

**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 1**

**Parte 1 - Exercícios conceituais**

**1. Explique com suas palavras o que é um neurônio artificial e o seu funcionamento traçando uma analogia com o neurônio biológico.**

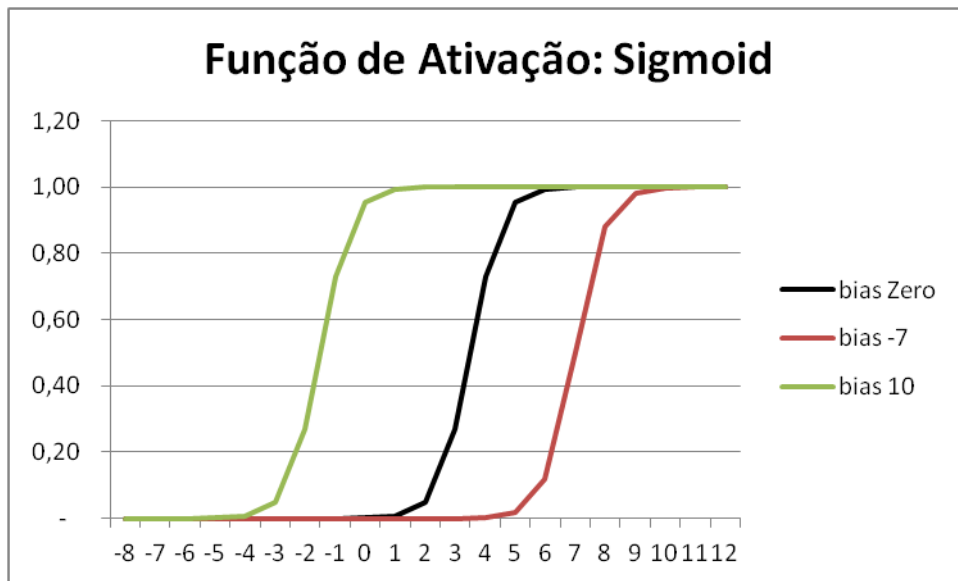
Um neurônio artificial é uma estrutura que recebe informações, as processa e repassa para o próximo neurônio. Cada informação recebida tem um peso que é ponderado durante o processamento. É semelhante ao neurônio biológico que recebe estímulos elétricos advindos de reações físico-químicas de várias fontes, como receptor (tato, visão, olfato, etc.) e outros neurônios. O corpo celular é responsável por processar as informações recebidas com potencial de ativação que indicará se o neurônio poderá disparar um impulso elétrico ao longo de seu axônio para transmitir estímulos para outras estruturas (células, tecidos, nervos, neurônios, etc.).

**2. Explique o objetivo do uso do Bias.**

O objetivo do uso do Bias é aumentar ou diminuir a influencia das entradas no neurônio. Essa influencia está diretamente relacionada com a função de ativação. Por exemplo: se a função de ativação for bipolar que é ativada apenas quando o valor  $fx$  for maior que 0, então somente será maior que 0 se  $(w \cdot x > -b)$ .

**3. Explique matematicamente em um plano o efeito do uso do Bias na definição de um hiperplano para um problema hipotético.**

O valor de bias tem efeito na função de ativação, ou seja, a partir de qual valor a função será ativada. Veja abaixo exemplo da função sigmoid. Quando o bias é igual a zero, a função de ativação começa a crescer a partir do ponto 0. Já quando o bias é negativo, no caso -7, a função de ativação começa a crescer depois o ponto 0. Quando o bias é positivo, no caso 10, a função de ativação começa a crescer antes do ponto 0.



#### 4. O que são e para que servem os notebooks?

Notebook é uma interface que combina código e documentação, ou ainda, é uma combinação de codificação com Wikipedia onde se pode escrever textos, fórmulas, gráficos, etc. Como a codificação é disposta em várias células, é possível executar trechos de códigos avaliando os resultados obtidos em cada passo. Também é possível executar códigos na ordem e quantas vezes desejar. Nos notebooks é possível selecionar a linguagem bem como sua versão.

#### 5. Explique conceitualmente o funcionamento do tensorflow estabelecendo uma relação entre nós e arestas de um grafo.

O tensorflow funciona com uma estrutura de grafos que é definido por arestas (tensors ou arrays de dados multidimensionais) e no nó são definidas as operações matemáticas a serem realizadas. Ao final da operação no nó, o resultado é enviado para o próximo grafo que recebe as informações em suas arestas. Assim o processamento é propagado ao longo da rede de grafos.

#### 6. Quais as principais diferenças no fluxo de execução no tensorflow em relação a estruturas de dados convencionais?

Em estruturas de dados convencionais costumam ser executadas de forma sequencial enquanto no tensorflow tem a flexibilidade de quebrar a execução em um ou vários CPUs ou GPUs em computadores, servidores ou dispositivos móveis utilizando a mesma API.

#### 7. Descreva o objetivo do uso da função de ativação de um neurônio.

A função de ativação faz o mapeamento de uma entrada em um domínio conhecido, por exemplo, uma saída binária. Assim cada neurônio executa sua operação matemática e ao invés de transmitir o resultado desta operação para o próximo neurônio, o valor passa por uma função de ativação que retorno o valor mapeado.

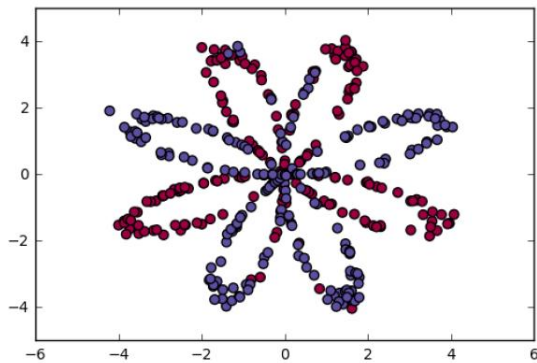
#### 8. Qual o objetivo do uso de pesos nos neurônios artificiais? Dica: Faça uma analogia entre sistemas biológicos e artificiais.

