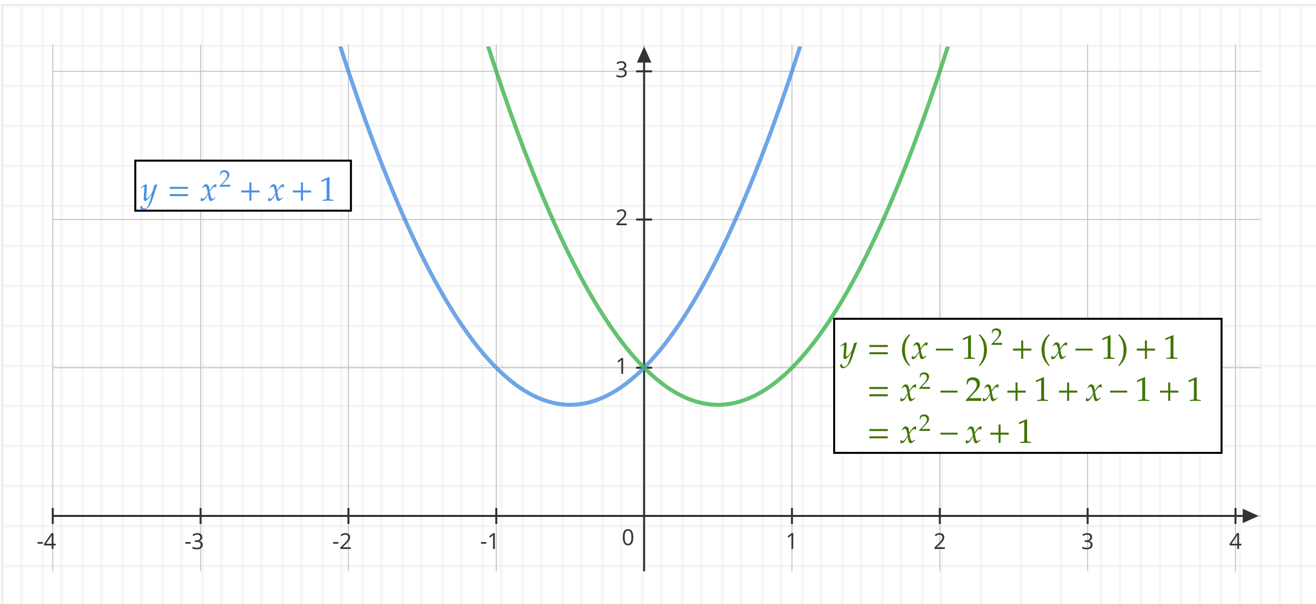
函数的变换

为什么线性函数，通过改变系数可以实现坐标的变换。首先线性函数都可以表示为多项式的形式：

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^x \tag{1}$$

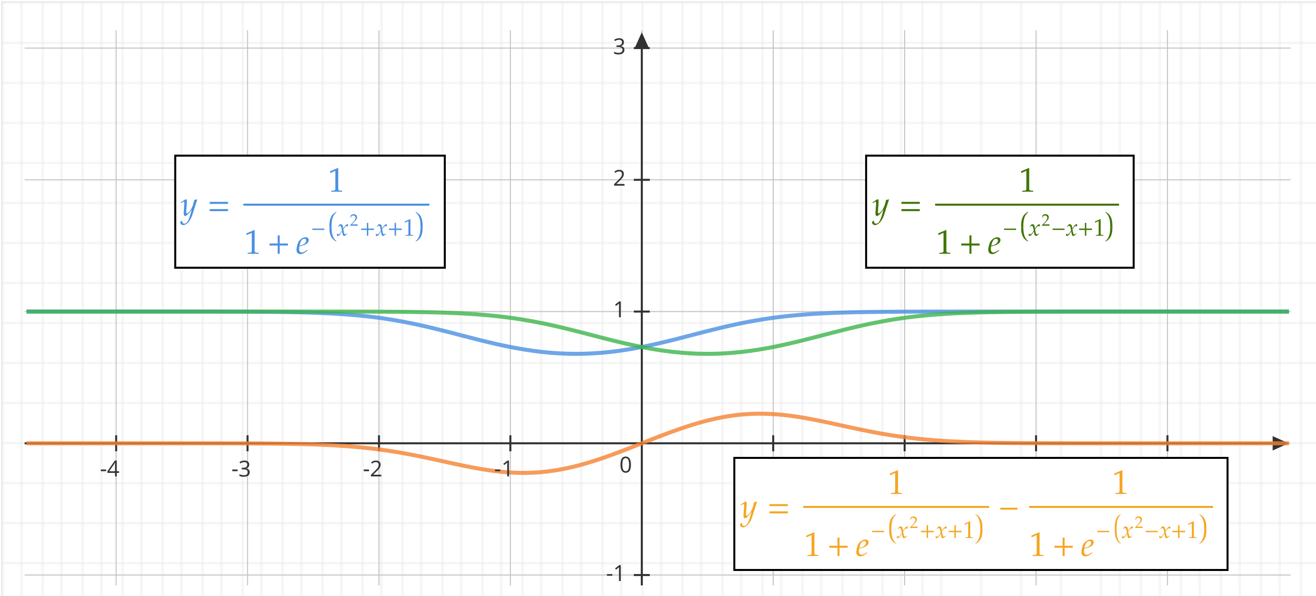
那么对多项式图像的坐标变换，到最后都可以化简到系数的变化上面。如果是n 元一次函数就更容易理解了。



