基于 softmax 函数的多分类模型

假设一个多分类模型，共输出 C 中类型，则对于单个样本的交叉熵损失函数为：

$$L=-\sum_{j=1}^{C} y_j \ln f\left(x_j\right)$$

上式中，y_j 表示该样本实际是否属于第 j 类，属于为 1，不属于为 0，x_j 表示样本在第 j 个输出节点(每个输出节点对应一类)的输入，$f(x_j)$（激活函数）表示该样本属于第 j 类的概率，即 softmax 函数：$f\left(x_j\right)=\frac{e^{x_j}}{\sum_{k=1}^{C} e^{x_k}}$，因此：

(1).当 i 等于 j 时：

$$\begin{aligned}\frac{\partial f\left(x_j\right)}{\partial x_i}=\frac{\partial f\left(x_j\right)}{\partial x_j}&=\frac{e^{x_j}\sum_{k=1}^{C} e^{x_k}-e^{x_j}e^{x_j}}{\left(\sum_{k=1}^{C} e^{x_k}\right)^2}\\&=\frac{e^{x_j}}{\sum_{k=1}^{C} e^{x_k}}-\left(\frac{e^{x_j}}{\sum_{k=1}^{C} e^{x_k}}\right)^2\\&=f\left(x_j\right)\left[1-f\left(x_j\right)\right]\end{aligned}$$

(2).当 i 不等于 j 时：

$$\begin{aligned}\frac{\partial f\left(x_j\right)}{\partial x_i}&=\frac{-e^{x_i}e^{x_j}}{\left(\sum_{k=1}^{C} e^{x_k}\right)^2}\\&=-\frac{e^{x_i}}{\sum_{k=1}^{C} e^{x_k}}\cdot\frac{e^{x_j}}{\sum_{k=1}^{C} e^{x_k}}\\&=-f\left(x_i\right)f\left(x_j\right)\end{aligned}$$

所以：

$$\frac{\partial f\left(x_j\right)}{\partial x_i}=\begin{cases}f\left(x_j\right)\left[1-f\left(x_j\right)\right] & i=j\\-f\left(x_i\right)f\left(x_j\right) & i\neq j\end{cases}$$

因此单个样本的损失函数 L 对 x_i 的偏导为：

$$\begin{aligned}\frac{\partial L}{\partial x_i} &= -\sum_{j=1}^{C} y_j \frac{1}{f\left(x_j\right)} \cdot \frac{\partial f\left(x_j\right)}{\partial x_i} \\ &= -y_i \frac{1}{f\left(x_i\right)} f\left(x_i\right)\left[1-f\left(x_i\right)\right] + \sum_{j=1,\, j\neq i}^{C} y_j \frac{1}{f\left(x_j\right)} f\left(x_i\right) f\left(x_j\right) \\ &= -y_i\left[1-f\left(x_i\right)\right] + \sum_{j=1,\, j\neq i}^{C} y_j f\left(x_i\right) \\ &= -y_i + \sum_{j=1}^{C} y_j f\left(x_i\right)\end{aligned}$$

由于在所有的类别中，只有一个类别为 1，其他都为 0，所以：

$$\frac{\partial L}{\partial x_i} = f\left(x_i\right) - y_i$$

因此，另 $\boldsymbol{x}=(x_1,x_2,\ldots,x_c)^T$，$\boldsymbol{x}$ 表示样本所有输出节点的输入向量，$\boldsymbol{y}=(y_1,y_2,\ldots,y_c)^T$，$\boldsymbol{y}$ 表示该样本实际类别向量，其中 $\boldsymbol{y}$ 里面只有一个元素为 1，其他都为 0，所以：

$$\frac{\partial L}{\partial \boldsymbol{x}} = f\left(\boldsymbol{x}\right) - \boldsymbol{y}$$