1.求交叉熵的反向传播函数

交叉熵损失函数常用于分类问题中，在二分类的问题中，由于只有两种情况1或者0，对应的概率为p,1-p。则表达式为：

$$L = -[y \cdot log(p) + (1-y) \cdot log(1-p)]$$

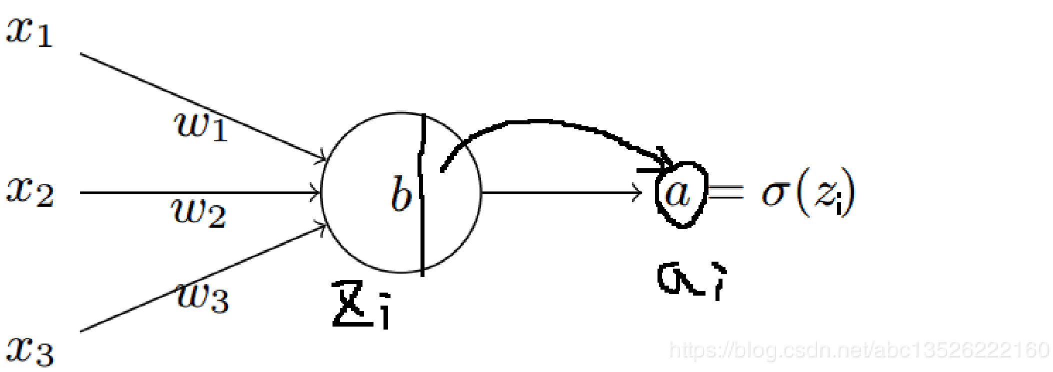
y为样本标签，p为样本预测为1或者正的概率。那么在多分类中的损失函数为：

$$L = -\sum_{i=1}^{n} y_i log(p_i)$$

n为类别数量。

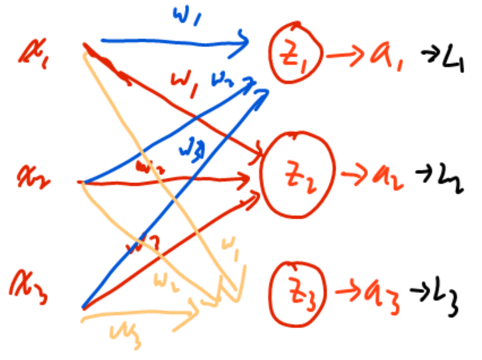
预测概率$a_i = p_i = e^{-(w^T x_i + b)} / \sum_k^n e^{-(w^T x_k + b)}$，为softmax函数。那么在神经网络中，将$a_i^l$表示第l层第i个神经元的输出，令$z_i^l = -(w^T x_i + b)$表示第l层第i个神经元的输入,即l-1层的输出，$\sum_k e^{z_k^l}$,表示第l层所有神经元的输入之和。则

$$a_i^l = e^{z_i^l} / \sum_k e^{z_k^l}$$



求什么值：在反向传播中依据微积分中的链式法则，沿着从输出层到输入层的顺序，依次计算并存储目标函数有关神经网络各层的中间变量以及参数的梯度，也就是每一层z_i的梯度，即$\frac{\partial L}{\partial z_i}$：

根据复合函数求导法则：$\frac{\partial L}{\partial z_i} = \sum_{j=1} \left(\frac{\partial L_j}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_i}\right)$，由于每一层神经元不止一个，所以j表示该层神经元下标，$L = L_1 + L_2 + \cdots + L_m$，m为神经元个数。



1) i=j时：

$$\frac{\partial L_j}{\partial a_j} = \frac{\partial(-y_j \ln a_j)}{\partial a_j} = -y_j \frac{1}{a_j}$$

$$\frac{\partial a_j}{\partial z_i} = \frac{\partial\left(\frac{e^{z_i}}{\sum_k e^{z_k}}\right)}{\partial z_i} = \frac{\sum_k e^{z_i} e^{z_k} - (e^{z_i})^2}{(\sum_k e^{z_k})^2}$$

$$= \frac{e^{z_i}}{\sum_k e^{z_k}}\left(1 - \frac{e^{z_i}}{\sum_k e^{z_k}}\right)$$

$$= a_i(1-a_i)$$

由于L包含所有神经元输出，当i=j时，分子分母都要对z_i求导，i≠j时只需分母对z_i求导

2) i≠j时：

$$\frac{\partial a_j}{\partial z_i} = \frac{\partial\left(\frac{e^{z_j}}{\sum_k e^{z_k}}\right)}{\partial z_i} = e^{z_j} \frac{-1}{(\sum_k e^{z_k})^2} \cdot z_i$$

$$= -\frac{e^{z_j}}{\sum_k e^{z_k}} \cdot \frac{e^{z_i}}{\sum_k e^{z_k}} = -a_i a_j$$

