Softmax 回归

1. 线性回归

https://promhub.top/%E7%BA%BF%E6%80%A7%E5%9B%9E%E5%BD%92.html

2. 逻辑回归

在逻辑回归中可以通过一条直线来划分两种不同的类型，以此可以解决二分类问题。 在逻辑回归中通过表示为：

$$\hat{y} := \begin{cases} 1 \text{ if } f(x) > 0.5 \\ 0 \text{ if } f(x) \leq 0.5 \end{cases} \tag{2 - 1}$$

以输出值为 0.5 作为界限进行区分。

拟合的方程$f(x)$ 写为：

$$f(x) = w \cdot x \tag{2 - 2}$$

与线性回归相似，在单特征时有一个特征 w，一个输入 w，一个输出 y，偏置 b。

这样进而就可以由 $f(x)$ 写出一个 x 关于$\hat{y}$的表达式。

$$\hat{y} = \begin{cases} 1 \text{ if } x > f^{-1}(0.5) \\ 0 \text{ if } x \leq f^{-1}(0.5) \end{cases} \tag{2 - 3}$$

这样就可以通过输入 x 得出输出 y。

这个函数是间断的，不适用于使用梯度下降法进行处理，这样就可以使用更加平滑的 sigmond 函数来代替。sigmond 函数可以写为

$$\begin{gathered} \mathrm{f(z)} = \frac{1}{1 + e^{-z}} \\ z = w \cdot x \end{gathered} \tag{2 - 4}$$

2.1. 极大似然估计

概率描述的是一个事件发生的可能性，似然性是放过来在事件发生后反推在什么参数条件下，这个事件发生的概率最大（例如下雨（果）的时候有乌云（因/证据/观察的数据）的概率，即已经有了果，对证据发生的可能性描述）。数学语言可以将其写为：

$$\begin{gathered} P(x|\beta) = a \\ L(\beta|x) = a \\ P(x|\beta) = L(\beta|x) \end{gathered} \tag{2 - 5}$$

二者是相等的。

在二分类问题中可以概率写为：

$$\begin{gathered} P(y|x, \beta) = P(y = 1)^y P(Y = 0)^{1-y} \\ P(y|x, \beta) = \frac{1}{1 + e^{-z}}^y \left(1 - \frac{1}{1 + e^{-z}}\right)^{1-y} \end{gathered} \tag{2 - 6}$$

。

得到最大值：

$$L(\beta) = \Pi_{i=1}^{n} P(y_i|x_i, \beta) = \Pi_{i=1}^{n} \frac{1}{1 + e^{-w \cdot x_i}}^{y_i} \left(1 - \frac{1}{1 + e^{-w \cdot x_i}}\right)^{1-y_i} \tag{2 - 7}$$

对其取对数可得：

$$\log(L(\beta)) = \sum_{i=1}^{n}\left(y_i \cdot \log\left(\frac{1}{1+e^{-w\cdot x_i}}^{y_i}\right) + \log\left((1-y_i)\cdot\left(1-\frac{1}{1+e^{-w\cdot x_i}}\right)^{1-y_i}\right)\right) \tag{2 - 8}$$

2.2. 损失函数

损失函数是用于衡量预测值与实际值的偏离程度，即模型预测的错误程度。通过寻找误差最小时的参数来确定出最佳的模型。

使用平方损失：

$$Q = \sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{2 - 9}$$

待入到逻辑回归中损失函数就是：

$$Q = \sum_{i=1}^{n}\left(y_i - \frac{1}{1+e^{-w*x}}\right) \tag{2 - 10}$$

因为这个函数不死凸函数，所以使用对数损失函数。

$$Q = -\log(L(\beta)) \tag{2 - 11}$$

3. 多元线性回归

多元线性回归相比于线性回归，特征与输入多余线性回归。形式为：

$$\hat{y} = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + ... + b \tag{3 - 1}$$

可以将其写成矩阵的形式：$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} + (b) = w_1x_1 + w_2x_2 + w_3x_3 + b$ 相当于是一个点乘。

进而可以用矩阵$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$ 来表示数据集，一行表示一组数据，一列对应与某特征值。使用矩阵$W = \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix}$ 来表示特征, 偏置可以表示为：(b)。这样计算矩阵：

$$\hat{y} = X \times W^T + b = \begin{pmatrix} \begin{pmatrix} x_{11} & x_{12} & x_{13} \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} + b \\ \begin{pmatrix} x_{21} & x_{22} & x_{23} \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} + b \\ \begin{pmatrix} x_{31} & x_{32} & x_{33} \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} + b \end{pmatrix} \tag{3 - 2}$$

使用平方损失函数：

$$Q = \sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{3 - 3}$$

4. Softmax 回归

在前面的逻辑回归，它主要应用与二分类问题，对于多分类问题，有多个输入，多个特征，多个输出。

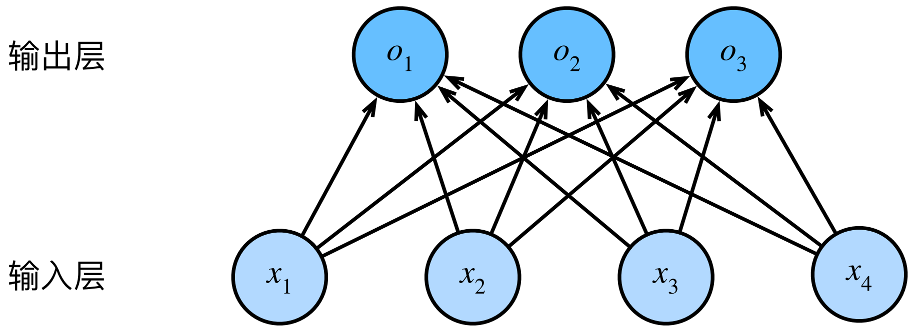

Figure 4 - 1: 网络图

可以表示为：

$$
\begin{aligned}
o_1 &= x_1 w_{11} + x_2 w_{12} + x_2 w_{13} + b_1 \\
o_2 &= x_1 w_{21} + x_2 w_{22} + x_2 w_{23} + b_2 \\
o_3 &= x_1 w_{31} + x_2 w_{32} + x_2 w_{33} + b_3
\end{aligned}
\tag{4 - 1}
$$

使用三个线性的表达式表示。

将其写成矩阵形式，可以表示为：

因为由上述的线性表达式得出的结果可能为正可能为负，三个输出之和也可能大于 1，这样都会违背概率的定理，所以使用 softmax 函数来规范得到的结果。