Neurônios recebem estímulos de vários sensores como visão, olfato, paladar, etc. Nossa rede neural pondera o peso de cada um destes estímulos de acordo com o indivíduo ou de acordo com a situação de momento. Por exemplo, se sentir cheiro de gás dentro de casa, o indivíduo

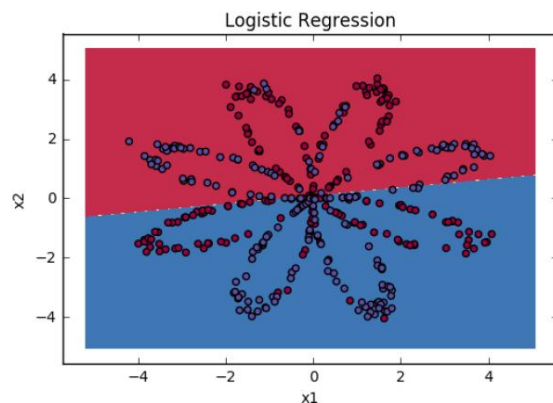
irá dar maior importância ao olfato. Os neurônios artificiais funcionam de forma semelhante no sentido que recebem entradas a serem processadas considerando o peso de cada uma. Assim é possível ponderar que entrada tem maior peso sobre as demais.

9. Um neurônio artificial que usa uma função de ativação não-linear é capaz de tratar problemas não-lineares? Justifique sua resposta.

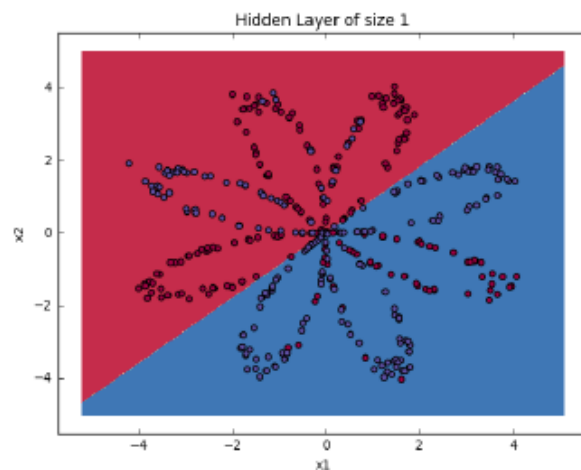
Não, problemas não-lineares não podem ser tratados apenas com um neurônio artificial com função de ativação não linear. Pois não é possível fazer a classificação de pontos (2 dimensões) utilizando apenas uma reta. Exemplo de funções não-lineares: função XOR, exemplo abaixo, etc.



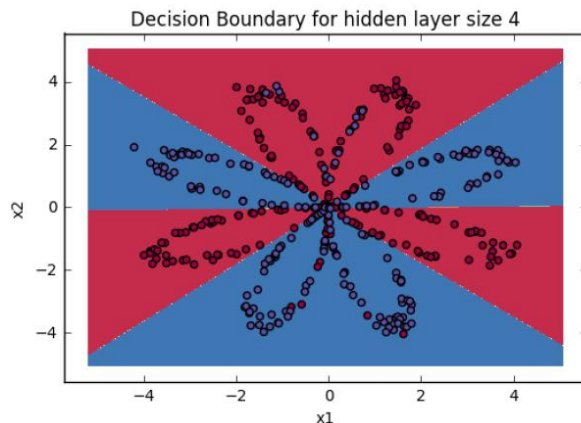
Aplicação de Regressão Logística Simples:



Aplicação de rede neural de 1 camada apenas:



Aplicação de rede neural de 4 camadas:



**10.** Se um neurônio artificial do tipo perceptron for inicializado com todos os pesos iguais a zero ele irá convergir para uma solução aceitável no problema?

Sim, mesmo iniciando com valor zeros para os pesos se obterá uma solução aceitável. Para os valores de peso zero, a função de ativação retornará zero (ativo quando maior que 0,5 por exemplo). Porém a diferença entre o resultado esperado e o retornado será zero ou um, para uma classificação binária. Logo o Perceptron ajustará os pesos nas amostras que forem divergentes, neste momento os pesos vão deixar de ser zero e iram caminhar para uma solução aceitável.

**11.** Um neurônio artificial após o seu treinamento apresenta todos os pesos iguais. Como você interpretaria esse resultado?

Como os pesos são iguais, significa dizer que a linha de separação da classificação cruza os eixos no mesmo valor formando um triângulo equilátero com o ponto zero.

**12.** Em caso do NÃO-USO do bias em um neurônio artificial o resultado do treinamento será o mesmo?

Não, o resultado não será o mesmo, porém terão resultados aceitáveis sem o bias. A parte da função que calcula dos pesos tem maior relevância no calculo da função. O bias tem peso menor já que apenas é somado ao resultado final da função.

**13.** Para que serve a taxa de aprendizado no processo de treinamento de um neurônio artificial?

A taxa de aprendizado é um valor aplicado na atualização dos pesos de cada neurônio. Deve ser calibrado para não ser tão pequeno para não levar a uma grande demora na convergência dos resultados, como também não deve ser tão grande para não impactar na chegada de um valor ótimo de convergência.

