SoftMax 回归

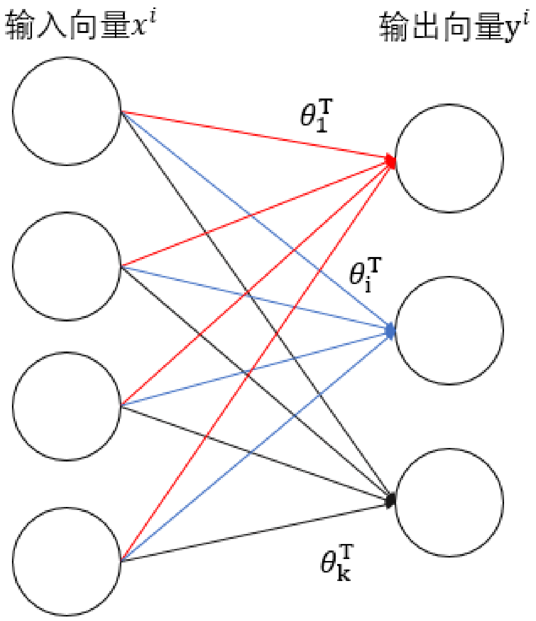
机器学习笔记 create by siwanghu v1.0

Softmax 回归模型是 logistic 回归模型在多分类问题上的推广，在多分类问题中，类标签可以取两个以上的值

假设训练集为：$\{(x^1,y^1),(x^2,y^2),\ldots,(x^m,y^m)\}$，$x^i$代表输入特征，是一个 n+1 维的向量，n 代表输入数据的特征数目（其中$x_0=1$，对应偏置），标签$y^i\epsilon(1,2,\ldots,k)$，代表标签 y 的取值集合。

对于给定的测试输入x^i，我们想用假设函数估算出x^i属于每个类别 j 的概率值$\mathrm{p}(y=j|x^i)$，因此我们的假设函数需要输出一个 k 维向量（向量元素和为 1）来表示这 k 个估计的概率值。

Softmax 模型可以认为是只有输入层和输出层，激活函数为$\mathrm{y}=e^x$的神经网络模型，θ_i^T是第 i 个输出神经元与各个输入神经元连接的权值值，以向量形式定义，维数与输入向量x^i相同



$$h(x^i)=\begin{bmatrix}p(y^i=1|x^i;\theta_1^T)\\ \ldots\\ p(y^i=k|x^i;\theta_k^T)\end{bmatrix}=\frac{1}{\sum_{\mathrm{j}=1}^{k}e^{\theta_j^T x^i}}\begin{bmatrix}e^{\theta_1^T x^i}\\ \ldots\\ e^{\theta_k^T x^i}\end{bmatrix}$$

$\frac{\mathbf{1}}{\sum_{\mathrm{j}=1}^{k}e^{\theta_j^T x^i}}$是对输出向量进行概率归一化，使得所有概率之和为 1

Softmax 的损失函数，为此，定义示性函数：

$$l\{值为真的表达式\} = 1$$

所以，损失函数为：

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

m 为训练数据批次的大小

$$\boxed{\begin{gathered}\frac{\partial \mathrm{L}(\theta)}{\partial \theta} = -\frac{1}{\mathrm{m}}\sum_{i=1}^{m}\left[x^i\left(l\{y^i = j\} - p(y^i = j|x^i;\theta)\right)\right] \\ p(y^i = j|x^i;\theta) = \frac{e^{\theta_j^T x^i}}{\sum_{\mathrm{l}=1}^{k} e^{\theta_l^T x^i}}\end{gathered}}$$