

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

RNP Redes Neurais Profundas - Deep Learning

Aluno: Fabiano Rodrigues Barbosa

LISTA DE EXERCÍCIOS 2

1. Sabe-se que cada neurônio, de forma individual, estabelecerá um hiperplano de separação em um dado problema. Se uma quantidade relativamente grande de neurônios forem colocados em paralelo em uma única camada, tal rede seria capaz de separar problemas não-lineares? Justifique sua resposta.

Se estes neurônios do tipo perceptron em paralelo forem colocados em paralelo e adicionar uma camada de saída com função de ativação bipolar, sim a rede seria capaz de separar problemas não-lineares. A proposta de adicionar mais neurônios na camada é uma forma de se resolver problemas mais complexos que não são simplesmente separados linearmente. Para redes do tipo perceptron, existe uma relação da quantidade de linhas de separação geradas com a quantidade de neurônios da rede. Cada neurônio é responsável por um segmento de separação do problema. Os resultados destes neurônios são combinados para se obter uma saída de separação:

1 neurônio → 1 linha de classificação, 2 neurônios → 2 linhas de classificação, 3 neurônios → 3 linhas de classificação, e assim por diante.

2. Qual a importância da escolha da função de ativação em redes neurais artificiais?

A função de ativação nas redes neurais artificiais tem a função de definir a saída de cada neurônio. Esta saída pode levar a uma ativação ou não em sua saída. Cada função de ativação tem um conjunto de saídas diferentes, por exemplo: a sigmoid retorna apenas valores positivos entre 0 e 1, a tangente hiperbólica retorna valores entre -1 e 1, já a função softmax retorna a probabilidade dos dados serem de uma das classes definidas. Em uma situação que a rede precisa retornar várias classes, um função softmax é mais recomendada em relação a uma sigmoid que retorna apenas duas classes.

O tipo de função de ativação monótona ou não também define a saída da ativação. Uma função de ativação monótona como a binária é capaz de mapear uma saída diferente de uma não monótona com uma função senoidal. Um exemplo neste caso seria a classificação abaixo, na qual a figura a esquerda é o resultado de uma rede perceptron de 3 camadas com 2 neurônios na camada de entrada, 2 neurônios na camada oculta e 1 neurônio de saída. Já na figura a direita é o resultado de uma rede de base radial com somente 1 neurônio com função de ativação Gaussiana (não monótona).

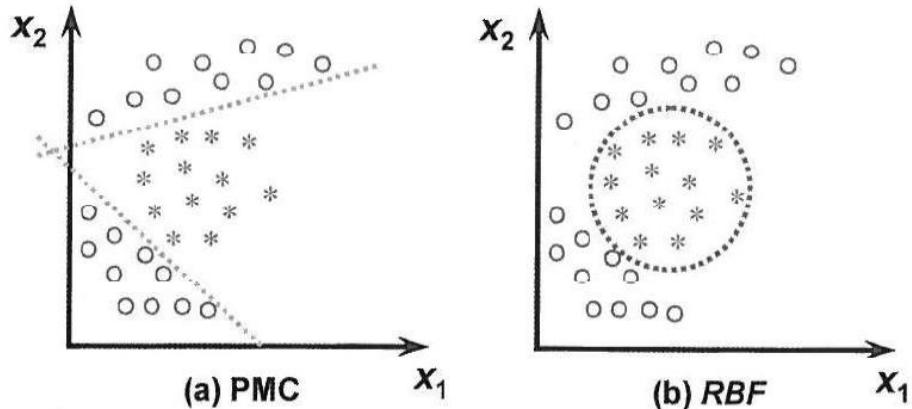


Figura 6.5 – Fronteira de separabilidade do PMC e RBF

Outro ponto importante a se considerar são as propriedades das funções de ativação. Funções do tipo não-lineares não conseguem resolver problemas mesmo em redes com vários neurônios distribuídos em várias camadas. Quando se utiliza apenas funções não-lineares, obtém apenas o mesmo resultado de um único neurônio.

A combinação de função de ativação com o método de otimização também deve ser levado em conta. Se utilizar função de ativação binária com método de otimização gradiente, não será possível resolver o problema. A derivada de uma função binária no ponto zero é indefinida e fora do ponto zero é uma constante, logo o método gradiente não consegue resolver a questão. Esta mesma situação aplica-se a função de ativação identidade. Para aplicar o gradiente é necessário funções contínuas que sejam possíveis aplicar derivada, tais como sigmoid, tangente hiperbólica ou arco tangente.

Mais um ponto importante são os possíveis valores de retorno das funções de ativação. Quando estes valores são finitos (por exemplo: sigmoid retorna valores entre 0 e 1) os treinamentos baseados em gradiente tentam a ser mais estáveis, pois afetam de forma limitada os pesos dos neurônios. Quando são infinitos, as atualizações de pesos são mais intensas levando a uma maior eficiência do treinamento. Porém essa eficiência tem o custo de serem mais instáveis necessitando ajustar melhor a taxa de aprendizagem do treinamento.

3. Qual o possível ganho, em comparação a um único neurônio, que se pode obter em uma rede neural artificial com duas camadas de neurônios?

A quantidade de neurônios e camadas utilizadas para resolução de problemas utilizando redes neurais é diretamente relacionada com a complexidade do problema a ser resolvido. Um único neurônio perceptron é capaz de separar apenas linearmente o problema. Porém quando combinado com outros neurônios em rede cada um poderá gerar uma hiperplano de separação. Assim podemos utilizar um conjunto de neurônios devidamente configurados para resolver problemas mais complexos através de combinação de vários hiperplanos de separação.

4. Quais são as dificuldades em usar uma rede neural de múltiplas camadas?

As redes neurais de múltiplas camadas tem o potencial de resolver problemas mais complexos quando comparadas com redes mais simples. Porém as redes de múltiplas camadas tendem a ter um ajuste de pesos durante a aprendizagem muito baixa para as primeiras camadas mais próximas à entrada. Os pesos das últimas camadas, mais próximas à saída da rede, tentam a ser ajustados com maior eficiência em relação às primeiras camadas. Isso é devido ao ajuste de

vários níveis da regra da cadeia. Dentre as técnicas possíveis, a técnica de inicialização de pesos chama Xavier pode ser utilizada para obter melhores resultados.

Além disso, as redes de múltiplas camadas devem ser treinadas com maior cuidado por existir vários parâmetros a serem devidamente configurados para se obter um resultado aceitável, tais como: número de camadas, número de neurônios em cada camada, taxa de aprendizagem, inicialização dos pesos, definição de funções de ativação em cada camada, número de épocas, etc.