#### 14. O que é e para que serve o gradiente?

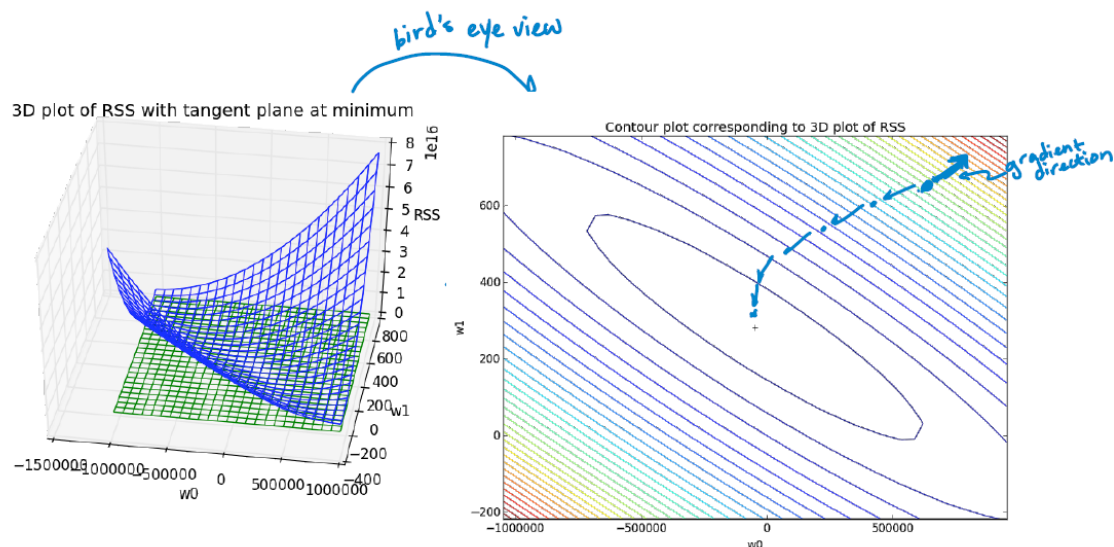
a. Exemplifique conceitualmente. Dica: Exponha sua resposta para uma criança de 10 anos e verifique se ela entendeu.

O gradiente é como uma nota vermelha no boletim da criança. Quando surgir uma nota vermelha no boletim, indica para a criança que ela não está no caminho certo naquela matéria. Logo a direção dela tem que ser oposta à aquela que estava quando obteve uma nova vermelha.

b. Exemplifique matematicamente

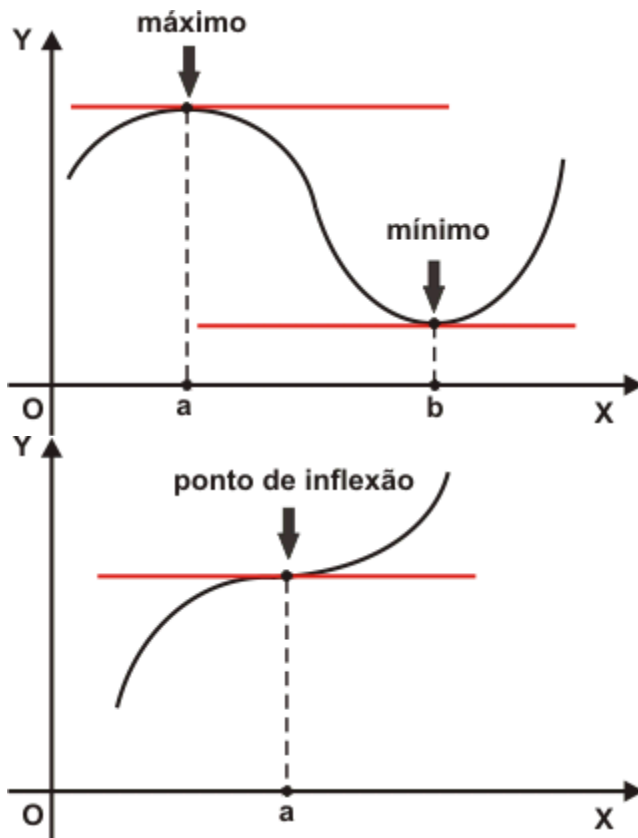
O gradiente é um valor que direciona para chegar a um resultado, máximo ou mínimo de uma função. Por exemplo: com o gradiente mínimo deseja-se chegar ao menor valor de erro possível. Assim o gradiente fornece o valor que direciona a função a chegar ao mínimo possível, ajustando a cada época os pesos dos neurônios.

Na figura abaixo representa um exemplo de ajuste de dois pesos:  $w_1$  e  $w_2$ . No gráfico à esquerda, é apresentado estes dois pesos com o RSS que é o erro quadrático. Veja que na figura à direita, a partir de um ponto qualquer inicial, o gradiente informa a direção para se obter o mínimo. Neste caso obter o mínimo erro RSS.



15. Sempre que a derivada de uma função é nula podemos afirmar que a função passa por um máximo ou mínimo? Dicas nas imagens abaixo:

Não, sempre que a derivada de uma função é zero significa dizer que naquele ponto pode representar o máximo, mínimo ou uma inflexão.



16. Dado a função:

$$y = f(x) = x^4 * 3x^3 + 2$$

escolha um valor aleatório de  $x$  e mostre com um exemplo numérico como o cálculo da derivada leva ao ponto de mínimo da função com uma determinada taxa de aprendizado.

Sendo  $y = x^4 * 3x^3 + 2$

A derivada de  $y = (x^4)' * (3x^3) + (x^4) * (3x^3)'$

$$y' = (4x^3) * (3x^3) + (x^4) * (9x^2)$$

$$y' = 12x^6 + 9x^6$$

$$y' = 21x^6$$

A derivada de  $y = 21x^6$

Definindo um ponto qualquer  $x = 3$

Derivada de  $y$  em relação a  $x$  quando  $x = 3 \rightarrow 15309$

Como o valor da derivada é maior que 0, no ponto  $x = 3$  a função está crescendo.

