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。

$$h_\theta(x)=\begin{bmatrix} P(y=1|x;\theta) \\ P(y=2|x;\theta) \\ \vdots \\ P(y=K|x;\theta) \end{bmatrix}=\frac{1}{\sum_{j=1}^{K}\exp\left(\theta_j^T x\right)}\begin{bmatrix} \exp\left(\theta_1^T x\right) \\ \exp\left(\theta_2^T x\right) \\ \vdots \\ \exp\left(\theta_K^T x\right) \end{bmatrix}$$

一般形式：

$$p\left(y^{(i)}=j|x^{(i)};\theta\right)=\frac{\exp\left(\theta_j^T x^{(i)}\right)}{\sum_{l=1}^{K}\exp\left(\theta_l^T x^{(i)}\right)}$$

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

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

其中$1\{\cdot\}$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

$$\begin{aligned} J(\theta)&=-\frac{1}{m}\left[\sum_{i=1}^{m}\left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)+y^{(i)}\log h_\theta\left(x^{(i)}\right)\right] \\ &=-\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=0}^{1}1\left\{y^{(i)}=j\right\}\log p\left(y^{(i)}=j|x^{(i)};\theta\right)\right] \end{aligned}$$

softmax回归中，将x分类为类别j的概率是：

$$p\left(y^{(i)}=j|x^{(i)};\theta\right)=\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k}e^{\theta_l^T x^{(i)}}}$$

对式（2.1）求导，结果为

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[x^{(i)}\left(1\left\{y^{(i)}=j\right\}-p\left(y^{(i)}=j|x^{(i)};\theta\right)\right)\right]$$

之后可以利用梯度下降法来求解参数。