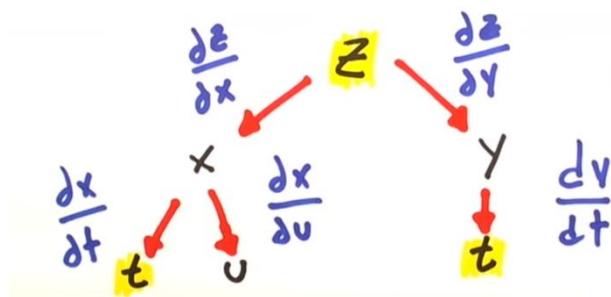
Adicionalmente redes neurais de múltiplas camadas tem maior facilidade de ocorrer overfitting, que pode ser tratado utilizando de técnicas que adicionam ainda mais parâmetros como: regularização (L1 weight decay e L2 sparsity), dropout.

Para finalizar, além da grande quantidade de parâmetros a serem devidamente configurados, há várias correlações entre estes diversos parâmetros. Por exemplo, para se obter melhor aprendizagem os pesos iniciais devem ser de tal forma que estejam condizentes com a faixa de ativação das funções de ativação escolhidas, permitindo a saída ser sensibilizada sem saturar a função de ativação. Assim deve-se pensar na rede como um todo para avaliar a melhor combinação de parâmetros para obter os melhores resultados.

5. Explique intuitivamente o que é a regra da cadeia.

A regra da cadeia é um método para se chegar ao ajuste de pesos necessário nas diversas camadas das redes neurais multicamadas. O primeiro passo consiste em calcular o erro utilizando de pesos iniciais. Ao final do processo de propagação, é calculado o erro que deverá ser utilizado para ajustar os pesos para a próxima época. Durante a retropropagação os pesos da última camada vão ser atualizados de acordo com o erro final. Já da penúltima camada serão ajustados de acordo com o erro da última camada e assim até chegar a primeira camada. É semelhante a engrenagens conectadas entre si, na qual o ajuste do peso da última camada, ou melhor, o giro da engrenagem da última camada faz girar as demais engrenagens conectadas entre si até chegar a engrenagem da primeira camada. Como se tratam de engrenagens de tamanhos diferentes, a razão de giro delas é diferente em cada uma.

Uma outra forma de entender a regra da cadeia é visualizar na forma de uma árvore. No exemplo abaixo Z é uma função que depende de X e Y. Já X depende de t e u, e Y de t. Se quisermos obter a derivada de Z em relação a t devemos aplicar a derivada de cada parcela do nó. Assim a derivada em cada nó é a soma das derivadas de seus filhos.



Desta forma a derivada de Z em relação a t é igual a:

$$\boxed{\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t}}$$

Já ser a derivada que quisermos obter fosse Z em relação a u, teríamos a equação abaixo. Neste caso não temos a segunda parcela da adição pelo fato da derivada dela ser uma constante zero.

$$\boxed{\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial u}}$$

6. Explique matematicamente o que é a regra da cadeia.

Em cálculo, a regra da cadeia é uma fórmula para a derivada da função composta de duas funções.

Por exemplo, sejam $y = f(x)$ e $u = g(x)$ duas funções tais que suas derivadas existem e

exista a derivada da função $y = f(g(x))$ que indicaremos por $\frac{dy}{dx}$, então

$$y' = \frac{dy}{dx} = f'(g(x)) \cdot g'(x)$$

$$\text{ou ainda, } y' = \frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

Logo,

$$y = f(g(x)) \Rightarrow y' = f'(g(x)) \cdot g'(x)$$

A derivada obtida acima da função composta também é conhecida como regra da cadeia.

7. Qual a importância das derivadas parciais para o processo de treinamento de uma rede neural artificial multicamadas.

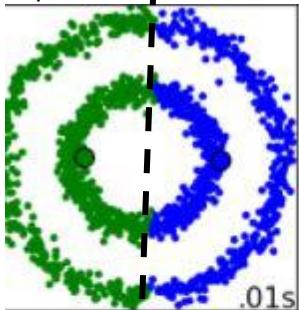
As derivadas parciais são utilizadas para se calcular o valor de ajuste dos pesos de cada neurônio na rede. Assim o erro calculado ao final do processo de propagação é refletido das últimas camadas próximas à saída da rede para as primeiras até a camada de entrada da rede, camada a camada. Na camada de saída (última) se utiliza da derivada simples, já na penúltima é a aplicação dupla da derivada parcial em relação ao erro. Desta forma os pesos são ajustados de acordo com as várias derivadas parciais em cadeia e o ajuste de cada camada depende do resultado da derivada parcial da anterior. Este processo é feito em cascata da última camada até a primeira.

8. Quanto mais camadas uma rede neural artificial possuir, melhor o seu desempenho?

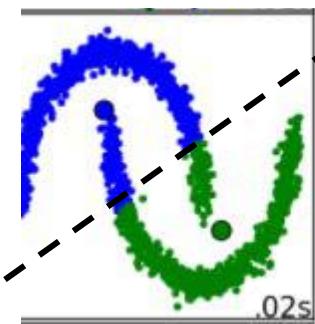
Não necessariamente, pois quanto mais camadas se adicionam à rede, problemas mais complexos podem ser resolvidos. Por outro lado, utilizar mais camadas que o necessário para resolver o problema pode levar a efeitos colaterais como overfitting e maior tempo

computacional, que poderão impactar em resultados até piores daquelas redes com menor número de camadas.

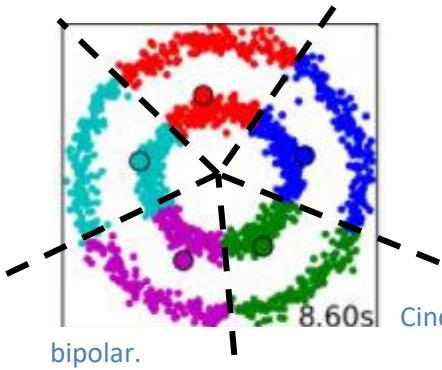
9. Dados quatro problemas diferentes, ilustrados na figura abaixo. Como você projetaria a arquitetura de uma rede neural artificial de modo a economizar neurônios utilizados na rede?



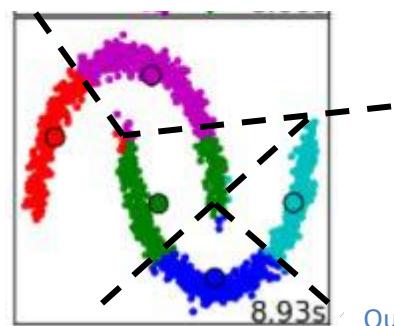
Apenas um neurônio perceptron com função de ativação bipolar.



Apenas um neurônio perceptron com função de ativação bipolar.



Cinco neurônios perceptron em paralelo com função de ativação bipolar.



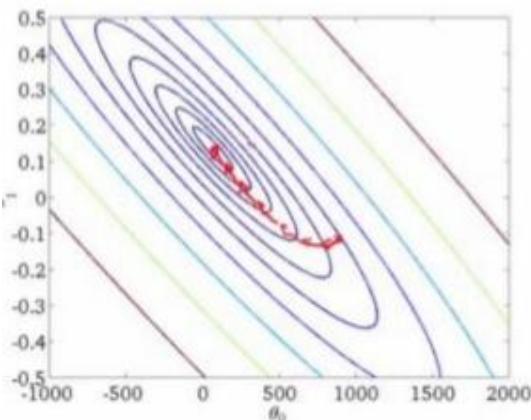
Quatro neurônios perceptron em paralelo com função de ativação bipolar.