Para encontrar o mínimo, devemos reduzir o valor de  $x$  (reduzir porque o gradiente é positivo).

Vamos reduzir de 3 para 2.

Derivada de  $y$  em relação a  $x$  quando  $x = 2 \rightarrow 1344$

Como o valor da derivada é maior que 0, no ponto  $x = 2$  a função está crescendo.

Vamos reduzir de 2 para 1.

Derivada de  $y$  em relação a  $x$  quando  $x = 1 \rightarrow 21$

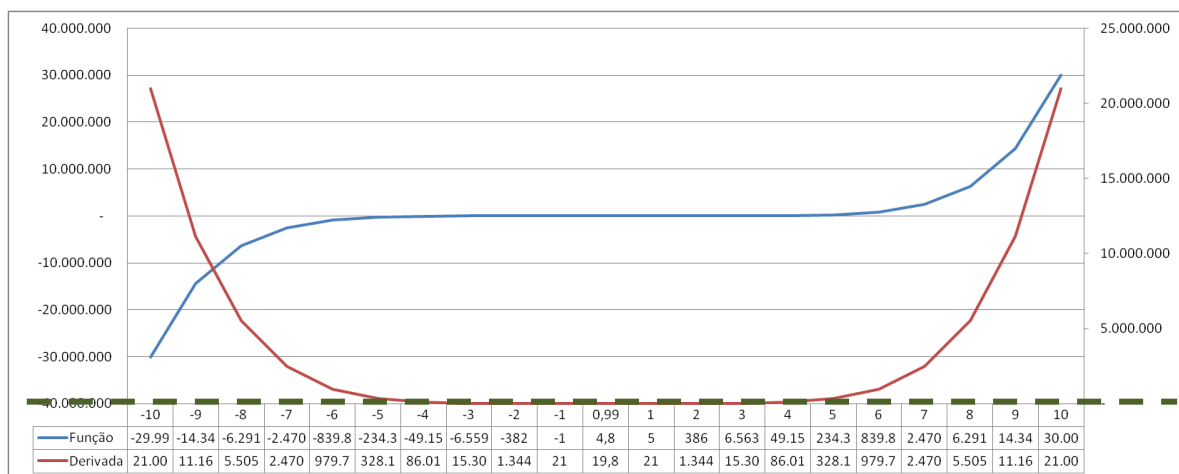
Como o valor da derivada é maior que 0, no ponto  $x = 1$  a função está crescendo.

Vamos reduzir de 1 para 0.

Derivada de  $y$  em relação a  $x$  quando  $x = 0 \rightarrow 0$

Logo  $x = 0$  pode ser o ponto mínimo da função ou sua inflexão. Neste caso é uma inflexão, pois se reduzirmos mais uma vez o valor de  $x$  de 0 para -1, temos uma derivada de 21. Como é positivo, devemos reduzir ainda mais o valor de  $x$  para encontrar o mínimo. Assim o mínimo da função  $x$  é mínimo quando  $x$  tende a menos infinito.

Mais detalhes no gráfico abaixo:



Se a função fosse:

$$y = f(x) = x^4 - 3x^3 + 2$$

Sendo  $y = x^4 - 3x^3 + 2$

A derivada de  $y = 4x^3 - 9x^2$

Definindo um ponto qualquer  $x = 3$

Derivada de  $y$  em relação a  $x$  quando  $x = 3 \rightarrow 27$

Como o valor da derivada é maior que 0, no ponto  $x = 3$  a função está crescendo.

Para encontrar o mínimo, devemos reduzir o valor de  $x$  (reduzir porque o gradiente é positivo).

Vamos reduzir de 3 para 2.

Derivada de  $y$  em relação a  $x$  quando  $x = 2 \rightarrow -4$

Como o valor da derivada é menor que 0, no ponto  $x = 2$  a função está decrescendo.

Vamos aumentar de 2 para 2,5.

Derivada de  $y$  em relação a  $x$  quando  $x = 2,5 \rightarrow 6,25$

Como o valor da derivada é maior que 0, no ponto  $x = 2,5$  a função está crescendo.

Vamos reduzir de 2,5 para 2,4.

Derivada de  $y$  em relação a  $x$  quando  $x = 2,4 \rightarrow 3,46$

Como o valor da derivada é maior que 0, no ponto  $x = 2,4$  a função está crescendo.

Vamos reduzir de 2,4 para 2,3.

Derivada de  $y$  em relação a  $x$  quando  $x = 2,3 \rightarrow 1,06$

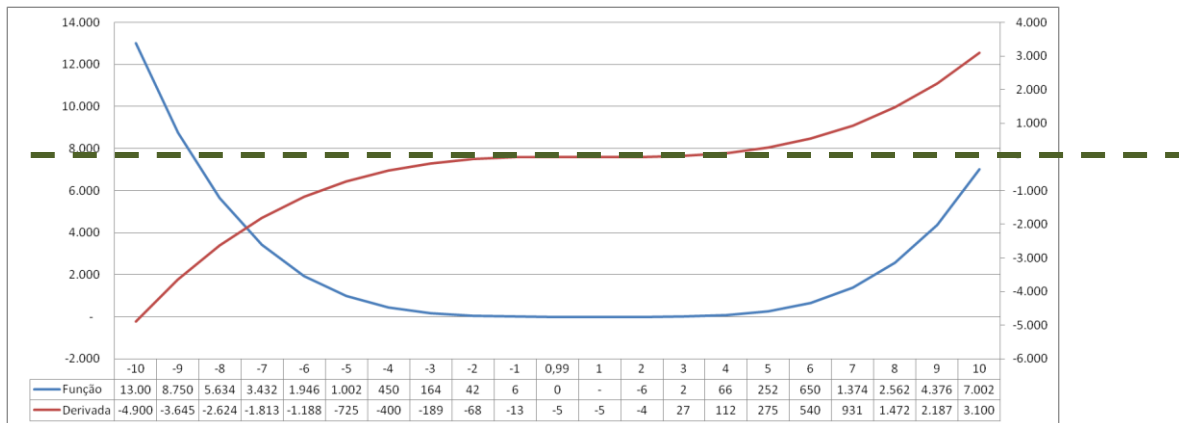
Como o valor da derivada é maior que 0, no ponto  $x = 2,3$  a função está crescendo.