则 $\frac{\partial L}{\partial z_i} = \sum_j \left(\frac{\partial L_j}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_i}\right) = \sum_{j \neq i} \left(\frac{\partial L_j}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_i}\right) + \sum_{j=i} \left(\frac{\partial L_j}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_i}\right)$

$$= \sum_{j \neq i} -\frac{y_j}{a_j} \cdot (-a_i a_j) + \left(-\frac{y_i}{a_i}\right) \cdot a_i(1-a_i)$$

$$= \sum_{j \neq i} y_j a_i + y_i(a_i - 1)$$

$$= \sum_{j \neq i} y_j a_i + y_i a_i - y_i$$

$$= \sum_j a_i y_j - y_i$$

而对于 z_i，在分类中只有一个类别为1，其它全为0，因此 $\frac{\partial L}{\partial z_i} = a_i - y_i$

2.证明逻辑斯蒂是凸优化问题

逻辑函数的目标函数为$L = -\sum_i^n [y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i)]$

，其中y_i为样本标签取0或1，p_i为对应样本标签的概率，$p_i = 1/1 + e^{-(w^T x_i + b)}$,将参数(**w**,b)的b融入到**w**中，则$w = (w_0, w_1, w_2, \ldots, w_n)$,n为样本大小，$w_0$即为b的值，则$x_i = (1, x_{i1}, x_{i2} \ldots x_{in})$，则$p_i = 1/1 + e^{-w^T x_i}$

若要证明目标函数为凸函数，则对**w**二次求导后，二阶导函数大于0即可。

证明：令 $\delta(x) = \frac{1}{1+e^{-x}}$ $\quad \delta'(x) = \frac{0-(-1) \times e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2} = \delta(x) \cdot (1 - \delta(x))$

$$\frac{\partial L}{\partial w} = \frac{\partial \left[-\sum_{i=1}^n y_i \log \delta(w^T x_i) + (1-y_i) \log[1 - \delta(w^T x_i)]\right]}{\partial w}$$

$$= -\sum_{i=1}^n y_i \cdot \frac{1}{\cancel{\delta(w^T x_i)}} \cdot \cancel{\delta(w^T x_i)} \cdot [1 - \delta(w^T x_i)] \cdot x_i + (1-y_i) \cdot \frac{-\delta(w^T x_i)\cancel{[1-\delta(w^T x_i)]} \cdot x_i}{\cancel{1-\delta(w^T x_i)}}$$

$$= -\sum_{i=1}^n y_i [1 - \delta(w^T x_i)] x_i + (y_i - 1)\delta(w^T x_i) x_i$$

$$= -\sum_{i=1}^n x_i y_i - \cancel{x_i y_i \delta(w^T x_i)} + \cancel{x_i y_i \delta(w^T x_i)} - x_i \delta(w^T x_i)$$