10. Qual a relação entre gradiente descendente, gradiente descendente estocástico e mínimos locais e globais?

O gradiente descendente (batch gradient) é o método de minimizar o erro de uma função através de atualização dos pesos utilizando toda a amostra a ser treinada. Ou seja, o erro é calculado para o todo o conjunto de dados a ser treinado, e este erro de toda amostra é utilizada para ajustar os pesos. Já no gradiente descendente estocástico o erro e o ajuste de peso são calculados utilizando apenas uma amostra por vez.

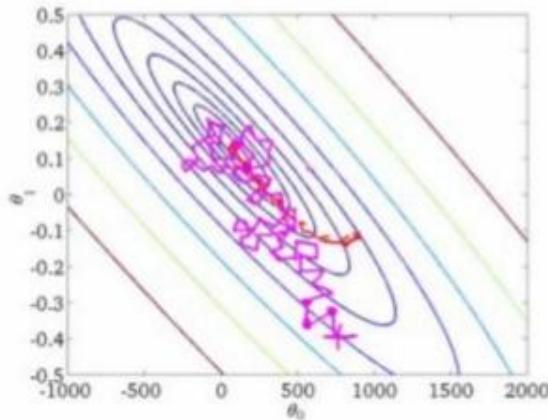
Conforme mostrado no gráfico abaixo, o gradiente descendente tende a convergir ao um mínimo de forma mais rápida, enquanto o estocástico realiza vários ajustes: um passo de ajuste a cada amostra o que leva a demora para encontrar um resultado mínimo satisfatório. Assim o método estocástico tente de passar por pontos de mínimo locais com maior facilidade quando comparado com o batch. Este fato pode levar a encontrar um mínimo local e se satisfazer com o resultado já que o erro foi minimizado.

Cada um destes dois métodos de gradiente tem suas vantagens e desvantagens. O gradiente descendente, por exemplo, requer mais computação por cada passo de atualização dos pesos. Por outro lado possui várias técnicas de aceleração que podem ser aplicadas no batch gradient que não podem ser aplicadas no estocástico.



Batch: gradient

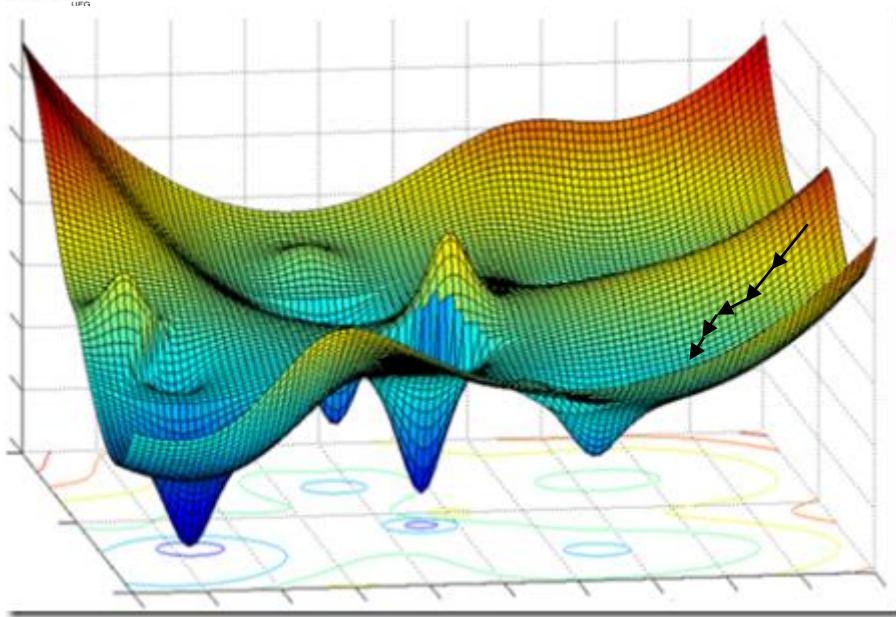
$$x \leftarrow x - \eta \nabla F(x)$$



Stochastic: single-example gradient

$$x \leftarrow x - \eta \nabla F_i(x)$$

O gráfico abaixo apresenta alguns pontos de mínimo local. Diferente do gráfico acima, neste é mais fácil visualizar que dependendo do ponto de partida para o cálculo do gradiente, pode-se chegar a um mínimo local a invés do mínimo global.



O estocástico é mais recomendável para um grande número de amostras a serem treinadas, pois geralmente a capacidade funcional pode ser um fator limitante para treinar toda a amostra. Com o método estocástico também pode ser utilizado para rastrear as mudanças de gradiente.

Ambos os métodos possuem suas vantagens e desvantagens, talvez valha aplicar um meio termo. Ou seja, nem treinar a rede com apenas uma amostra por vez nem treinar com toda a amostra ao mesmo tempo. Poderia se utilizar o conceito de mini-batches para treinar a rede, tentando obter assim o melhor de cada método.

11. Dado as entradas abaixo, mostre passo a passo com letra de próprio punho os cálculos da iteração da segunda época de treinamento para um peso de cada camada (a primeira foi mostrada em sala), ou seja, considere os pesos no instante ($t+1$). Pode-se suprimir os mesmos passos no exemplo mostrado em sala.

Entrada:

$I_1 = 0.05$

$I_2 = 0.1$

Saída:

$O_1: 0.01$

$O_2: 0.99$

[Veja resposta após do exercício 12.](#)

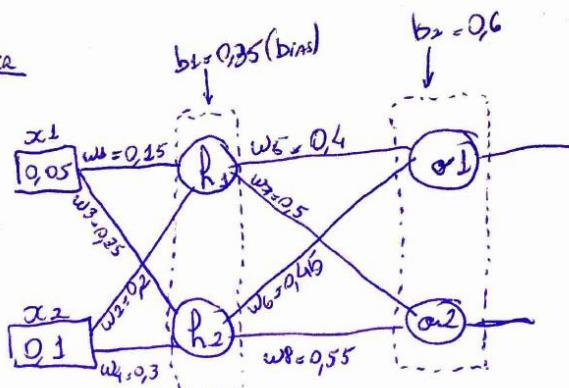
12. Implemente o algoritmo backpropagation em linguagem python e apresente um notebook correspondente com os comentários. Por fim, mostre o gráfico de erro em relação às épocas para o conjunto de dados do exercício anterior.