Vamos reduzir de 2,3 para 2,2.

Derivada de  $y$  em relação a  $x$  quando  $x = 2,2 \rightarrow -0,97$

Logo  $x$  entre 2,2 e 2,3 é o ponto mínimo.





Como apresentado no gráfico acima a linha azul representa a função, já a linha vermelha a sua derivada. Note que a derivada é positiva quando a função é crescente, negativa quando a função é decrescente e zero quando ocorre a inflexão da curva.

### 17. Qual a diferença entre o algoritmo de treinamento do neurônio perceptron e o Adaline?

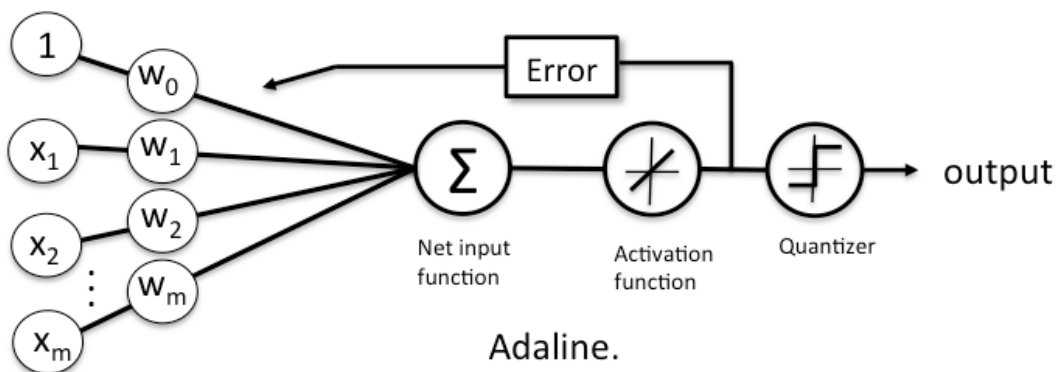
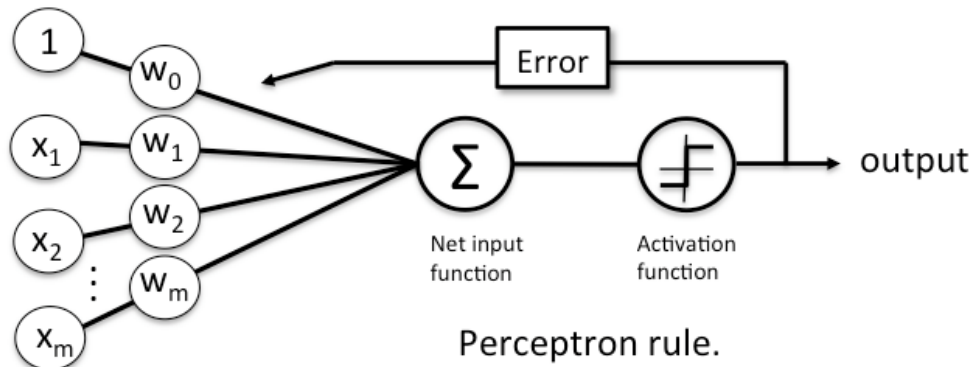
Primeiramente, o que perceptron e Adaline têm em comum:

- Ambos são classificadores binários
- Ambos têm limite de decisão linear
- Ambos podem aprender de forma iterativa, amostra por amostra. Porém Adaline utiliza-se a via descendente de gradiente estocástico.

As diferenças são:

- O Adaline é similar ao Perceptron, com diferença apenas pelo seu algoritmo de treinamento. Enquanto o Perceptron ajusta os pesos somente quando um padrão é classificado incorretamente, o Adaline utiliza a regra Delta para minimizar o erro médio (MSE) após cada padrão ser apresentado, ajustando os pesos proporcionalmente ao erro.
- Perceptron utiliza rótulos de classes para aprender os coeficientes.
- Adaline utiliza valores preditos contínuos (da entrada de rede) para aprender os coeficientes do modelo, que é mais poderoso, uma vez que nos diz "quanto" estávamos certo ou errado.
- Adaline tem a vantagem de ser mais preciso.

Para ilustrar essa diferença, veja na figura abaixo. O Perceptron calcula o erro após a função de ativação degrau. Já o Adaline calcula o erro após a função de ativação linear.



**18.** Independente dos valores iniciais assumidos para o vetor de pesos do Adaline, uma mesma configuração final para  $w$  (final) será sempre obtida após a sua convergência. Essa afirmação é falsa ou verdadeira? Justifique.

Verdadeira, o Adaline irá ter uma mesma configuração final após a sua convergência mesmo que os valores iniciais forem aleatórios. O que poderá impactar é o custo para se obter a convergência, ou seja a quantidade de épocas para se obter o resultado convergente

## Parte II - Exercícios de Implementação

**19.** Considere o código do arquivo `exercicio14.py`. Construa um notebook e comente o que cada trecho de código indicado está realizando.

