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	0800	1	0
1	TEOU-	0	

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

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

$$y = sigmoid(w)$$

Loss Function

Loss function

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

$$p(y_{i}|x_{i}) = h_{w}(x_{i})^{y_{i}} (1 - h_{w}(x_{i}))^{1-y_{i}}$$

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

梯度

• 梯度

$$\frac{\partial loss}{\partial w_{j}} = \frac{\partial loss}{\partial h_{w}(x_{i})} \frac{\partial h_{w}(x_{i})}{\partial w} \frac{\partial w}{\partial w_{j}}$$

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

$$\frac{\partial h_{w}(x_{i})}{\partial w} \frac{\partial w}{\partial w_{j}} = h_{w}(x_{i}) (1 - h_{w}(x_{i})) x_{i}^{j}$$

梯度下降

• 梯度下降

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

$$w_{j} = w_{j} - \alpha \frac{\partial loss}{\partial w_{j}}$$

正则化

- 什么是过拟合
- L1正则化与L2正则化

$$loss_new = loss + \alpha \sum_{i=1}^{n} |w_i| \quad loss_new = loss + \alpha |w|^2$$







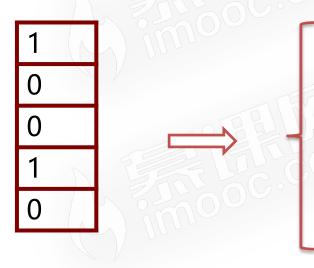


Personal Recommendation Algorithm

Corpus

- 样本选择规则
- 样本过滤规则

Example



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

Example

0.25分位点:18

Hour : 23 **□**

0.75分位点:40

Country: China (China, USA, Japan) (1,0,0)