[Ver código anexado ou:](#)

https://github.com/fabianorb/mestrado_deep_learning/blob/master/lista02/exercicio12.ipynb

Resposta da Questão 11:
Como treinar uma rede multi-layer

ENTRADAS		SAÍDA DESEJADA	
x_1	x_2	y_1	y_2
0,05	0,1	0,01	0,99


FASE 1 - PROPAGAÇÃO

$$u_{h1} = x_1 \cdot w_1 + x_2 \cdot w_2 + b_1 \cdot 1$$

$$u_{h1} = 0,05 \cdot 0,25 + 0,1 \cdot 0,3 + 0,35 \cdot 1$$

$$u_{h1} = 0,3775$$

$$u_{h2} = x_1 \cdot w_3 + x_2 \cdot w_4 + b_2 \cdot 1$$

$$u_{h2} = 0,05 \cdot 0,35 + 0,1 \cdot 0,3 + 0,35 \cdot 1$$

$$u_{h2} = 0,3925$$

$$g(h_1) = g(u_{h1}) = \frac{1}{1 + e^{-u_{h1}}} = 0,593269992$$

$$g(h_2) = g(u_{h2}) = \frac{1}{1 + e^{-u_{h2}}} = 0,596884378$$

$$u_{o1} = g(h_1) \cdot w_5 + g(h_2) \cdot w_6 + b_2 \cdot 1$$

$$u_{o1} = 0,593269992 \cdot 0,4 + 0,596884378 \cdot 0,55 + 0,6$$

$$u_{o1} = 1,105905967$$

$$u_{o2} = g(h_1) \cdot w_7 + g(h_2) \cdot w_8 + b_2 \cdot 1$$

$$u_{o2} = 0,593269992 \cdot 0,5 + 0,596884378 \cdot 0,55 + 0,6$$

$$u_{o2} = 1,224928404$$

$$g(o_1) = g(u_{o1}) = \frac{1}{1 + e^{-u_{o1}}} = 0,75136507$$

$$g(o_2) = g(u_{o2}) = \frac{1}{1 + e^{-u_{o2}}} = 0,772928465$$

ERRO TOTAL

$$E_{\text{TOTAL}} = \frac{1}{N} \sum_{k=1}^N (y_k - g(o_k))^2 = E_{o1} + E_{o2}$$

$$E_{o1} = \frac{1}{2} (y_1 - g(o_1))^2 = \frac{1}{2} \cdot (0,01 - 0,75136507)^2 = 0,274811083$$

$$E_{o2} = \frac{1}{2} (y_2 - g(o_2))^2 = \frac{1}{2} \cdot (0,99 - 0,772928465)^2 = 0,023560026$$

$$E_{\text{TOTAL}} = E_{o1} + E_{o2} = 0,298371109$$

FASE 2 - RETRORROLAGEM

TAXA DE Aprendizado $n = 0,5$

AJUSTAR w_5 :

$$\frac{\partial E_{\text{Total}}}{\partial w_5} = \frac{\partial E_{\text{Total}}}{\partial g_{01}} \cdot \frac{\partial g_{01}}{\partial u_{01}} \cdot \frac{\partial u_{01}}{\partial w_5}$$

$$E_{\text{Total}} = \frac{1}{2} \cdot (y_1 - g_{01})^2 + \frac{1}{2} \cdot (y_2 - g_{02})^2$$

$$\frac{\partial E_{\text{Total}}}{\partial g_{01}} = 2 \cdot \frac{1}{2} (y_1 - g_{01})^{2-1} \cdot (-1) + 0 = -(y_1 - g_{01}) = -0,01 + 0,75136507 = 0,74136507$$

$$g_{01} = \frac{1}{1 + e^{-u_{01}}}$$

$$\frac{\partial g_{01}}{\partial u_{01}} = g_{01} (1 - g_{01}) = 0,75136504 \cdot (1 - 0,75136504) = 0,186815602$$

$$u_{01} = w_5 \cdot g_{h1} + w_6 \cdot g_{h2} + b_2 \cdot 1$$

$$\frac{\partial u_{01}}{\partial w_5} = 1 \cdot g_{h1} \cdot w_5^{1-1} + 0 + 0 = g_{h1} = 0,59326992$$

$$\frac{\partial E_{\text{Total}}}{\partial w_5} = 0,74136507 \cdot 0,18681562 \cdot 0,59326992 = 0,0822167041$$

$$w_5(t+1) = w_5(t) - n \cdot \frac{\partial E_{\text{Total}}}{\partial w_5} = 0,4 - 0,5 \cdot 0,0822167041 = 0,35791648$$

AJUSTAR w_6 :

$$\frac{\partial E_{\text{Total}}}{\partial w_6} = \frac{\partial E_{\text{Total}}}{\partial g_{02}} \cdot \frac{\partial g_{02}}{\partial u_{02}} \cdot \frac{\partial u_{02}}{\partial w_6}$$

$$u_{02} = w_5 \cdot g_{h1} + w_6 \cdot g_{h2} + b_2 \cdot 1$$

$$\frac{\partial u_{02}}{\partial w_6} = 0 + 1 \cdot g_{h2} \cdot w_6^{1-1} + 0 = g_{h2} = 0,596884378$$

$$\frac{\partial E_{\text{Total}}}{\partial w_6} = 0,74136507 \cdot 0,18681562 \cdot 0,596884378 = 0,082667628$$

$$w_6(t+1) = w_6(t) - n \cdot \frac{\partial E_{\text{Total}}}{\partial w_6} = 0,45 - 0,5 \cdot 0,082667628 = 0,408666386$$

Ajustar w_7 :

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_7} = \frac{\partial E_{\text{TOTAL}}}{\partial g_{02}} \cdot \frac{\partial g_{02}}{\partial u_{02}} \cdot \frac{\partial u_{02}}{\partial w_7}$$

$$E_{\text{TOTAL}} = \frac{1}{2} \cdot (y_t - g_{02})^2 + \frac{1}{2} (y_p - g_{02})^2$$

$$\frac{\partial E_{\text{TOTAL}}}{\partial g_{02}} = 2 \cdot \frac{1}{2} (y_t - g_{02})^{2-1} \cdot (-1) + 0 = -(y_t - g_{02}) = -(0,99 - 0,772928465) = \boxed{-0,217071535}$$

$$g_{02} = \frac{1}{1 + e^{-u_{02}}}$$

$$\frac{\partial g_{02}}{\partial u_{02}} = g_{02}(1 - g_{02}) = 0,772928465(1 - 0,772928465) = \boxed{0,175510053}$$

$$u_{02} = w_7 \cdot g_{h1} + w_8 \cdot g_{h2} + b_2$$

$$\frac{\partial u_{02}}{\partial w_7} = 1 \cdot w_7^{1-1} \cdot g_{h1} + 0 + 0 = g_{h1} = \boxed{0,593269992}$$

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_7} = -0,217071535 \cdot 0,175510053 \cdot 0,593269992 = \boxed{-0,022602540}$$

$$w_7(t+) = w_7(t) - n \cdot \frac{\partial E_{\text{TOTAL}}}{\partial w_7} = 0,5 - 0,5 \cdot (-0,022602540) = \boxed{0,5113052}$$

