## Softmax回归原理总结

## 【参考资料】

Softmax回归

## 1. Softmax函数

softmax函数是logistic函数的一般形式,作用是将取值范围在 $(-\infty, +\infty)$ 上的K维向量,转换为取值范围在(0,1]的 K维向量,即一个概率分布的形式。

$$h_{ heta}(x) = egin{bmatrix} P(y = 1 | x; heta) \ P(y = 2 | x; heta) \ dots \ P(y = K | x; heta) \end{bmatrix} = rac{1}{\sum_{j=1}^K \exp\left( heta_j^T x
ight)} egin{bmatrix} \exp\left( heta_1^T x
ight) \ \exp\left( heta_2^T x
ight) \ dots \ \exp\left( heta_1^T x
ight) \end{bmatrix}$$

一般形式:

$$p\left(y^{(i)} = j|x^{(i)}; heta
ight) = rac{\exp\left( heta_j^T x^{(i)}
ight)}{\sum_{l=1}^K \exp\left( heta_l^T x^{(i)}
ight)}$$

## 2. 代价函数

softmax回归是logistic回归在多分类问题上的拓展形式,其代价函数为

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^{m} \sum_{j=1}^{k} 1\left\{ y^{(i)} = j \right\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}} \right]$$
(2.1)

其中1{.}是指示函数。

式 (2.1) 又被称为多分类交叉熵损失函数,它是logistic回归代价函数的推广,logistic回归的代价函数可以改写为

$$egin{aligned} J( heta) &= -rac{1}{m} \left[ \sum_{i=1}^m \left( 1 - y^{(i)} 
ight) \log \left( 1 - h_ heta \left( x^{(i)} 
ight) 
ight) + y^{(i)} \log h_ heta \left( x^{(i)} 
ight) 
ight] \ &= -rac{1}{m} \left[ \sum_{i=1}^m \sum_{j=0}^1 1 \left\{ y^{(i)} = j 
ight\} \log p \left( y^{(i)} = j | x^{(i)}; heta 
ight) 
ight] \end{aligned}$$

softmax回归中,将x分类为类别i的概率是:

$$p\left(y^{(i)}=j|x^{(i)}; heta
ight)=rac{e^{ heta_{j}^{T}x^{(i)}}}{\sum_{l=1}^{k}e^{ heta_{l}^{T}x^{(i)}}}$$

对式 (2.1) 求导, 结果为

$$abla_{ heta_j}J( heta) = -rac{1}{m}\sum_{i=1}^m \left[x^{(i)}\left(1\left\{y^{(i)}=j
ight\} - p\left(y^{(i)}=j|x^{(i)}; heta
ight)
ight)
ight]$$

之后可以利用梯度下降法来求解参数。