Personal Recommendation Algorithm

Main Flow

- LR(logistic regression)背景知识介绍
- LR算法数学原理解析
- 样本选择与特征构建基本知识介绍

LR背景知识介绍

- 点击率预估与分类模型
- 什么是LR
- Sigmoid函数

Example

- LR模型工作流程

买礼物	说早安	陪吃饭	女朋友开心
1	1	1	1
0	0	1	0
1	1	0	1

LR Model训练总流程

- 从Log中获取训练样本与特征
- Model 参数学习
- Model 预测

LR Model优缺点总结

- 易于理解，计算代价小
- 容易欠拟合，需要特征工程

Personal Recommendation Algorithm

Sigmoid Function

- 阶跃函数及其导数

$$f(x)=\frac{1}{1+\exp(-x)}$$

$$f'(x)=\frac{\exp(-x)}{(1+\exp(-x))^2}$$

$$f'(x)=\frac{1}{1+\exp(-x)}*\frac{1+\exp(-x)-1}{1+\exp(-x)}$$

LR Model Function

- model function

$$w = w_1 \times x_1 + w_2 \times x_2 + \square + w_n \times x_n$$

$$y = sigmoid(w)$$

Loss Function

- Loss function

$$loss = \log \prod_{i=1}^{n} p(y_i | x_i)$$

$$p(y_i | x_i) = h_w(x_i)^{y_i} (1 - h_w(x_i))^{1-y_i}$$

$$loss = -\left(y_i \log h_w(x_i) + (1 - y_i) \log(1 - h_w(x_i)) \right)$$

梯度

- 梯度

$$\frac{\partial loss}{\partial w_j} = \frac{\partial loss}{\partial h_w(x_i)} \frac{\partial h_w(x_i)}{\partial w} \frac{\partial w}{\partial w_j}$$

$$\frac{\partial loss}{\partial h_w(x_i)} = -\left(\frac{y_i}{h_w(x_i)} + \frac{y_i - 1}{1 - h_w(x_i)} \right)$$

$$\frac{\partial h_w(x_i)}{\partial w} \frac{\partial w}{\partial w_j} = h_w(x_i)\left(1 - h_w(x_i)\right) x_i^j$$

梯度下降

- 梯度下降

$$\frac{\partial loss}{\partial w_j} = \left(h_w\left(x_i\right) - y_i\right)x_i^j$$

$$w_j = w_j - \alpha\frac{\partial loss}{\partial w_j}$$

正则化

- 什么是过拟合
- L1正则化与L2正则化

$$loss_new = loss + \alpha \sum_{i=1}^{n} |w_i|$$

$$loss_new = loss + \alpha |w|^2$$

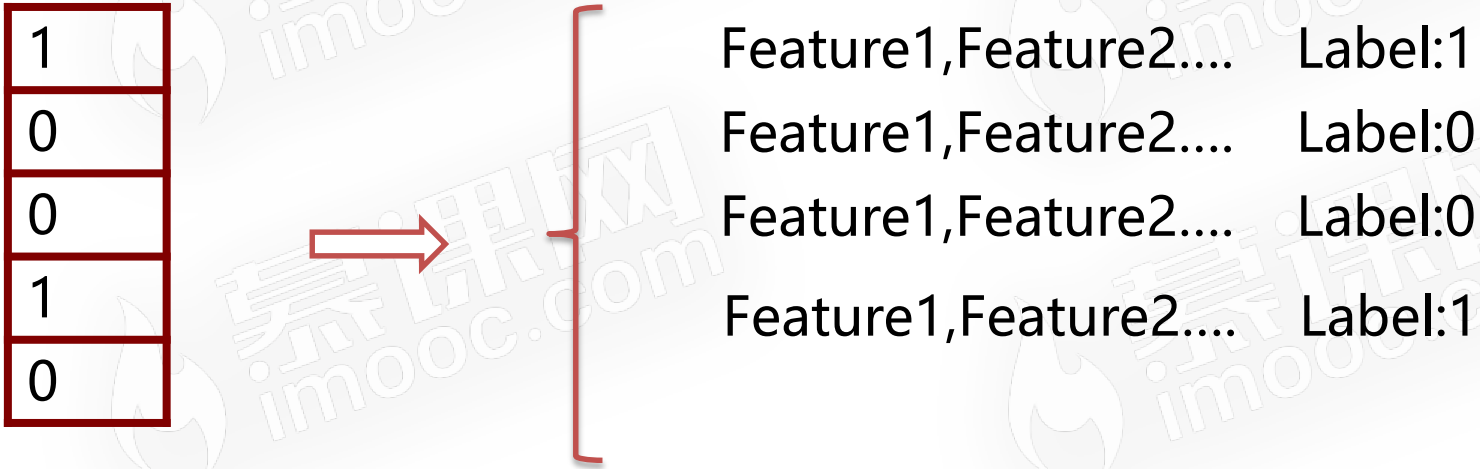
Personal Recommendation Algorithm

Corpus

- 样本选择规则
- 样本过滤规则

慕课网 imooc.com

Example


1
0
0
1
0
Feature1,Feature2.... Label:1
Feature1,Feature2.... Label:0
Feature1,Feature2.... Label:0
Feature1,Feature2.... Label:1

Feature

- 特征的统计与分析
- 特征的选择
- 特征的预处理

慕课网 imooc.com

Example

Hour：23 ⟹ 0.25分位点：18
0.50分位点：25
0.75分位点：40 ⟹ [0,1,0,0]

Country：China ⟹ [China, USA, Japan] ⟹ [1,0,0]