$$= -\sum_{i=1}^n x_i [y_i - \delta(w^T x_i)]$$

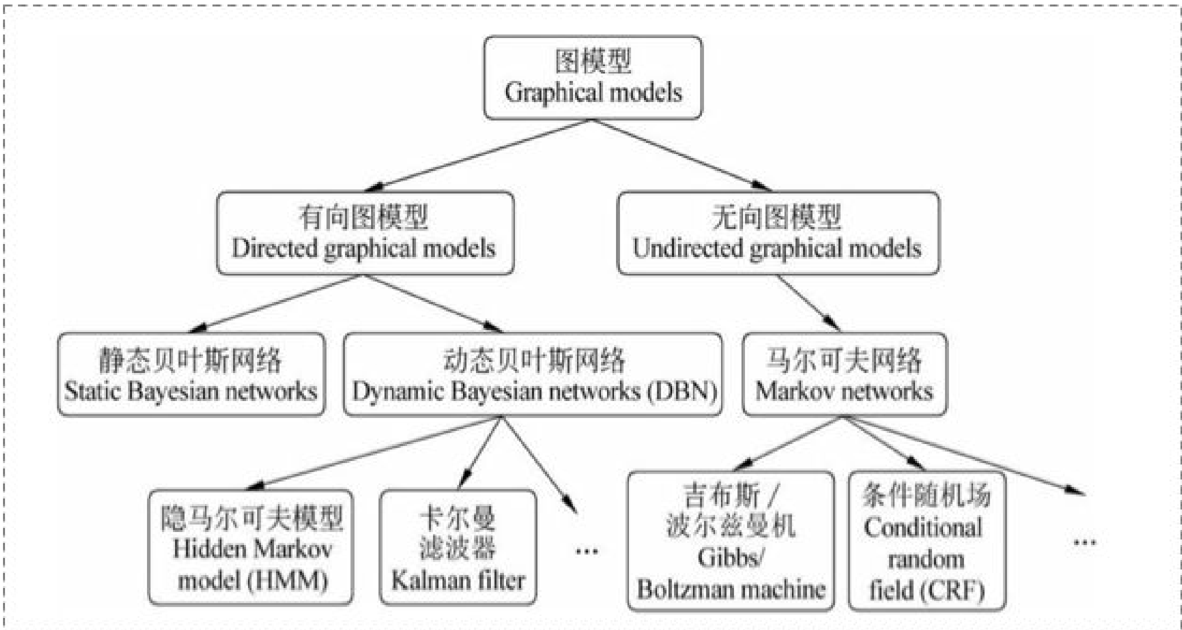
→ 去掉与w无关项

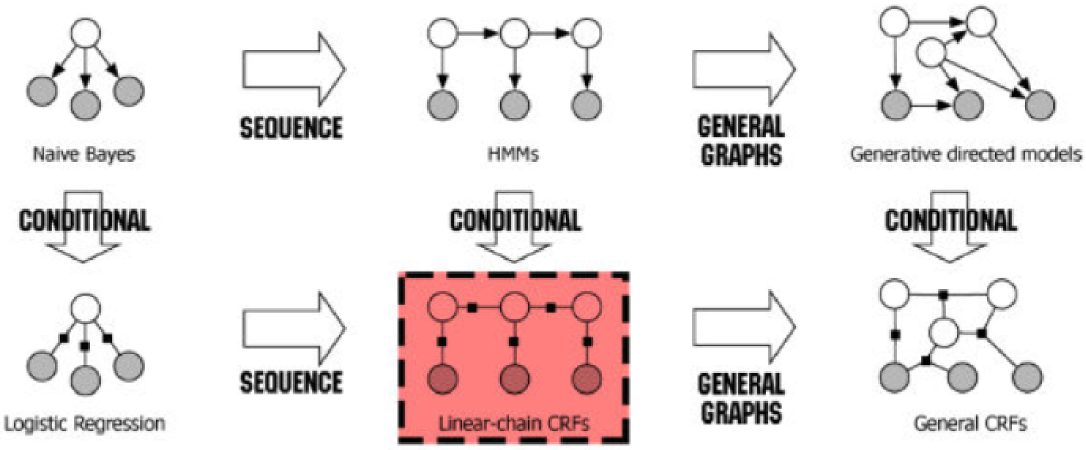
$$\frac{\partial^2 L}{\partial w^2} = \frac{\partial \left(\sum_{i=1}^n \delta(w^T x_i) x_i\right)}{\partial w} = x_i^2 \cdot \delta(w^T x_i) \cdot [1 - \delta(w^T x_i)]$$

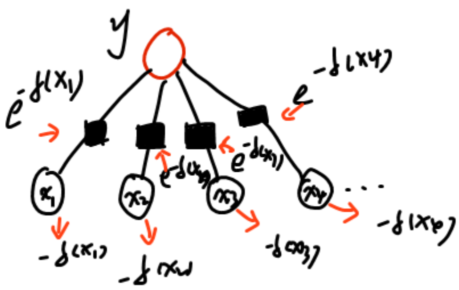
$\because 0 < \delta(w^T x_i) = 1/1+e^{-w^T x_i} < 1$，而 $x_i^2 > 0$ $\quad \therefore \frac{\partial^2 L}{\partial w^2} > 0$，则 logistic 为凸优化问题.

3.画出逻辑回归的概率图模型，并写出各点、边的势函数

在概率图模型中，根据生成模型、判别模型来分类，一般有向图为生成模型，无向图为判别模型







逻辑回归，$P(y=k|x,w)=\dfrac{e^{-(w^Tx_k+b)}}{\sum_{i=1}^{k}e^{-(w^Tx_i+b)}}$

将b整合到w中，令$\delta(x)=e^{-w^Tx}$

$$P(y=k|x,w)=\frac{e^{-\delta(x_k)}}{\sum_{i=1}^{k}e^{-\delta(x_i)}}$$