Ajustar w_8 :

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_8} = \frac{\partial E_{\text{TOTAL}}}{\partial g_{02}} \cdot \frac{\partial g_{02}}{\partial u_{02}} \cdot \frac{\partial u_{02}}{\partial w_8}$$

$$u_{02} = w_7 \cdot g_{h1} + w_8 \cdot g_{h2} + b_2$$

$$\frac{\partial u_{02}}{\partial w_8} = 0 + 1 \cdot w_8^{1-1} \cdot g_{h2} + 0 = g_{h2} = \boxed{0,596884378}$$

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_8} = -0,217071535 \cdot 0,175510053 \cdot 0,596884378 = \boxed{-0,022740242}$$

$$w_8(t+) = w_8(t) - n \cdot \frac{\partial E_{\text{TOTAL}}}{\partial w_8} = 0,55 - 0,5 \cdot (-0,022740242) = \boxed{0,561370221}$$

Ajustar b_{201} :

$$\frac{\delta E_{\text{TOTAL}}}{\delta b_{201}} = \frac{\delta E_{\text{TOTAL}}}{\delta g_{01}} \cdot \frac{\delta g_{01}}{\delta u_{01}} \cdot \frac{\delta u_{01}}{\delta b_{201}}$$

$$u_{01} = g_{h1} \cdot w_5 + g_{h2} \cdot w_6 + b_{201}$$

$$\frac{\delta u_{01}}{\delta b_{201}} = 0 + 0 + 1 \cdot b_{201}^{1,1} = 1$$

$$\frac{\delta E_{\text{TOTAL}}}{\delta b_{201}} = 0,74136507 \cdot 0,18681562 \cdot 1 = \boxed{0,138498575}$$

$$b_{201}(t+1) = b_{201}(t) - \eta \frac{\delta E_{\text{TOTAL}}}{\delta b_{201}}$$

$$b_{201}(t+1) = 0,6 - 0,5 \cdot 0,138498575 = \boxed{0,530750319}$$

Ajustar b_{202} :

$$\frac{\delta E_{\text{TOTAL}}}{\delta b_{202}} = \frac{\delta E_{\text{TOTAL}}}{\delta g_{02}} \cdot \frac{\delta g_{02}}{\delta u_{02}} \cdot \frac{\delta u_{02}}{\delta b_{202}}$$

$$\approx -0,217071535 \cdot 0,375510053 \cdot 1 = -0,038098237$$

$$b_{202}(t+1) = 0,6 - 0,5 \cdot (-0,038098237) = \boxed{0,619049118}$$

Ajustar w_2 :

$$\frac{\partial E_{TOTAL}}{\partial w_2} = \underbrace{\frac{\partial E_{TOTAL}}{\partial g_{h1}}}_{\textcircled{a}} \cdot \underbrace{\frac{\partial g_{h1}}{\partial u_{02}}}_{\textcircled{b}} \cdot \underbrace{\frac{\partial u_{02}}{\partial w_2}}_{\textcircled{c}}$$

(a) $E_{TOTAL} = E_{01} + E_{02} = \frac{1}{2} \cdot (y_1 - g_{01})^2 + \frac{1}{2} \cdot (y_2 - g_{02})^2$

$$\frac{\partial E_{TOTAL}}{\partial g_{h1}} = \underbrace{\frac{\partial E_{01}}{\partial g_{h1}}}_{\textcircled{a1}} + \underbrace{\frac{\partial E_{02}}{\partial g_{h1}}}_{\textcircled{a2}}$$

(a1) $\frac{\partial E_{01}}{\partial g_{h1}} = \frac{\partial E_{01}}{\partial u_{01}} \cdot \frac{\partial u_{01}}{\partial g_{h1}}$

$$\frac{\partial E_{01}}{\partial u_{01}} = \frac{\partial E_{01}}{\partial g_{01}} \cdot \frac{\partial g_{01}}{\partial u_{01}} = 0,74136507 \cdot 0,186815602 = \boxed{0,138498562}$$

$$u_{01} = w_5 \cdot g_{h1} + w_6 \cdot g_{h2} + b_2$$

$$\frac{\partial u_{01}}{\partial g_{h1}} = w_5 \cdot 1 \cdot g_{h1}^{-1} + 0 + 0 = w_5 = \boxed{0,4}$$

$$\frac{\partial E_{01}}{\partial g_{h1}} = 0,138498562 \cdot 0,4 = \boxed{0,055399425}$$

(a2) $\frac{\partial E_{02}}{\partial g_{h1}} = \frac{\partial E_{02}}{\partial u_{02}} \cdot \frac{\partial u_{02}}{\partial g_{h1}}$

$$\frac{\partial E_{02}}{\partial u_{02}} = \frac{\partial E_{02}}{\partial g_{02}} \cdot \frac{\partial g_{02}}{\partial u_{02}} = (-0,217071535) \cdot 0,175510053 = \boxed{-0,038098237}$$

$$u_{02} = w_7 \cdot g_{h1} + w_8 \cdot g_{h2} + b_2$$

$$\frac{\partial u_{02}}{\partial g_{h1}} = w_7 \cdot 1 \cdot g_{h1}^{-1} + 0 + 0 = w_7 = \boxed{0,5}$$

$$\frac{\partial E_{02}}{\partial g_{h1}} = (-0,038098237) \cdot 0,5 = \boxed{-0,019049118}$$

Ⓐ $\frac{\partial E_{TOTAL}}{\partial g_{h_1}} = 0,055399425 + (-0,019049118) = \boxed{0,036350307}$

Ⓑ $\frac{\partial g_{h_1}}{\partial u_{h_2}}$

$$g_{h_1} = \frac{1}{1 + e^{-u_{h_2}}}$$

$$\frac{\partial g_{h_1}}{\partial u_{h_2}} = g_{h_1} \cdot (1 - g_{h_1}) = 0,59326999 \cdot (1 - 0,59326999) = \boxed{0,241300709}$$

Ⓒ $\frac{\partial u_{h_1}}{\partial w_1}$

$$u_{h_1} = w_1 \cdot x_1 + w_2 \cdot x_2 + b_1$$

$$\frac{\partial u_{h_1}}{\partial w_1} = 1 \cdot w_1 \cdot x_1 + 0 + 0 = x_1 = \boxed{0,05}$$

$$\frac{\partial E_{TOTAL}}{\partial w_1} = 0,036350307 \cdot 0,241300709 \cdot 0,05 = \boxed{0,000438568}$$

$$w_1(t+1) = w_1(t) - \eta \cdot \frac{\partial E_{TOTAL}}{\partial w_1} = 0,15 - 0,5 \cdot 0,000438568 = \boxed{0,149780736}$$

AJUSTAR w_2 :