[Verificar notebook em pdf.](#)

**20.** Faça a implementação do algoritmo de treinamento do neurônio perceptron. Realize o treinamento do neurônio para a solução do problema do operador lógico OU partindo da inicialização dos pesos abaixo:

$w_1$ : 0.3092  
 $w_2$ : 0.3092  
 bias: -0.8649

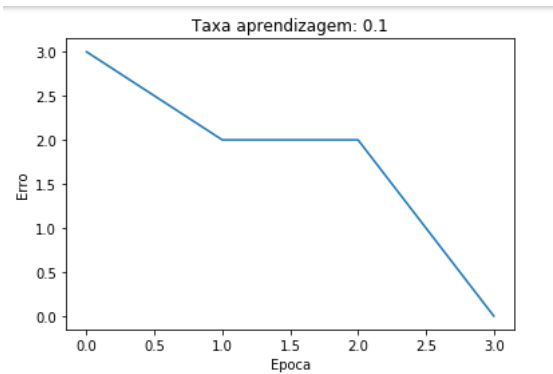
a) Em quantos passos o treinamento convergiu?

Convergiu na época 3, para taxa de aprendizagem = 0,1. Com os seguintes pesos:

w1: 0.7092

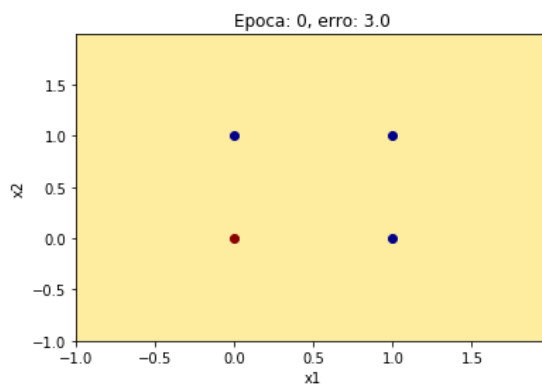
w2: 0.7092

bias: -0.1649

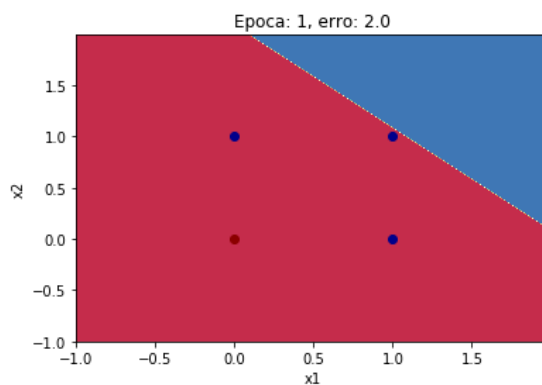


b) Exiba algumas figuras (pelo menos 2) mostrando o hiperplano de separação do problema nas diferentes épocas do treinamento.

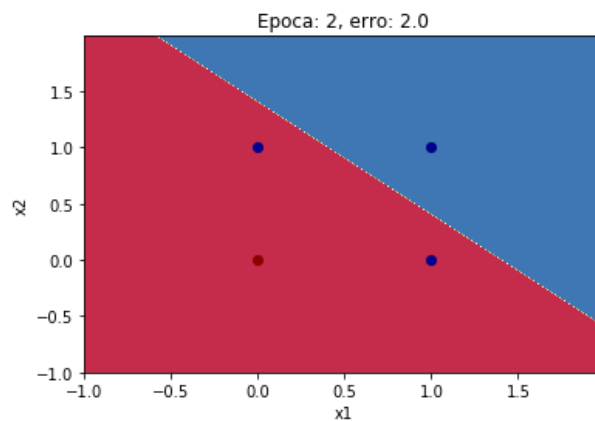
Epoca: 0, erro: 3.0  
w: [ 0.3092 0.3092 -0.8649]



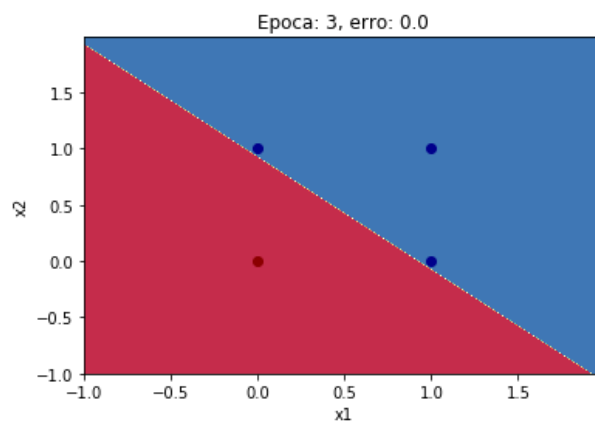
Epoca: 1, erro: 2.0  
w: [ 0.5092 0.5092 -0.5649]



Epoca: 2, erro: 2.0  
w: [ 0.6092 0.6092 -0.3649]



Epoca: 3, erro: 0.0  
w: [ 0.7092 0.7092 -0.1649]



w\_calc: [ 0.7092 0.7092 -0.1649]  
erro na ultima epoca 4 : 0.0

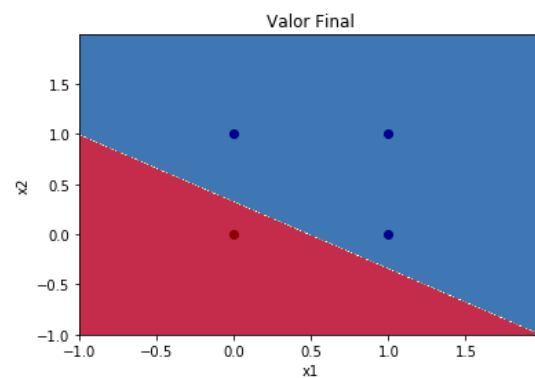
c) Defina outro conjunto de valores de inicialização dos pesos e repita os itens a) e b)

