

# Logistic Regression

给定数据  $X = \{x_1, x_2, \dots\}$ ,  $Y = \{y_1, y_2, \dots\}$

考虑二分类任务, 即  $y_i \in \{0, 1\}$ ,  $i = 1, 2, \dots$ ,

## 假设函数(Hypothesis function)

假设函数就是其基本模型, 如下:

$$h_{\theta}(x) = g(\theta^T x)$$

其中  $\theta^T x = w^T x + b$ , 而  $g(z) = \frac{1}{1+e^{-z}}$  为 *sigmoid* 函数, 也称激活函数。

## 损失函数

损失函数又叫代价函数, 用于衡量模型的好坏, 这里可以用极大似然估计法来定义损失函数。

似然与概率的区别以及什么是极大似然估计, [一文搞懂极大似然估计](#)

代价函数可定义为极大似然估计, 即  $L(\theta) = \prod_{i=1} p(y_i = 1|x_i) = h_{\theta}(x_1)(1 - h_{\theta}(x_2)) \dots$ , 其中  $x_1$  对应的标签  $y_1 = 1$ ,  $x_2$  对应的标签  $y_2 = 0$ , 即设定正例的概率为  $h_{\theta}(x_i)$ :

$$p(y_i = 1|x_i) = h_{\theta}(x_i)$$

$$p(y_i = 0|x_i) = 1 - h_{\theta}(x_i)$$

根据极大似然估计原理, 我们的目标是

$$\theta^* = \arg \max_{\theta} L(\theta)$$

为了简化运算, 两边加对数, 得到

$$\theta^* = \arg \max_{\theta} L(\theta) \Rightarrow \theta^* = \arg \min_{\theta} -\ln(L(\theta))$$

化简可得(这里只为了写代码, 具体推导参考西瓜书):

$$-\ln(L(\theta)) = \ell(\theta) = \sum_{i=1} (-y_i \theta^T x_i + \ln(1 + e^{\theta^T x_i}))$$

## 求解: 梯度下降

根据凸优化理论, 该函数可以由梯度下降法, 牛顿法得出最优解。

对于梯度下降来讲, 其中  $\eta$  为学习率:

$$\theta^{t+1} = \theta^t - \eta \frac{\partial \ell(\theta)}{\partial \theta}$$

其中

$$\frac{\partial \ell(\theta)}{\partial \theta} = \sum_{i=1} (-y_i x_i + \frac{e^{\theta^T x_i} x_i}{1 + e^{\theta^T x_i}}) = \sum_{i=1} x_i (-y_i + h_{\theta}(x_i)) = \sum_{i=1} x_i (-error)$$

这里梯度上升更方便点:

$$\theta^{t+1} = \theta^t + \eta \frac{-\partial \ell(\theta)}{\partial \theta}$$

其中

$$\frac{-\partial \ell(\theta)}{\partial \theta} = \sum_{i=1} (y_i x_i - \frac{e^{\theta^T x_i} x_i}{1 + e^{\theta^T x_i}}) = \sum_{i=1} x_i (y_i - h_{\theta}(x_i)) = \sum_{i=1} x_i * error$$

## 伪代码

---

训练算法如下：

- 输入：训练数据  $X = \{x_1, x_2, \dots, x_n\}$ , 训练标签  $Y = \{y_1, y_2, \dots\}$ , 注意均为矩阵形式
- 输出：训练好的模型参数  $\theta$ , 或者  $h_{\theta}(x)$
- 初始化模型参数  $\theta$ , 迭代次数  $n\_iters$ , 学习率  $\eta$
- **FOR**  $i\_iter$  in range( $n\_iters$ )
  - **FOR**  $i$  in range( $n$ )  $\rightarrow n = len(X)$ 
    - $error = y_i - h_{\theta}(x_i)$
    - $grad = error * x_i$
    - $\theta \leftarrow \theta + \eta * grad$   $\rightarrow$  梯度上升
    - **END FOR**
- **END FOR**

## Refs

---

西瓜书

李宏毅笔记