Aproximações de função sigmoide

Luís Otávio Amorim

6 de julho de 2023

1 Aproximação do professor Angelo

Esta aproximação separa a sigmoide em 4 regiões sendo duas lineares e duas não lineares, desta forma há 3 parâmetros que devem ser ajustados: $x_{\min}$, x_{med} e $x_{\max}$.

Assim os intervalos $[-\infty, x_{\min})$ e $[x_{\max}, \infty]$ são tratados como lineares e aproximados por duas retas distintas. Já os intervalos $[x_{\min}, x_{med})$ e $[x_{med}, x_{\max})$ são considerados não lineares, sendo aproximados assim por parábolas. Então a sigmoide é aproximada por uma função definida em partes como mostra a equação 1.

$$y = \begin{cases} c_0 + b_0 x, & x < x_{\min} \\ c_1 + x(b_1 + a_1 x), & x_{\min} \leq x < x_{med} \\ c_2 + x(b_2 + a_2 x), & x_{med} \leq x < x_{\max} \\ c_3 + b_3 x, & x_{\max} < x \end{cases} \tag{1}$$

Os parâmetros podem ser calculados como nas equações abaixo. Para calculá-los é importante considerar também os valores de $x_{\sup}$ e $x_{\inf}$ que vai depender da quantidade de bits na parte inteira do seu número em ponto fixo. Como estamos utilizando, no geral 3 bits de parte inteira, temos $-x_{\inf} = x_{\sup} = 8$

$$\begin{aligned} b_0 =& \frac{-y_{\min}}{x_{\inf} - x_{\min}} \\ c_0 =& -b_0 x_{\inf} \end{aligned} \tag{2}$$

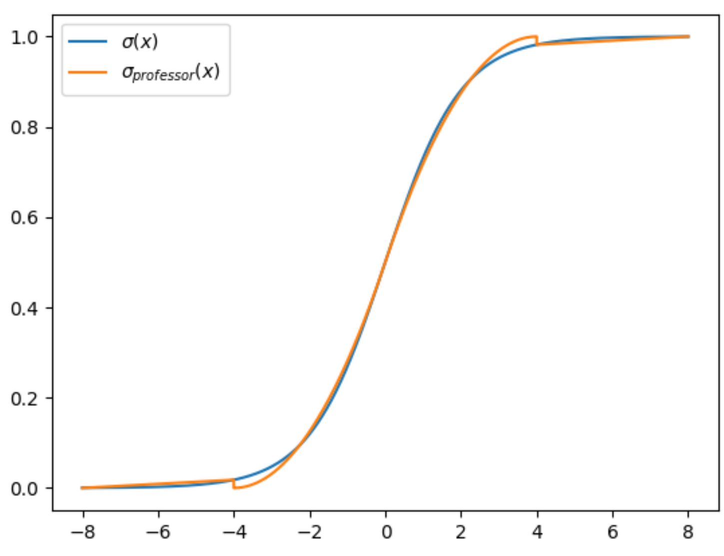
$$\begin{aligned} b_1 =& \frac{x_{\min}}{2x_{\min}x_{med} - x_{med}^2 - x_{\min}^2} \\ a_1 =& \frac{-b_1}{2x_{\min}} \\ c_1 =& \frac{b_1^2}{4a_1} \end{aligned} \tag{3}$$

$$
\begin{aligned}
b_2 =& \frac{-x_{\max}}{2x_{\max}x_{med} - x_{med}^2 - x_{\max}^2} \\
a_2 =& \frac{-b_2}{2x_{\max}} \\
c_2 =& 1 + \frac{b_2^2}{4a_2}
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
b_3 =& \frac{1 - y_{\max}}{x_{\sup} - x_{\max}} \\
c_3 =& 1 - b_3 x_{\sup}
\end{aligned} \tag{5}
$$

Realizei diversos testes porém, como no nosso caso a sigmoide é sempre simétrica, fixei x_{med} em 0. Assim os testes foram variando os parâmetros $x_{\max}$ e $x_{\min}$. Os melhores resultados que obtive foi quando $-x_{\min} = x_{\max} = 4$. Neste caso obtivemos um erro máximo de $2,16\%$ e um erro médio de $0,78\%$. A figura 1 mostra uma comparação entre o resultado obtido e uma sigmoide real.

