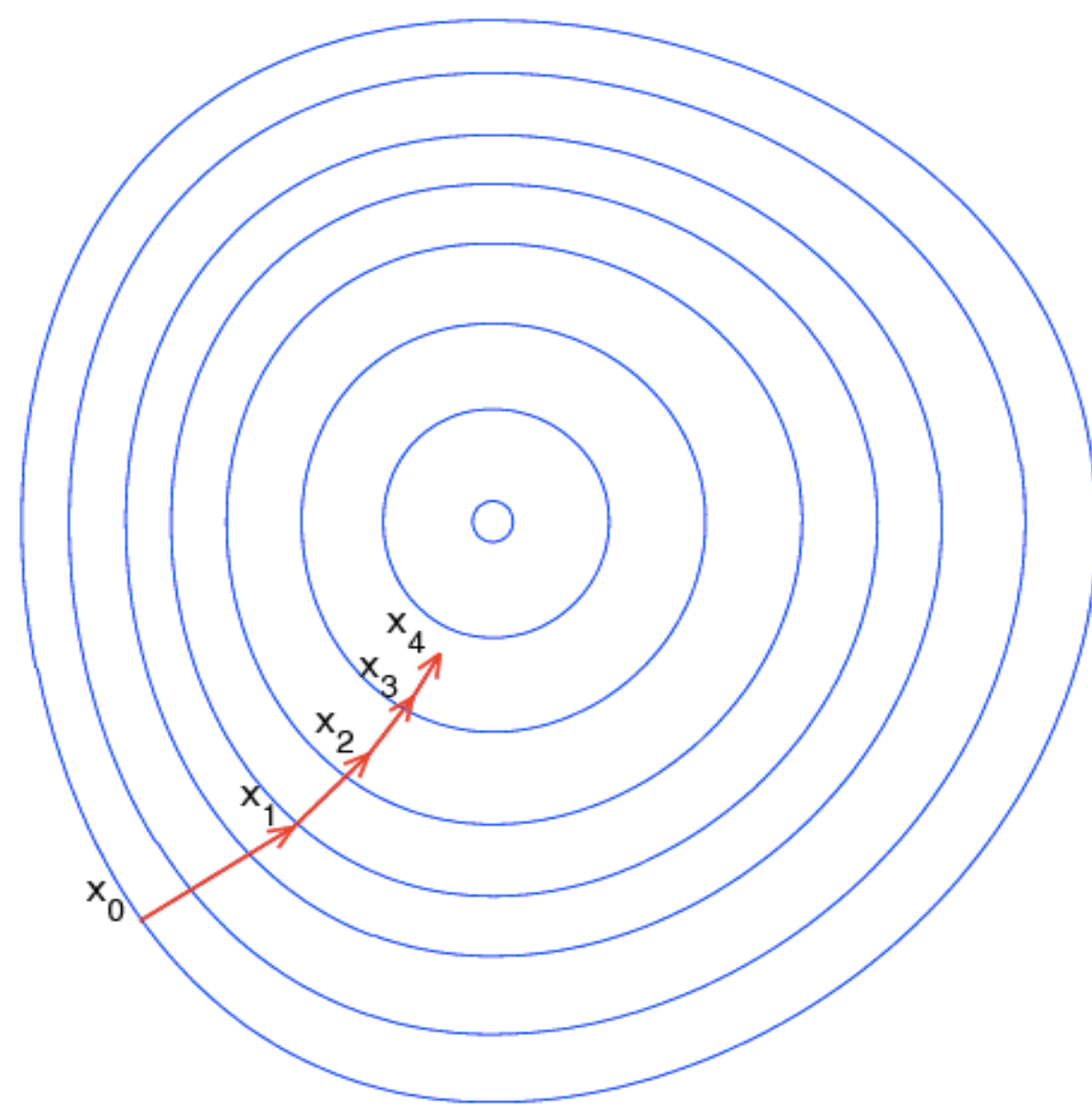
逻辑回归算法介绍-Sigmoid函数

$$S(t) = \frac{1}{1 + e^{-(w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \dots + C)}}$$



