PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA

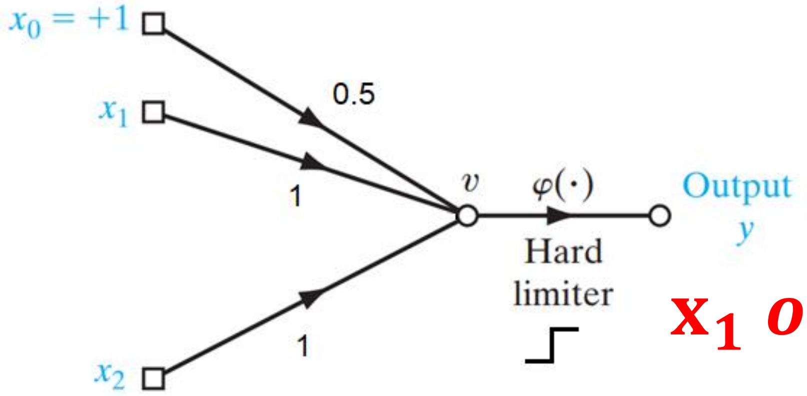
Machine Learning Theory
and Applications Lab

Aprendizado de Máquina

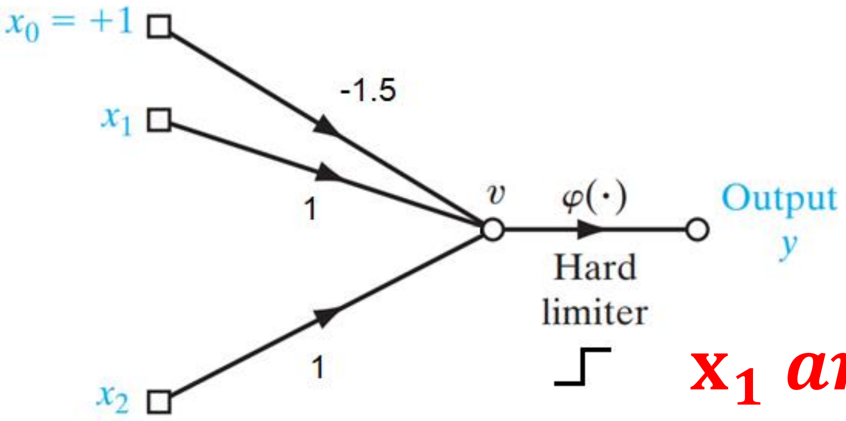
Paradigma baseado em Otimização
Redes Neurais II: Backpropagation