Figura 1: Comparação do primeiro método com uma sigmoide real



2 Nova aproximação

A segunda aproximação que realizei foi uma que eu mesmo criei. A ideia aqui é utilizar as propriedades simétricas da sigmoide disponíveis na equação 6. Assim é necessário aproximar a sigmoide apenas para valores positivos, podendo extrapolar esta aproximação para valores negativos.

$$\sigma(x) = 1 - \sigma(-x) \tag{6}$$

Esta aproximação é feita dividindo o intervalo $[0, \infty]$ em três partes sendo duas lineares e uma não linear e, para isso é necessário ajustar dois parêmtros chamados aqui de fim_l e fim_n. Assim, a equação 7 define a função de aproximação.

$$y = \begin{cases} 1 + a_3x - b_3, & x_{\text{inf}} \leq x \leq -fim_n \\ 1 - c_2 + x(-a_2x + b_2), & -fim_n < x \leq -fim_l \\ 1 + a_1x - b1, & -fim_l < x \leq 0 \\ a_1x + b_1, & 0 < valor \leq fim_l \\ c_2 + x(a_2x + b_2), & fim_l < x \leq fim_n \\ a_3x + b_3, & fim_n < x \leq x_{\text{sup}} \end{cases} \tag{7}$$

Desta forma, por mais que o valor de y é dividido em 6 intervalos distintos, utilizamos a simetria para ser necessário parametrizar apenas 3 desses 6 intervalos. Os valalores x_{inf} e x_{sup}, assim como na outra aproximação dependem das caracteristicas de ponto fixo utilizada, então não são parâmetros modificaveis, além disso nas definições das constantes utilizamos também o parâmetro x_{med} que é a média entre fim_l e fim_n. Os outros parâmetros podem ser calculados seguindo as fórmulas abaixo.

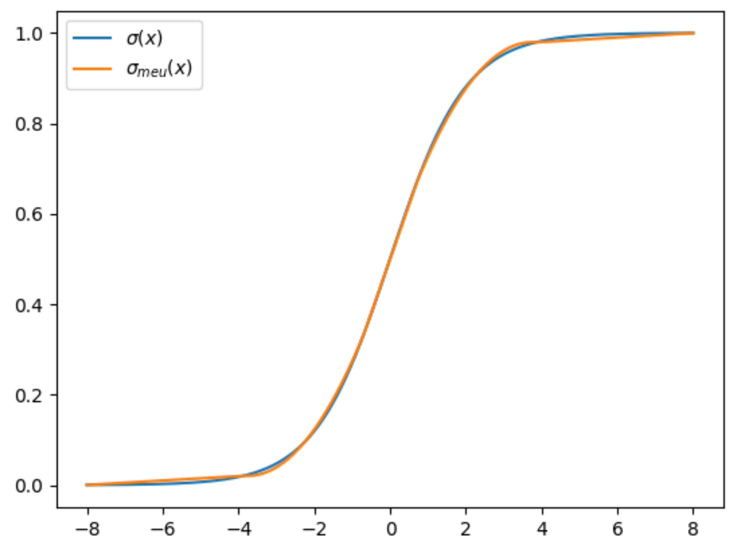
$$\begin{aligned} a_1 =& \frac{\sigma(fim_l) - 0,5}{fim_l} \\ b_1 =& \sigma(fim_l) \end{aligned} \tag{8}$$

$$\begin{aligned} a_2 &= \frac{2\sigma(fim_n) - 4\sigma(x_{med}) + 2\sigma(fim_n)}{(fim_n - fim_l)^2} \\ b_2 &= \frac{2(\sigma(x_{med}) - \sigma(fim_l))}{fim_n - fim_l} - \frac{a_2}{2}(3fim_l + fim_n) \\ c_2 &= \sigma(fim_n) - a_2 fim_l^2 - b_2 fim_l \end{aligned} \tag{9}$$

$$\begin{aligned} a_3 =& \frac{\sigma(fim_n) - x_{\text{sup}}}{fim_n - x_{\text{sup}}} \\ b_3 =& \sigma(fim_n) - fim_n a_3 \end{aligned} \tag{10}$$

Dentro dos testes que realizei, os melhores resultados foram quando $fim_l = 0,7$ e $fim_n = 3,85$ obtendo assim um erro máximo de $0,87\%$ e um erro médio de $0,51\%$. A figura 2 mostra a comparação de uma sigmoide real e desta aproximação.

Figura 2: Comparação do segundo método com uma sigmoide real



2.1 Deduções

Abordarei aqui como as deduções das constantes foram feitas para esta aproximação. Como pode ser visto na equação 7 são criados 6 intervalos que, devido à simetria se tornam apenas 3. O primeiro intervalo linear ocorre com $x \in [0, fim_l]$, o segundo intervalo, esse não linear, ocorre com $x \in (fim_l, fim_n]$ e o último intervalo, linear, ocorre com $x \in (fim_n, x_{\text{sup}}]$.