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_2} = \frac{\partial E_{\text{TOTAL}}}{\partial g_{h1}} \cdot \frac{\partial g_{h1}}{\partial w_2} \cdot \frac{\partial u_{h1}}{\partial w_2}$$

$$u_{h1} = w_1 \cdot x_1 + w_2 \cdot x_2 + b_1$$

$$\frac{\partial u_{h1}}{\partial w_2} = 0 + 1 \cdot w_2^{1+} \cdot x^2 + 0 = x^2 = \boxed{0,1}$$

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_2} = 0,036350307 \cdot 0,241308209 \cdot 0,1 = \boxed{0,000877335}$$

$$w_2(t+1) = w_2(t) - n \cdot \frac{\partial E_{\text{TOTAL}}}{\partial w_2} = 0,2 - 0,5 \cdot 0,000877335 = \boxed{0,199561432}$$

AJUSTAR w_3

$$\frac{\partial E_{total}}{\partial w_3} = \underbrace{\frac{\partial E_{total}}{\partial g_{h2}}}_{\textcircled{d}} \cdot \underbrace{\frac{\partial g_{h2}}{\partial u_{h2}}}_{\textcircled{e}} \cdot \underbrace{\frac{\partial u_{h2}}{\partial w_3}}_{\textcircled{f}}$$

$$\textcircled{d} \quad \frac{\partial E_{total}}{\partial g_{h2}} = \underbrace{\frac{\partial E_{o1}}{\partial g_{h2}}}_{\textcircled{d1}} + \underbrace{\frac{\partial E_{o2}}{\partial g_{h2}}}_{\textcircled{d2}}$$

$$\textcircled{d1} \quad \frac{\partial E_{o1}}{\partial g_{h2}} = \underbrace{\frac{\partial E_{o1}}{\partial u_{o1}}}_{\textcircled{m1}} \cdot \underbrace{\frac{\partial u_{o1}}{\partial g_{h2}}}_{\textcircled{m2}}$$

$$\frac{\partial E_{o1}}{\partial u_{o1}} = 0,138498562$$

$$u_{o1} = w_5 \cdot g_{h2} + w_6 \cdot g_{h2} + b_2$$

$$\frac{\partial u_{o1}}{\partial g_{h2}} = 0 + w_6 \cdot 1 \cdot g_{h2}^{1-1} + 0 = w_6 = 0,45$$

$$\frac{\partial E_{o1}}{\partial g_{h2}} = 0,138498562 \cdot 0,45 = 0,06324353$$

$$\textcircled{d2} \quad \frac{\partial E_{o2}}{\partial g_{h2}} = \underbrace{\frac{\partial E_{o2}}{\partial u_{o2}}}_{\textcircled{m2}} \cdot \underbrace{\frac{\partial u_{o2}}{\partial g_{h2}}}_{\textcircled{m3}}$$

$$\frac{\partial E_{o2}}{\partial u_{o2}} = -0,038098237$$

$$u_{o2} = w_7 \cdot g_{h2} + w_8 \cdot g_{h2} + b_2$$

$$\frac{\partial u_{o2}}{\partial g_{h2}} = 0 + w_8 \cdot 1 \cdot g_{h2}^{1-1} + 0 = w_8 = 0,55$$

$$\frac{\partial E_{o2}}{\partial g_{h2}} = -0,038098237 \cdot 0,55 = -0,02095403$$

$$\textcircled{d} \frac{\partial E_{\text{Total}}}{\partial g_{h2}} = 0,06324353 + (-0,02095403) = \boxed{0,042370323}$$

$$\textcircled{e} \frac{\partial g_{h2}}{\partial w_{h2}}$$

$$g_{h2} = \frac{1}{1 + e^{-w_{h2}}}$$

$$\frac{\partial g_{h2}}{\partial w_{h2}} = g_{h2} \cdot (1 - g_{h2}) = 0,0596884378 \cdot (1 - 0,0596884378) = \boxed{0,240613417}$$

$$\textcircled{f} \frac{\partial w_{h2}}{\partial w_3}$$

$$w_{h2} = w_3 \cdot x_1 + w_4 \cdot x_2 + b_1$$

$$\frac{\partial w_{h2}}{\partial w_3} = 1 \cdot w_3 \cdot x_1 + 0 + 0 = x_1 = \boxed{0,05}$$

$$\frac{\partial E_{\text{Total}}}{\partial w_3} = 0,042370323 \cdot 0,240613417 \cdot 0,5 = \boxed{0,000497713}$$

$$w_3(t+1) = w_3(t) - \eta \cdot \frac{\partial E_{\text{Total}}}{\partial w_3} = 0,25 - 0,5 \cdot 0,000497713 = \boxed{0,249751144}$$

Ajustar w_4 :

$$\frac{\partial E_{\text{Total}}}{\partial w_4} = \frac{\partial E_{\text{Total}}}{\partial g_{h2}} \cdot \frac{\partial g_{h2}}{\partial w_{h2}} \cdot \frac{\partial w_{h2}}{\partial w_4}$$

$$w_{h2} = w_3 \cdot x_1 + w_4 \cdot x_2 + b_1$$

$$\frac{\partial w_{h2}}{\partial w_4} = 0 + 1 \cdot w_4 \cdot x_2 + 0 = x_2 = \boxed{0,1}$$

$$\frac{\partial E_{\text{Total}}}{\partial w_4} = 0,042370323 \cdot 0,240613417 \cdot 0,1 = \boxed{0,00095425}$$

$$w_4(t+1) = w_4(t) - \eta \cdot \frac{\partial E_{\text{Total}}}{\partial w_4} = 0,3 - 0,5 \cdot 0,00095425 = \boxed{0,299502287}$$

AJUSTAR $b_1 h_1$:

$$\frac{\delta E_{TOTAL}}{\delta b_1 h_1} = \frac{\delta E_{TOTAL}}{\delta g_{h1}} \cdot \frac{\delta g_{h1}}{\delta h_1} \cdot \frac{\delta w_{h1}}{\delta b_1 h_1}$$

$$\frac{\delta E_{TOTAL}}{\delta g_{h1}} = \frac{\delta E_{TOTAL}}{\delta g_{h2}} + \frac{\delta E_{TOTAL}}{\delta g_{h1}}$$

$$\frac{\delta E_{TOTAL}}{\delta b_1 h_1} = -(y_2 - g_{o1}) \cdot g_{o1} \cdot (1 - g_{o1}) \cdot w_5 + (-y_2 - g_{o2}) \cdot g_{o2} \cdot (1 - g_{o2}) \cdot w_7 \\ \cdot g_{h2} \cdot (1 - g_{h2}) \cdot 1$$

$$\frac{\delta E_{TOTAL}}{\delta b_1 h_1} = 0,008771355$$

$$b_1 h_1(t+1) = 0,35 - 0,5 \cdot 0,008771355 = \boxed{0,345614323}$$

AJUSTAR $b_1 h_2$:

$$\frac{\delta E_{TOTAL}}{\delta b_1 h_2} = \frac{\delta E_{TOTAL}}{\delta g_{h2}} \cdot \frac{\delta g_{h2}}{\delta h_2} \cdot \frac{\delta w_{h2}}{\delta b_1 h_2}$$