Prof. Me. Otávio Parraga

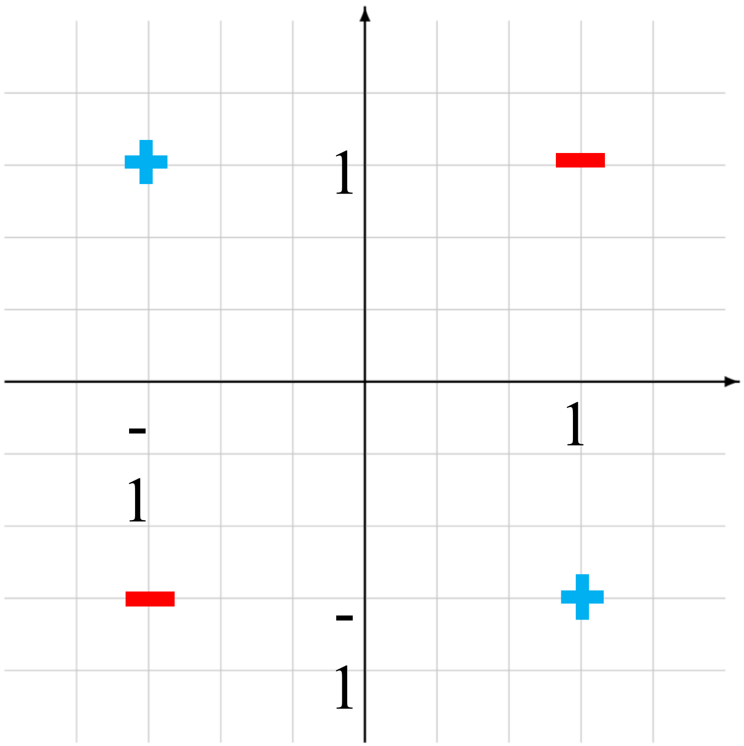
Aula Passada



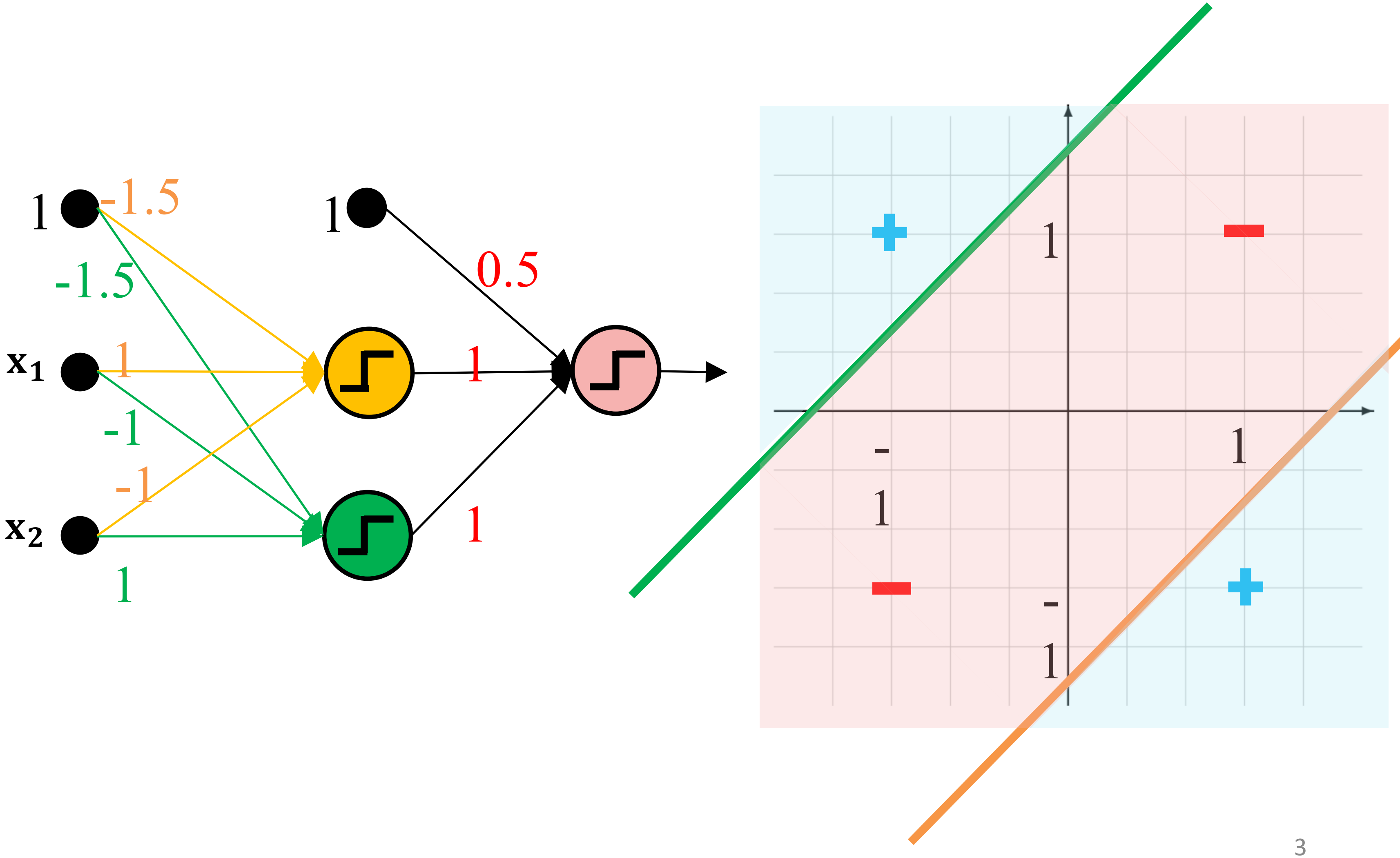
$\mathbf{x_1}$ ***or*** $\mathbf{x_2}$



$\mathbf{x_1}$ ***and*** $\mathbf{x_2}$

