

第二课时：机器学习线下演练

TIANCHI天池

主讲人：implus

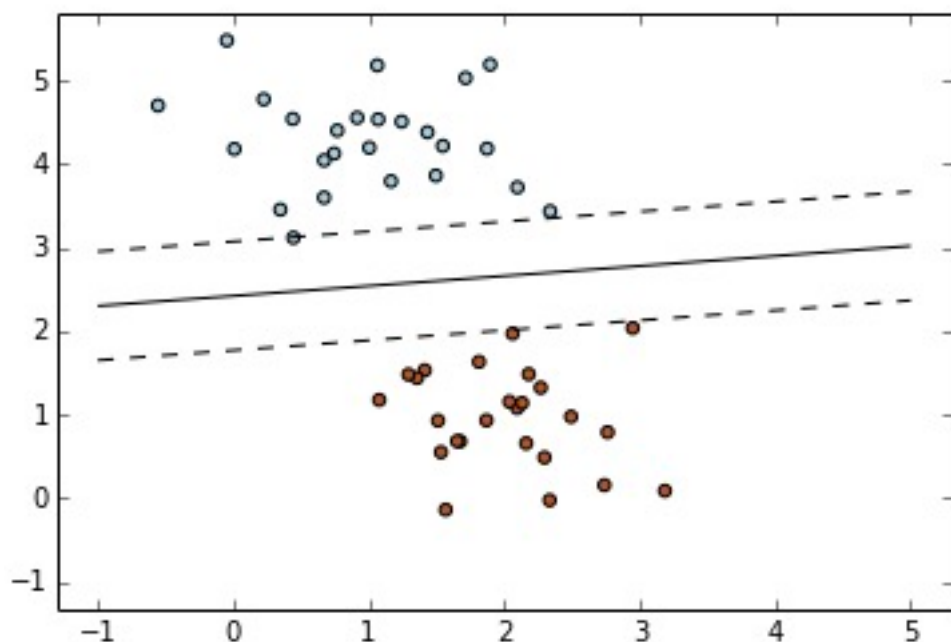


- 2015年天池大数据竞赛移动推荐算法全国冠军
- 2013 ~ 2014年ACM/ICPC亚洲区编程现场赛银奖
- 曾在SenseTime，现在MSRA实习
- 来自南京理工大学

提纲

1. 机器学习是如何解决二分类问题的
2. 我们如何将购买预测的问题转化为二分类问题
3. 我们线下如何实现

机器学习是如何解决二分类问题的



研究对象：点

对象特征：点坐标数值 (X_0, X_1)

对象标签：蓝色 (0) 或者红色 (1)

Logistic Regression

建模：

$$p(y=1|\mathbf{x};\theta) = \sigma(\theta^T \mathbf{x}) = \frac{1}{1+\exp(-\theta^T \mathbf{x})}$$

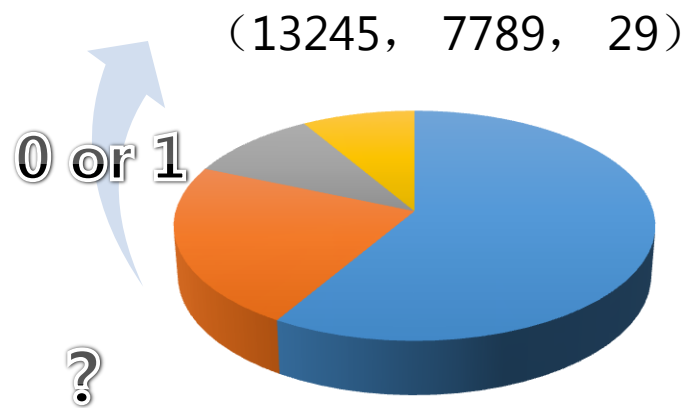
目标：

$$L(\theta) = \prod p(y^i=1|x^i)^{y^i} (1-p(y^i=1|x^i))^{1-y^i}$$

我们如何将购买预测的问题转化为二分类问题

给定0~30天的用户行为数据，预测31天的用户购买行为。

研究对象：点	→	(用户，商品，考察日)
对象特征：点坐标数值 (X0, X1)	→	在考察日之前该用户、该商品相关的各种行为的统计量
对象标签：蓝色 (0) 或者红色 (1)	→	该用户在考察日购买了该商品 (1) 或者没购买该商品 (0)



X0: 考察日前一天该用户对该商品进行了多少次浏览操作
(有多少条用户为13245，商品为7789，日期为28天，操作为浏览的记录)

X1: 考察日前一天该用户对该商品是否进行了加入购物车操作
(用户13245是否在28天将商品7789加入了购物车)

X2: 考察日一周该用户是否购买了该商品
(用户13245是否在第22天到第28天中对商品7789有购买的记录)

.....

(54312, 9383, 31) X0 | X1 | X2 ...

我们线下如何实现 —— python sklearn

1. 推荐使用 linux ubuntu, mac os
2. 更多的参考资料:
 1. Logistic Regression: <http://blog.csdn.net/zouxy09/article/details/20319673>
 2. Python basic: <http://www.dotnetperls.com/python>
 3. Sklearn with Python: <http://kukuruku.co/hub/python/introduction-to-machine-learning-with-python-andscikit-learn>