$$\frac{\delta E_{TOTAL}}{\delta b_1 h_2} = -(y_2 - g_{o1}) \cdot g_{o1} \cdot (1 - g_{o1}) \cdot w_6 + (-y_2 - g_{o2}) \cdot g_{o2} \cdot (1 - g_{o2}) \cdot w_8 \cdot g_{h2} \cdot (1 - g_{h2}),$$

$$\frac{\delta E_{TOTAL}}{\delta b_1 h_2} = 0,00982688$$

$$b_1 h_2(t+1) = 0,35 - 0,5 \cdot 0,00982688 = \boxed{0,345008656}$$

EPOCA = 2, PESOS INICIAIS DESTA EPOCA:

$$w_1 = 0,149780756$$

$$w_5 = 0,35893648$$

$$w_2 = 0,199561432$$

$$w_6 = 0,40866686$$

$$w_3 = 0,249751144$$

$$w_7 = 0,51130127$$

$$w_4 = 0,299502284$$

$$w_8 = 0,561370121$$

$$bh_1 = 0,345614323$$

$$b_{o1} = 0,530750719$$

$$bh_2 = 0,345008656$$

$$b_{o2} = 0,639049118$$

FASE 1 - PROPAGAÇÃO

$$uh_1 = x_1 \cdot w_1 + x_2 \cdot w_2 + bh_1$$

$$u_{o1} = g(h_1) \cdot w_5 + g(h_2) \cdot w_6 + b_{o1}$$

$$uh_1 = 0,373059502$$

$$u_{o1} = 0,986729661$$

$$g(h_1) = g(uh_1) = \frac{1}{1 + e^{-uh_1}}$$

$$g(h_1) = \boxed{0,728441485}$$

$$gh_1 = \boxed{0,592198055}$$

$$uh_2 = x_3 \cdot w_3 + x_4 \cdot w_4 + bh_2$$

$$u_{o2} = g(h_1) \cdot w_7 + g(h_2) \cdot w_8 + b_{o2}$$

$$uh_2 = 0,387446492$$

$$u_{o2} = 1,256230858$$

$$g(h_2) = g(uh_2) = \frac{1}{1 + e^{-uh_2}}$$

$$g(h_2) = \boxed{0,778376589}$$

$$gh_2 = \boxed{0,595667831}$$

ERRO TOTAL:

$$E_{\text{TOTAL}} = E_{01} + E_{02} = \frac{1}{2} (y_1 - g_{01})^2 + \frac{1}{2} (y_2 - g_{02})^2$$

$$E_{\text{TOTAL}} = 0,258079084 + 0,258079084$$

$$E_{\text{TOTAL}} = \boxed{0,280471318}$$

FASE 2 - RETROPROPAGAÇÃO TAXA DE APRENDIZADO = 0,5
ADJUSTAR W5:

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_5} = \frac{\partial E_{\text{TOTAL}}}{\partial g_{01}} \cdot \frac{\partial g_{01}}{\partial u_{01}} \cdot \frac{\partial u_{01}}{\partial w_5}$$

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_5} = -(y_1 - g_{01}) \cdot g_{01} \cdot (1 - g_{01}) \cdot g_{h1}$$

$$\frac{\partial E_{\text{TOTAL}}}{\partial w_5} = 0,084162083$$

$$w_5(t+1) = w_5(t) - \eta \cdot \frac{\partial E_{\text{TOTAL}}}{\partial w_5}$$

$$w_5(t+1) = 0,35891648 - 0,5 \cdot 0,084162083 = \boxed{0,316835438}$$

AJUSTAR w_6 :

$$\frac{\partial E_{TOTAL}}{\partial w_6} = \frac{\partial E_{TOTAL}}{\partial g_{01}} \cdot \frac{\partial g_{01}}{\partial u_{01}} \cdot \frac{\partial u_{01}}{\partial w_6}$$

$$\frac{\partial E_{TOTAL}}{\partial w_6} = -(y_s - g_{01}) \cdot g_{01} \cdot (1 - g_{01}) \cdot g_{h2}$$

$$\frac{\partial E_{TOTAL}}{\partial w_6} = 0,084655201$$

$$w_6(t+1) = w_6(t) - \eta \cdot \frac{\partial E_{TOTAL}}{\partial w_6}$$

$$w_6(t+1) = 0,408666386 - 0,5 \cdot 0,084655201 = \boxed{0,366338586}$$

AJUSTAR w_7 :

$$\frac{\partial E_{TOTAL}}{\partial w_7} = \frac{\partial E_{TOTAL}}{\partial g_{02}} \cdot \frac{\partial g_{02}}{\partial u_{02}} \cdot \frac{\partial u_{02}}{\partial w_7}$$

$$\frac{\partial E_{TOTAL}}{\partial w_7} = -(y_s - g_{02}) \cdot g_{02} \cdot (1 - g_{02}) \cdot g_{h1}$$

$$\frac{\partial E_{TOTAL}}{\partial w_7} = -0,021619024$$

$$w_7(t+1) = w_7(t) - \eta \cdot \frac{\partial E_{TOTAL}}{\partial w_7}$$

$$w_7(t+1) = 0,51130127 - 0,5 \cdot (-0,02619024) = \boxed{0,52213082}$$

Ajustar w8:

$$\frac{\partial E_{TOTAL}}{\partial w_8} = \frac{\partial E_{TOTAL}}{\partial g_{02}} \cdot \frac{\partial g_{02}}{\partial u_{02}} \cdot \frac{\partial u_{02}}{\partial w_8}$$

$$\frac{\partial E_{TOTAL}}{\partial w_8} = -(y_2 - g_{02}) \cdot g_{02} \cdot (1-g_{02}) \cdot g_{h2}$$

$$\frac{\partial E_{TOTAL}}{\partial w_8} = -0,021745693$$

Ajustar b₀₂:

$$\frac{\partial E_{TOTAL}}{\partial b_{02}} = \frac{\partial E_{TOTAL}}{\partial g_{02}} \cdot \frac{\partial g_{02}}{\partial u_{02}} \cdot \frac{\partial u_{02}}{\partial b_{02}}$$

$$\frac{\partial E_{TOTAL}}{\partial b_{02}} = -(y_2 - g_{02}) \cdot g_{02} \cdot (1-g_{02}) \cdot 1$$

$$\frac{\partial E_{TOTAL}}{\partial b_{02}} = 0,142118134$$

$$b_{02}(t+1) = 0,530750719 - 0,5 \cdot 0,142118134 = \boxed{0,459691652}$$

AJUSTAR b_{02} :

$$\frac{\delta E_{TOTAL}}{\delta b_{02}} = \frac{\delta E_{TOTAL}}{\delta g_{02}} \cdot \frac{\delta g_{02}}{\delta u_{02}} \cdot \frac{\delta u_{02}}{\delta b_{02}}$$