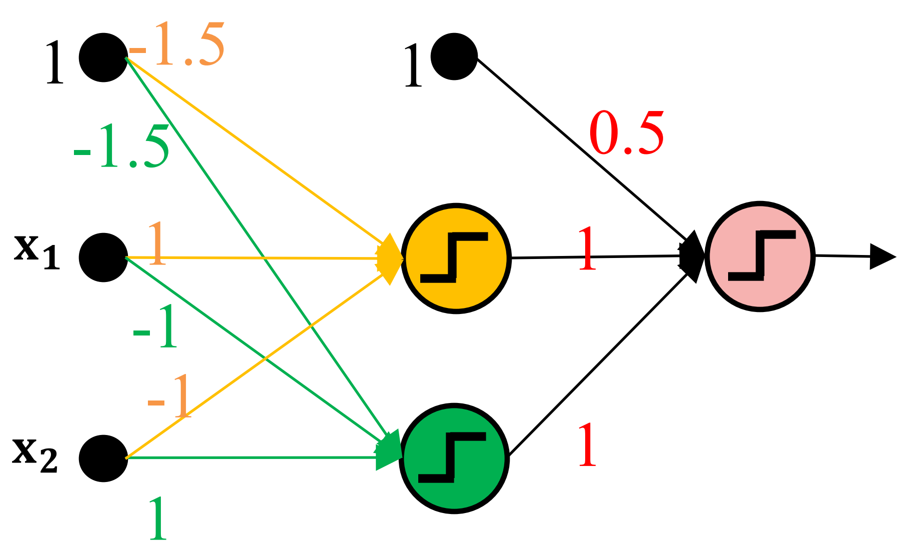


Aula Passada



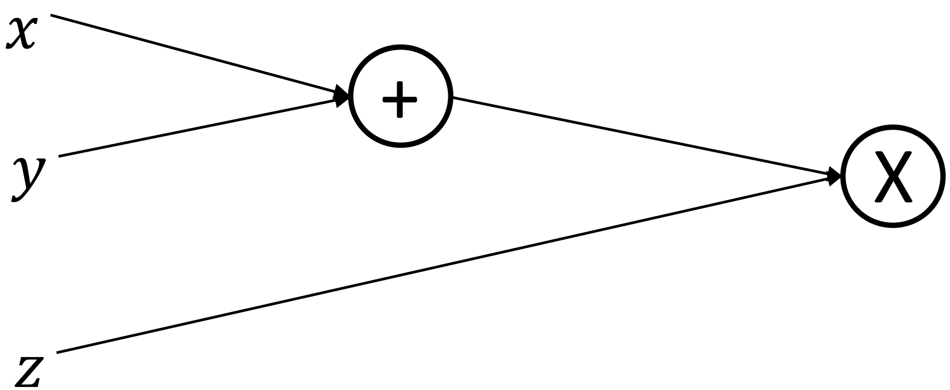
Perceptron Multi-Camada

- O problema é: como treinamos esse modelo?
  - Gradiente Descendente Estocástico
    - Stochastic Gradient Descent (SGD)