2.1.1 Intervalos lineares

Os intervalos lineares possuem uma dedução extremamente simples, apenas traçamos uma reta que passa pelos dois pontos que limitam o intervalo segundo a equação 11.

$$y - y_1 = \frac{y_1 - y_2}{x_1 - x_2}(x - x_1) \tag{11}$$

Realizaremos a dedução para a segunda parte linear, já que a da primeira é similar e inclusive mais simples. Substituindo na equação $y_1 = \sigma(fim_n)$, $y_2 = \sigma(x_{\text{sup}})$, $x_1 = fim_n$ e $x_2 = x_{\text{sup}}$ temos:

$$\begin{aligned} y - \sigma(fim_n) =& \frac{\sigma(fim_n) - \sigma(x_{\text{sup}})}{fim_n - x_{\text{sup}}}(x - fim_n) \\ y =& \frac{\sigma(fim_n) - \sigma(x_{\text{sup}})}{fim_n - x_{\text{sup}}}(x - fim_n) + \sigma(fim_n) \\ y =& \frac{\sigma(fim_n) - \sigma(x_{\text{sup}})}{fim_n - x_{\text{sup}}}x - fim_n\frac{\sigma(fim_n) - \sigma(x_{\text{sup}})}{fim_n - x_{\text{sup}}} + \sigma(fim_n) \end{aligned} \tag{12}$$

Considerando também a equação reduzida da reta: $y = ax+b$ podemos fazer as correspondências e obter que:

$$\begin{aligned} a =& \frac{\sigma(fim_n) - \sigma(x_{\text{sup}})}{fim_n - x_{\text{sup}}} \\ b = -fim_n\frac{\sigma(fim_n) - \sigma(x_{\text{sup}})}{fim_n - x_{\text{sup}}} +& \sigma(fim_n) = \sigma(fim_n) - fim_n a \end{aligned} \tag{13}$$

2.1.2 Intervalo não linear

Neste intervalo tentamos definir a sigmoide como uma parábola e para tanto seria necessário três pontos, para isso escolhemos os limites do intervalo: fim_l, fim_n e a média entre eles, assim definimos $x_1 = fim_l$, $x_2 = \frac{x_1 + x_2}{2}$ e $x_3 = fim_n$, além de que $y_1 = \sigma(x_1)$, $y_2 = \sigma(x_2)$ e $y_3 = \sigma(x_3)$. Desta forma foi possível escrever o sistema de equações abaixo:

$$\begin{aligned} y_1 =& ax_1^2 + bx_1 + c \\ y_2 =& ax_2^2 + bx_2 + c \\ y_3 =& ax_3^2 + bx_3 + c \end{aligned} \tag{14}$$

Esse sistema foi resolvido utilizando o método do escalonamento segundo as equações abaixo.

$$\begin{bmatrix} 1 & x_1 & x_1^2 & y_1 \\ 1 & x_2 & x_2^2 & y_2 \\ 1 & x_3 & x_3^2 & y_3 \end{bmatrix} \begin{matrix} \\ L_2 = L_2 - L_1 \\ L_3 = L_3 - L_1 \end{matrix}$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & y_1 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & y_2 - y_1 \\ 0 & x_3 - x_1 & x_3^2 - x_1^2 & y_3 - y_1 \end{bmatrix} L_3 = L_3 - \frac{x_3 - x_1}{x_2 - x_1} L_2$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & y_1 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & y_2 - y_1 \\ 0 & 0 & x_3^2 - x_1^2 - \frac{(x_3 - x_1)}{(x_2 - x_1)}(x_2^2 - x_1^2) & y_3 - y_1 - \frac{(x_3 - x_1)}{(x_2 - x_1)}(y_2 - y_1) \end{bmatrix}$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & y_1 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & y_2 - y_1 \\ 0 & 0 & (x_3 - x_1)(x_3 + x_1) - \frac{(x_3 - x_1)\cancel{(x_2 - x_1)}(x_2 + x_1)}{\cancel{(x_2 - x_1)}} & y_3 - y_1 - \frac{(x_3 - x_1)}{(x_2 - x_1)}(y_2 - y_1) \end{bmatrix}$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & y_1 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & y_2 - y_1 \\ 0 & 0 & (x_3 - x_1)\left[(x_3 + x_1) - (x_2 + x_1)\right] & y_3 - y_1 - \frac{(x_3 - x_1)}{(x_2 - x_1)}(y_2 - y_1) \end{bmatrix}$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & y_1 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & y_2 - y_1 \\ 0 & 0 & (x_3 - x_1)(x_3 - x_2) & y_3 - y_1 - \frac{(x_3 - x_1)}{(x_2 - x_1)}(y_2 - y_1) \end{bmatrix}$$

A partir daí foi possível encontrar os coeficientes a, b e c cujas expressões estão abaixo.

$$\begin{aligned} a =& \frac{y_3 - y_1}{(x_3 - x_1)(x_3 - x_2)} - \frac{y_2 - y_1}{(x_3 - x_2)(x_2 - x_1)} \\ b =& \frac{y_2 - y_1}{x_2 - x_1} - a(x_2 + x_1) \\ c =& y_1 - a x_1^2 - b x_1 \end{aligned} \tag{15}$$

Por fim, novamente utilizando o fato de que $x_2 = \dfrac{x_1 + x_3}{2}$ simplificamos estas expressões para obter as expressões em 9.