$$\frac{\delta E_{TOTAL}}{\delta b_{02}} = -(y_2 - g_{02}) \cdot g_{02} \cdot (1 - g_{02}) \cdot 1$$

$$\frac{\delta E_{TOTAL}}{\delta b_{02}} = -0,036506409$$

$$b_{02}(t+1) = 0,619019118 - 0,5 \cdot (-0,036506409) = \boxed{0,637302323}$$

$$w_8(t+1) = w_8(t) - \eta \cdot \frac{\delta E_{TOTAL}}{\delta w_8}$$

$$w_8(t+1) = 0,561370121 - 0,5 \cdot (-0,021745693) = \boxed{0,572242968}$$

AJUSTAR w_1 :

$$\frac{\delta E_{TOTAL}}{\delta w_1} = \frac{\delta E_{TOTAL}}{\delta g_{01}} \cdot \frac{\delta g_{01}}{\delta u_{01}} \cdot \frac{\delta u_{01}}{\delta w_1}$$

$$\frac{\delta E_{TOTAL}}{\delta w_1} = \left(\frac{\delta E_{01}}{\delta g_{01}} \cdot \frac{\delta g_{01}}{\delta u_{01}} \cdot \frac{\delta u_{01}}{\delta g_{01}} + \frac{\delta E_{02}}{\delta g_{02}} \cdot \frac{\delta g_{02}}{\delta u_{02}} \cdot \frac{\delta u_{02}}{\delta g_{01}} \right) \cdot g_{01} \cdot (1 - g_{01}) \cdot x_1$$

$$\frac{\delta E_{TOTAL}}{\delta w_1} = -(y_1 - g_{01}) \cdot g_{01} \cdot (1 - g_{01}) \cdot w_5 + (-y_2 - g_{02}) \cdot g_{02} \cdot (1 - g_{02}) \cdot w_7 \cdot g_{01} \cdot (1 - g_{01}) \cdot x_1$$

$$\frac{\delta E_{TOTAL}}{\delta w_1} = 0,000390538$$

$$w_1(t+1) = w_1(t) - \eta \frac{\delta E_{TOTAL}}{\delta w_1}$$

$$w_3(t+1) = 0,149780736 - 0,5 \cdot 0,000380538 = \boxed{0,149585447}$$

AJUSTAR w_2 :

$$\frac{\partial E_{TOTAL}}{\partial w_2} = -(y_1 - g_{01}) \cdot g_{01} \cdot (1-g_{01}) \cdot w_5 + -(y_2 - g_{02}) \cdot g_{02} \cdot (1-g_{02}) \cdot w_7 \cdot g_{12} \cdot g_{02} \cdot x_2$$

$$\frac{\partial E_{TOTAL}}{\partial w_2} = 0,000781076$$

$$w_2(t+1) = w_2(t) - \eta \frac{\partial E_{TOTAL}}{\partial w_2}$$

$$w_2(t+1) = 0,199573432 - 0,5 \cdot 0,000781076 = \boxed{0,199170894}$$

AJUSTAR w_3 :

$$\frac{\partial E_{TOTAL}}{\partial w_3} = -(y_1 - g_{01}) \cdot g_{01} \cdot (1-g_{01}) \cdot w_6 + -(y_2 - g_{02}) \cdot g_{02} \cdot (1-g_{02}) \cdot w_8 \cdot g_{12} \cdot (1-g_{12}) \cdot x_1$$

$$\frac{\partial E_{TOTAL}}{\partial w_3} = 0,000452636$$

$$w_3(t+1) = w_3(t) - \eta \frac{\partial E_{TOTAL}}{\partial w_3}$$

$$w_3(t+1) = 0,249753244 - 0,5 \cdot 0,000452636 = \boxed{0,249524836}$$

AJUSTAR w_4 :

$$\frac{\partial E_{TOTAL}}{\partial w_4} = -(y_1 - g_{01}) \cdot g_{01} \cdot (1-g_{01}) \cdot w_6 + -(y_2 - g_{02}) \cdot g_{02} \cdot (1-g_{02}) \cdot w_8 \cdot g_{12} \cdot (1-g_{12}) \cdot x_2$$

$$\frac{\partial E_{TOTAL}}{\partial w_4} = 0,000905232$$

$$w_4(t+1) = 0,299502284 - 0,5 \cdot 0,000905232 = \boxed{0,299049671}$$

Ajustar bh_1 :

$$\frac{\partial E_{TOTAL}}{\partial bh_1} = \frac{\partial E_{TOTAL}}{\partial g_{bh_1}} \cdot \frac{\partial g_{bh_1}}{\partial w_{bh_1}} \cdot \frac{\partial w_{bh_1}}{\partial bh_1}$$

$$\frac{\partial E_{TOTAL}}{\partial bh_1} = (-(y_1 - g_{\theta 1}) \cdot g_{\theta 1} \cdot (1 - g_{\theta 1}) \cdot w_5 + (-(y_2 - g_{\theta 2}) \cdot g_{\theta 2} \cdot (1 - g_{\theta 2}) \cdot w_7) g_{bh_1} \cdot (1 - g_{bh_1}) \cdot 1$$

$$\frac{\partial E_{TOTAL}}{\partial Bh_1} = 0,007810763$$

$$bh_1 = 0,345614323 - 0,5 \cdot 0,007810763 = \boxed{0,343708941}$$

AJUSTAR b_{h_2} :

$$\frac{\partial E_{TOTAL}}{\partial b_{h_2}} = \frac{\partial E_{TOTAL}}{\partial g_{h_2}} \cdot \frac{\partial g_{h_2}}{\partial w_{h_2}} \cdot \frac{\partial w_{h_2}}{\partial b_{h_2}}$$

$$\frac{\partial E_{TOTAL}}{\partial b_{h_2}} = \left(-(y_1 - g_{o1}) \cdot g_{o1} \cdot (1 - g_{o1}) \cdot w_6 + \left(-(y_2 - g_{o2}) \cdot g_{o2} \cdot (1 - g_{o2}) \cdot w_8 \right) \cdot g_{h_2} \cdot (1 - g_{h_2}) \cdot 1 \right)$$

$$\frac{\partial E_{TOTAL}}{\partial b_{h_2}} = 0,009076824$$

$$b_{h_2} = 0,345008656 - 0,5 \cdot 0,009076824 = \boxed{0,340470244}$$

PESOS AO FINAL DA EPOCA 2:

$$w_1 = 0,349585447$$

$$w_5 = 0,31683538$$

$$w_2 = 0,199170894$$

$$w_6 = 0,346338586$$

$$w_3 = 0,249524830$$

$$w_7 = 0,522110182$$

$$w_4 = 0,299049671$$

$$w_8 = 0,572242968$$

$$b_{h_1} = 0,341708941$$

$$b_{o1} = 0,459491652$$

$$b_{h_2} = 0,340470244$$

$$b_{o2} = 0,637302323$$