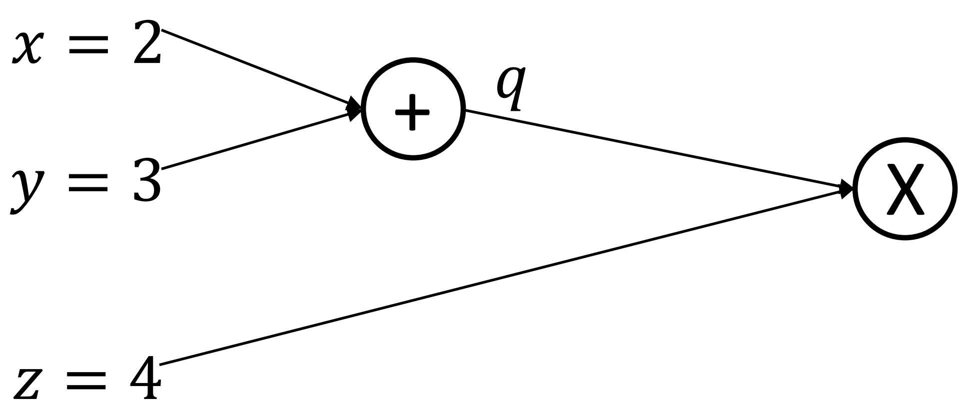
Regra da Cadeia

$$f(x, y, z) = (x + y) * z$$



Regra da Cadeia

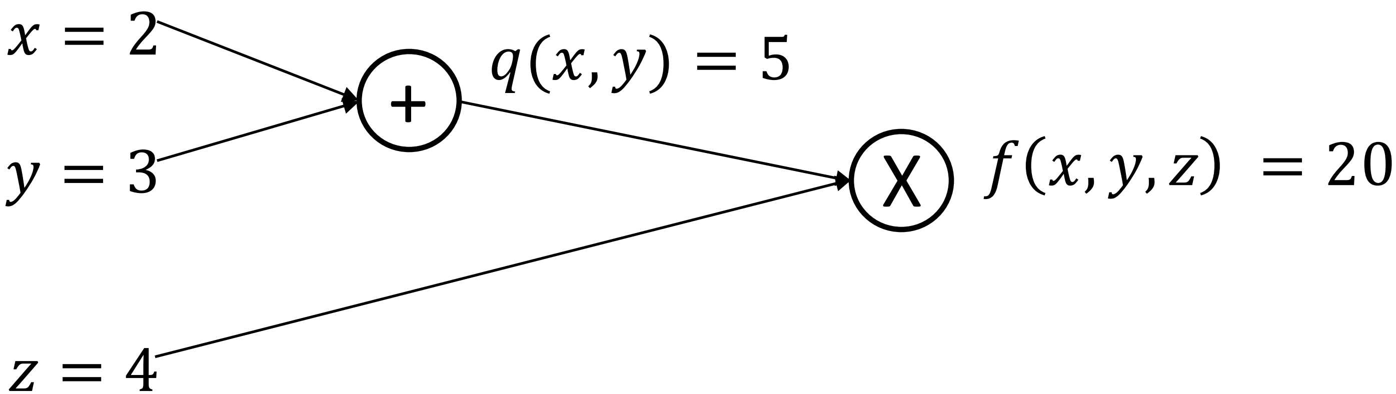
$$f(x, y, z) = (x + y) * z$$
$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$



Vamos dizer que $f(x, y, z)$
é uma função composta!

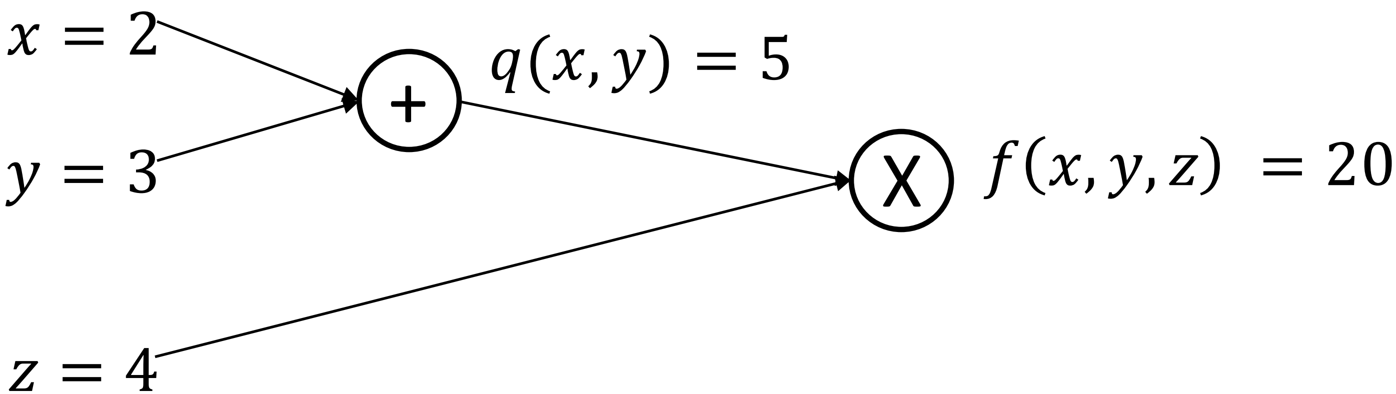
Regra da Cadeia

$$f(x, y, z) = (x + y) * z$$
$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$