那么对其复合函数来说呢？复合函数相当于用一个函数把原本是线性的输出硬给掰弯成其他形状，例如：

$$u = x^2 + x + 1 \tag{2}$$

$$y = \frac{1}{1 + e^{-u}} \tag{3}$$



我们同样可以看到，通过改变系数，同样也可以对复合函数进行坐标变换。甚至利用复合函数构造成更高阶的复合函数来拟合一条曲线。那么理论上来说，我们就可以通过足够多的复合函数来拟合任意形状的曲线，并且可以扩展到高维数据。有点像级数的意味了。

softmax

一般$softmax$ 函数后只有固定的误差函数，所以设计时可以作为一个单元考虑：$softmax-loss$。

$$z^{[L]}=a^{[L-1]}W^{[L]}+b^{[L]}$$

$$a^{[L]}=softmax\left(z^{[L]}\right)$$

$$softmax\left(z_j^{[L]}\right)=\frac{e^{z_j^{[L]}}}{\sum_{t=0}^{n}e^{z_t^{[L]}}}$$

$$y=[0,\quad 1,\quad 0,\quad 0\quad]^T$$

$$a^{[L]}=\hat{y}=[0.3,0.2,0.1,0.4]^T$$

$$\begin{aligned}Loss(\hat{y},y)&=-\sum_{i=1}^{n}y_i log\left(\widehat{y_i}\right)\\&=-log\widehat{y_2}\\&=-loga_2^{[L]}\end{aligned}$$

$$da_2^{[L]}=\frac{\partial Loss}{\partial a^{[L]}}=-y_2\frac{1}{a_2^{[L]}}$$

$$da^{[L]}=\left[0,-\frac{1}{a_2^{[L]}},0,0\right]$$

特殊构造的误差函数，
使得仅在概率最高的项时导数不为0

每一类的概率对z 的每一项都有偏导数
实际上是一个nxn 的雅可比矩阵

$$\frac{\partial a_j^{[L]}}{\partial z_i^{[L]}} = \frac{\partial softmax\left(z_j^{[L]}\right)}{\partial z_i^{[L]}} = \frac{e^{z_j^{[L]}}\sum_{t=0}^{n} e^{z_t^{[L]}} - \sum_{t=0}^{n} e^{z_t^{[L]}} e^{z_j^{[L]}}}{\left(\sum_{t=0}^{n} e^{z_t^{[L]}}\right)^2}$$

$$= \begin{cases} softmax\left(z_j^{[L]}\right) - softmax\left(z_j^{[L]}\right)^2 & i = j \\ -softmax\left(z_j^{[L]}\right) softmax\left(z_i^{[L]}\right) & i \neq j \end{cases}$$

$$= \begin{cases} a_j^{[L]}\left(1 - a_j^{[L]}\right) & i = j \\ -a_i^{[L]} a_j^{[L]} & i \neq j \end{cases}$$

$$dz_i^{[L]} = \begin{cases} -\frac{y_j}{a_j^{[L]}} a_j^{[L]}\left(1 - a_j^{[L]}\right) = a_j^{[L]} - 1 & i = j \\ \frac{y_j}{a_j^{[L]}} a_i^{[L]} a_j^{[L]} = a_i^{[L]} & i \neq j \end{cases}$$

$$= \left[a_1^{[L]}, a_2^{[L]}, \ldots, a_j^{[L]} - 1, \ldots, a_n^{[L]}\right] = a^{[L]} - y$$

但实际上我的问题还是没有解决，所有的偏导应该是个雅可比矩阵，如果其他项不为0 的话。应该怎么处理呢？求平均吗？