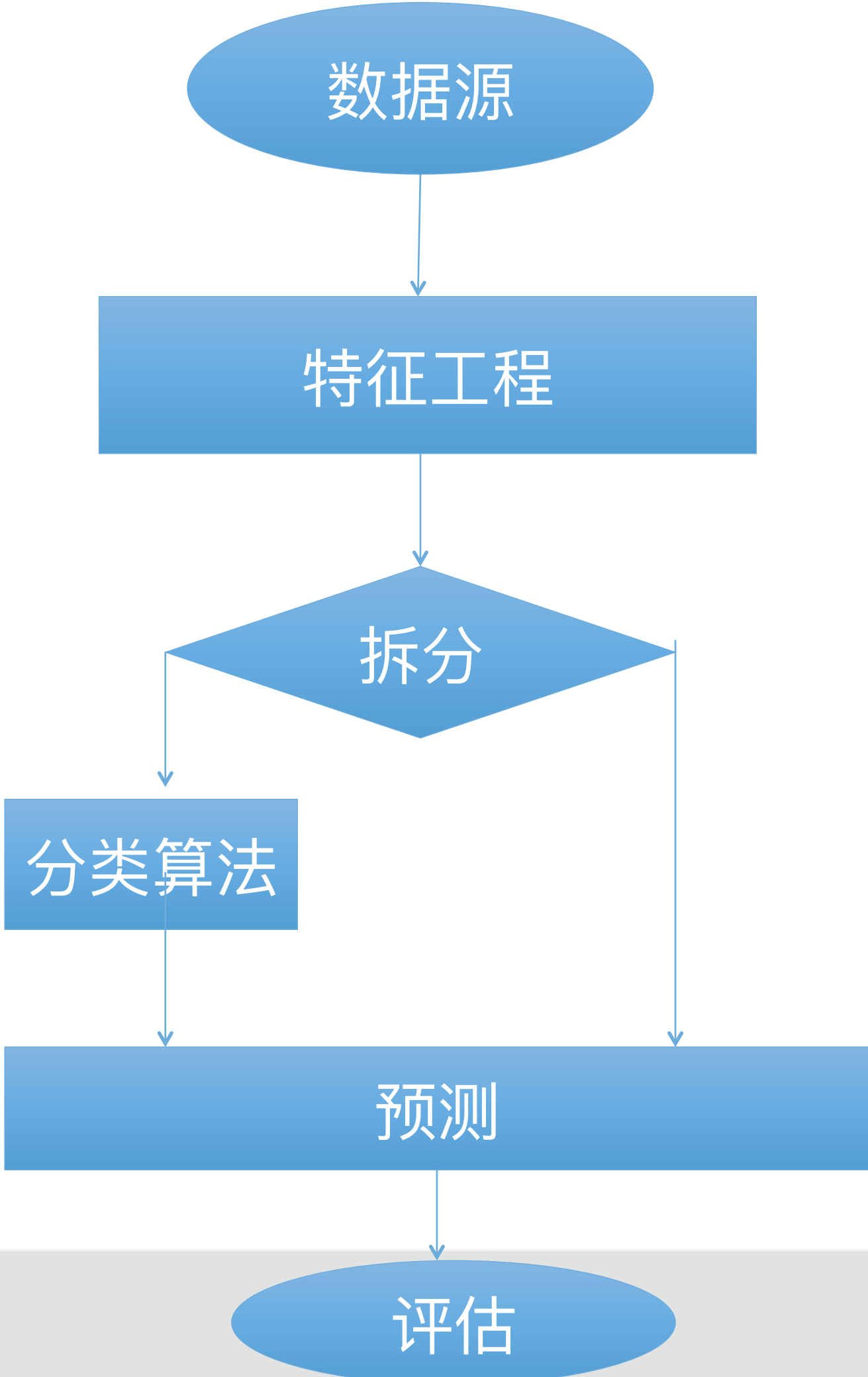
逻辑回归算法介绍-梯度下降

$$W^{t+1} = W^t - \partial * X^T * \text{error}$$

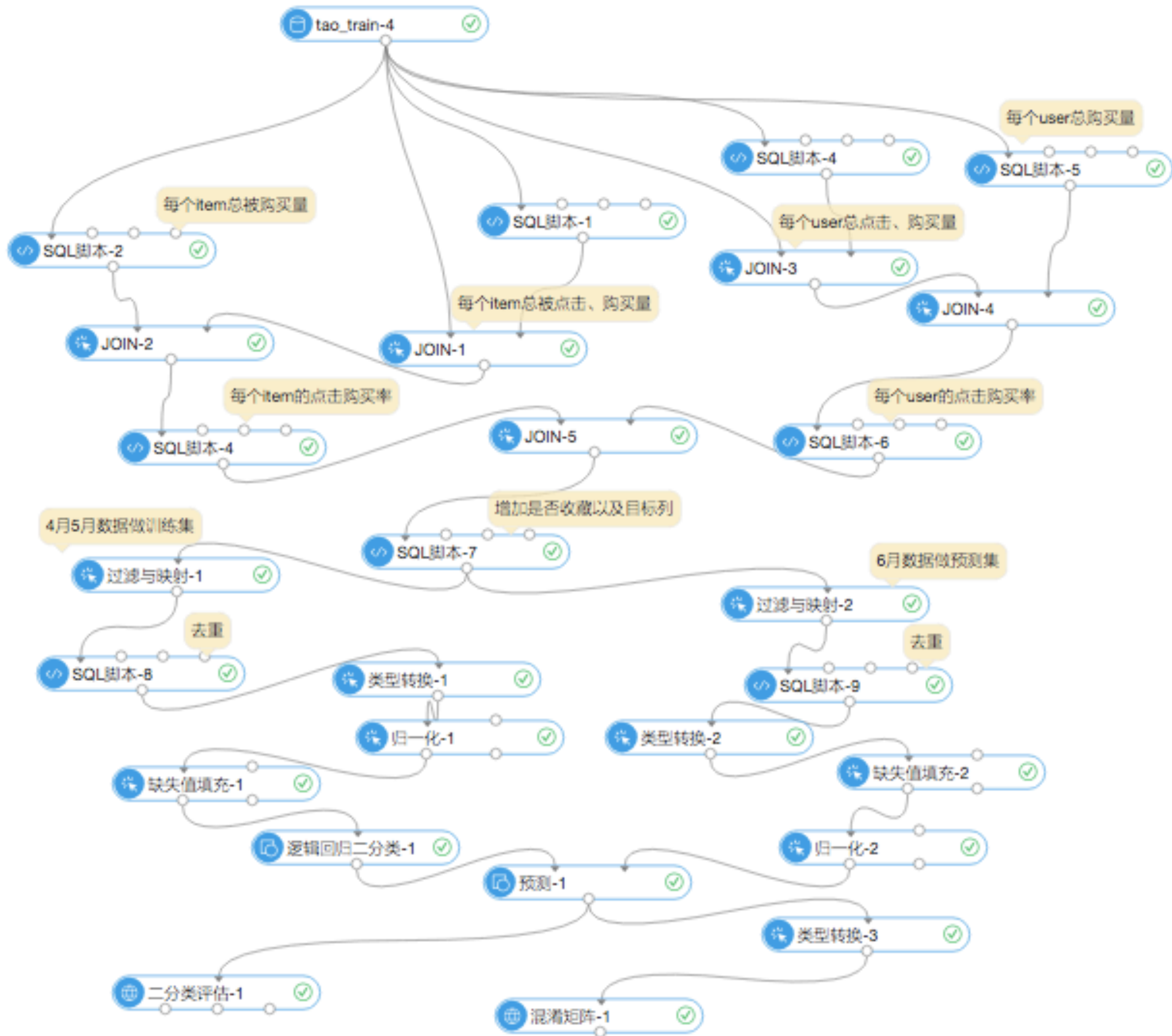


基于画像特征的商品推荐

基于画像特征的推荐架构图



PAI实现逻辑图



相关资料

推荐学习材料：

- 《机器学习实践》
- 《统计学习方法》
- 吴恩达的机器学习相关课程

推荐实验环境：机器学习PAI <https://data.aliyun.com/product/learn>

实验本案例： <https://yq.aliyun.com/articles/79297>

我的个人微信公众号（与我交流）：凡人机器学习

为了无法计算的价值 |  阿里云