Regra da Cadeia

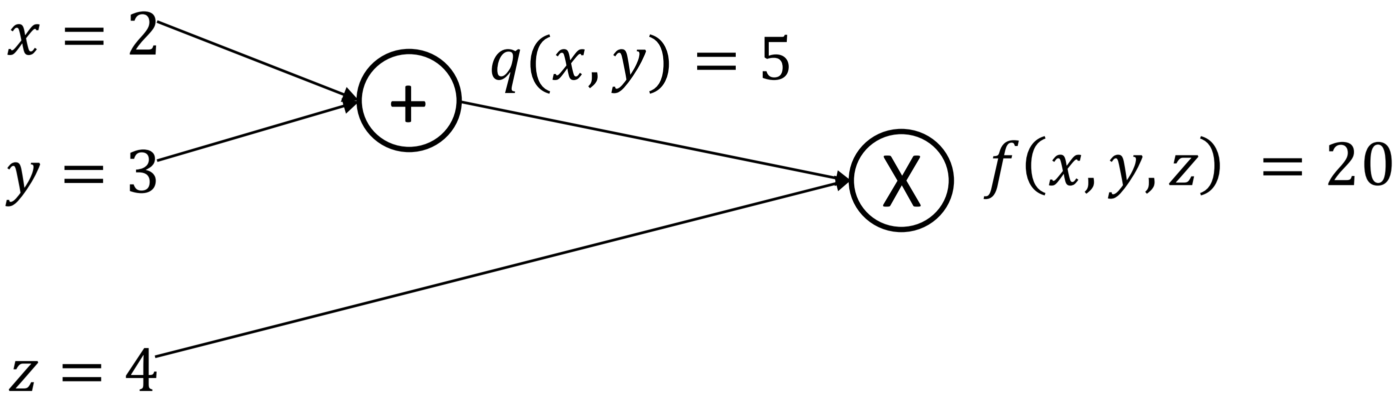
$$f(x, y, z) = (x + y) * z$$
$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$



Nosso objetivo é calcular as **derivadas parciais** de f com relação a x, y e z

Regra da Cadeia

$$f(x, y, z) = (x + y) * z$$
$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$



$$\frac{\partial f}{\partial x} =? \qquad \frac{\partial f}{\partial z} =?$$

$$\frac{\partial f}{\partial y} =? \qquad \frac{\partial f}{\partial q} =?$$

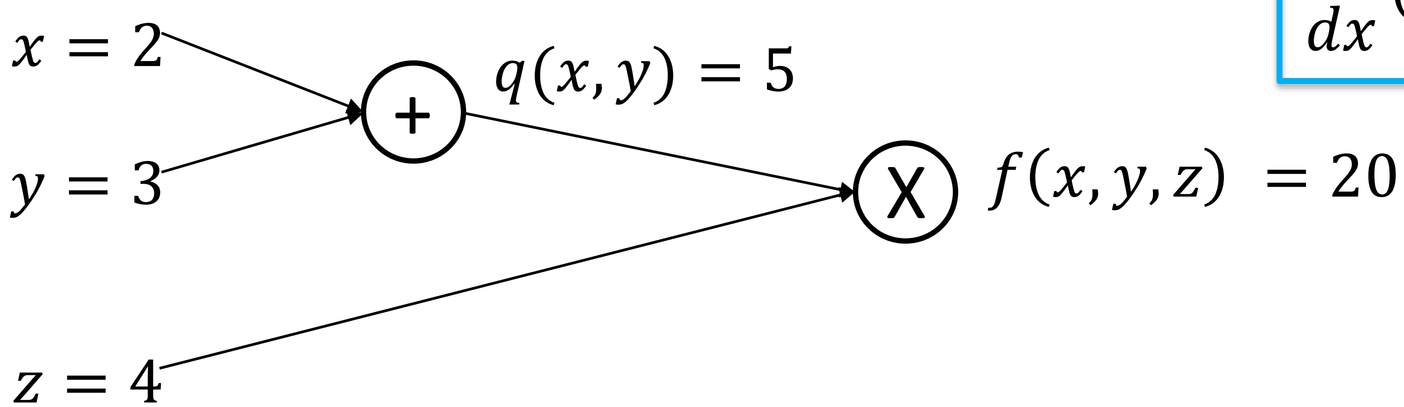
Nosso objetivo é calcular as **derivadas parciais** de f com relação a x, y e z