Com os pesos iniciais

w1: 0.0

w2: 0.0

bias: 0.0



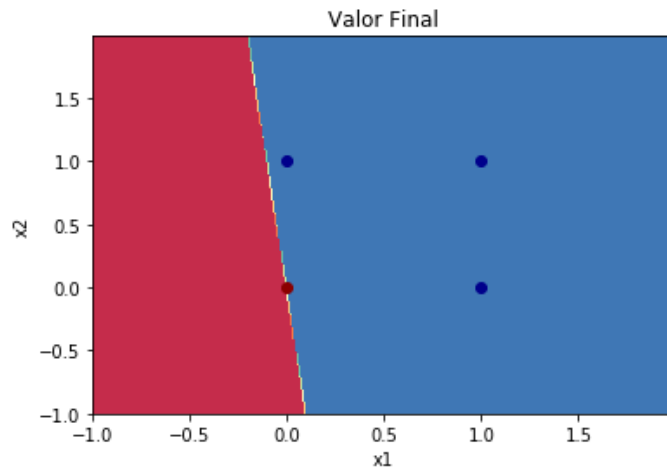
w\_calc: [ 0.2 0.3 0.4]

Com os pesos iniciais

$w_1: 1.0$

$w_2: -1.0$

bias: 2.0



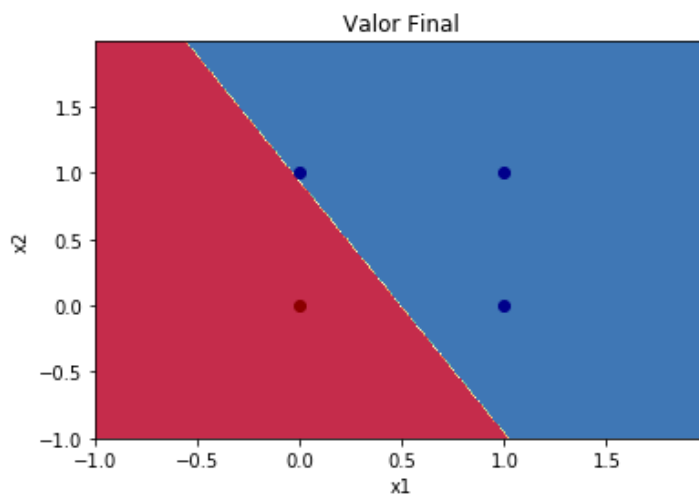
$w_{calc}: [1. \quad 0.1 \quad 0.5]$

Com os pesos iniciais

$w_1: 5.0$

$w_2: -0.1$

bias: -5.25



$w_{calc}: [5.3 \quad 2.8 \quad -2.15]$

Mais detalhes no notebook em pdf.

**21.** Faça a implementação do algoritmo de treinamento do neurônio Adaline. Realize o treinamento do neurônio para a solução do problema do operador lógico OU partindo da inicialização dos pesos abaixo:

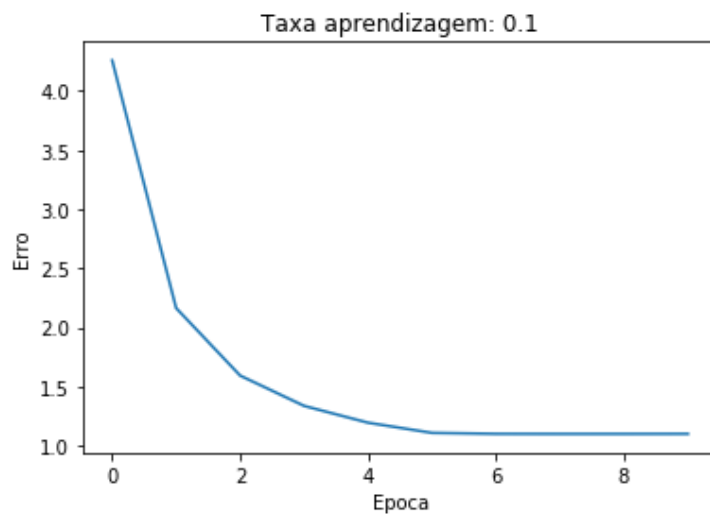
$w_1$ : 0.3192

$w_2$ : 0.3129

bias: -0.8649

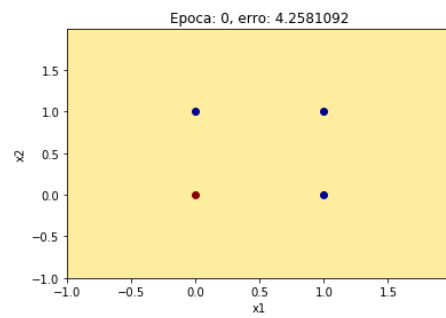
d) Em quantos passos o treinamento convergiu?

Convergiu na época máxima permitida, para taxa de aprendizagem = 0,1. O valor do erro vai reduzindo a quase zero a cada época. Abaixo segue gráfico com o valor do erro a cada época.

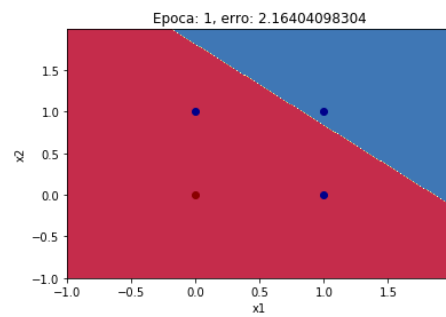


e) Exiba algumas figuras (pelo menos 2) mostrando o hiperplano de separação do problema nas diferentes épocas do treinamento.