Regra da Cadeia

$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$

Colinha

$$\frac{d}{dx}(2ax) = 2a$$
$$\frac{d}{dx}(x + a) = 1$$



$$\frac{\partial f}{\partial x} = ? \qquad \frac{\partial f}{\partial z} = ?$$

$$\frac{\partial f}{\partial y} = ? \qquad \frac{\partial f}{\partial q} = ?$$

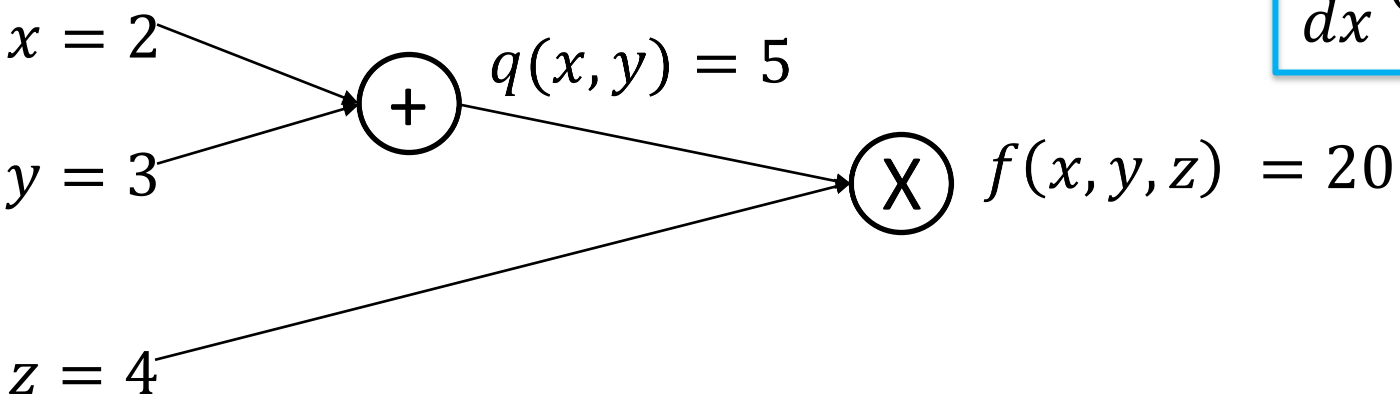
Nosso objetivo é calcular as **derivadas parciais** de f com relação a x, y e z

Regra da Cadeia

$$f(x,y,z) = q(x,y) * z$$
$$q(x,y) = x + y$$

Colinha

$$\frac{d}{dx}(2ax) = 2a$$
$$\frac{d}{dx}(x + a) = 1$$



$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial x} \qquad \frac{\partial f}{\partial z} = q$$