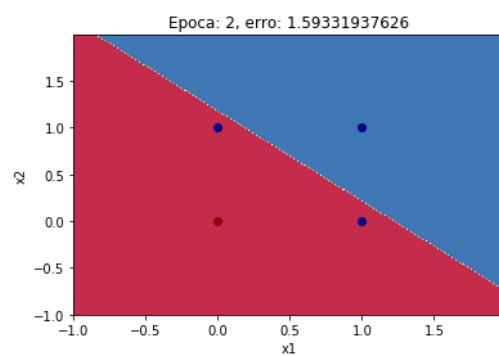
Epoca: 0, erro: 4.2581092  
w: [ 0.3092 0.3092 -0.8649]



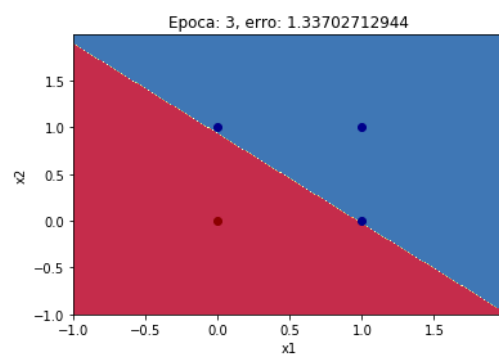
Epoca: 1, erro: 2.16404098304  
w: [ 0.50159992 0.51629202 -0.43908908]



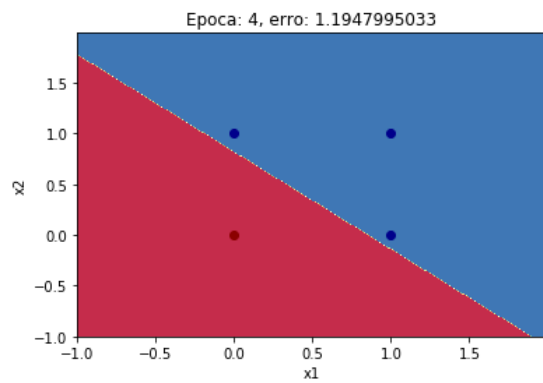
Epoca: 2, erro: 1.59331937626  
w: [ 0.5862063 0.60821807 -0.22268498]



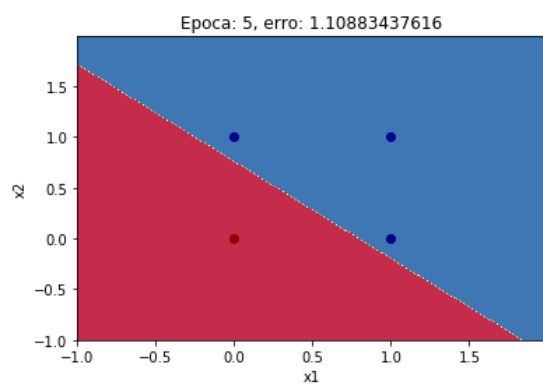
Epoca: 3, erro: 1.33702712944  
w: [ 0.61936077 0.64509335 -0.10804217]



Epoca: 4, erro: 1.1947995033  
w: [ 0.62820928 0.65589004 -0.04317498]



Epoca: 5, erro: 1.10883437616  
w: [ 0.62581386 0.65455623 -0.00295616]



f) Defina outro conjunto de valores de inicialização dos pesos e repita os itens a) e b)

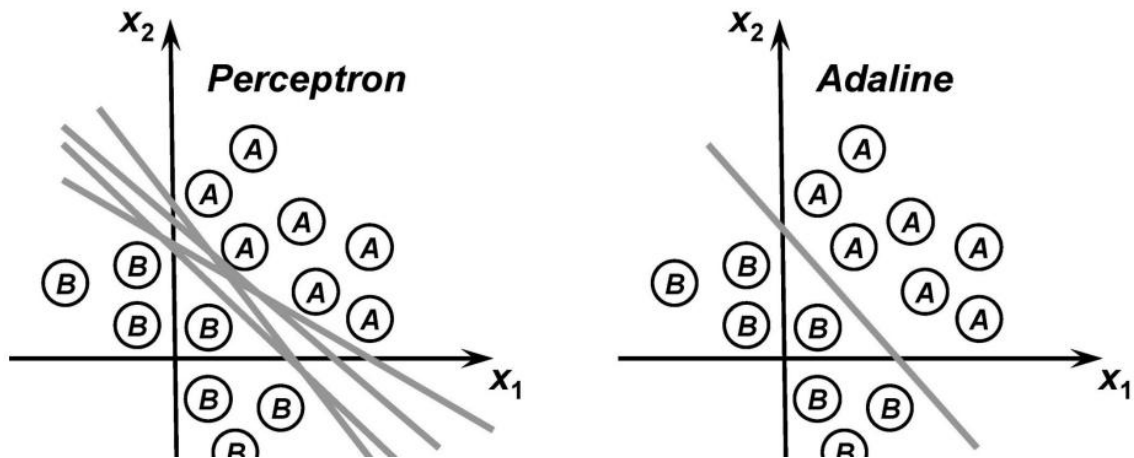
Para diferentes valores de pesos iniciais, os pesos finais após o Adaline foi próximo aos seguintes valores:

W1: 0.44476911  
W2: 0.47255152  
Bias: 0.27733622

Mais detalhes no notebook em pdf.



22. Explique a figura abaixo:



As linhas representadas no gráfico do Perceptron são referentes a diversas soluções possíveis para o problema. Dependendo da inicialização dos parâmetros de peso, se obtém um resultado diferente, porém plausível para o problema.

Já o Adaline se obtém usa solução única que independe dos valores da inicialização dos parâmetros de peso.

Em ambos cenários estamos considerando que foram executadas épocas suficientes para obter resultado satisfatório de cada método.