Nosso objetivo é calcular as **derivadas parciais** de f com relação a x, y e z

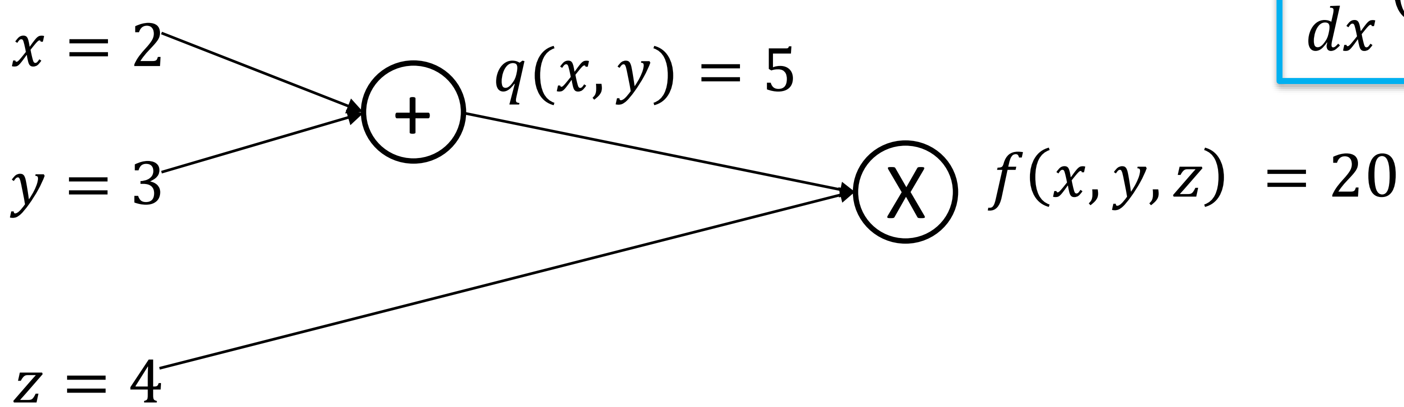
$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial y} \qquad \frac{\partial f}{\partial q} = z$$

Regra da Cadeia

$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$

Colinha

$$\frac{d}{dx}(2ax) = 2a$$
$$\frac{d}{dx}(x + a) = 1$$



$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial x} \qquad \frac{\partial f}{\partial z} = 5 \qquad \frac{\partial q}{\partial x} = 1$$

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial y} \qquad \frac{\partial f}{\partial q} = 4 \qquad \frac{\partial q}{\partial y} = 1$$

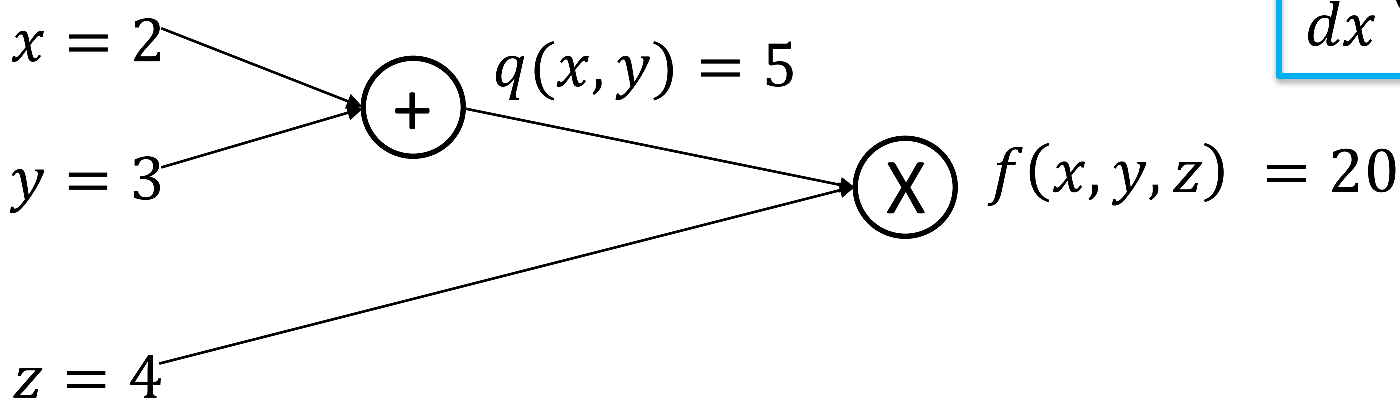
Nosso objetivo é calcular as **derivadas parciais** de f com relação a x, y e z

Regra da Cadeia

$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$

Colinha

$$\frac{d}{dx}(2ax) = 2a$$
$$\frac{d}{dx}(x + a) = 1$$



$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} 1 \qquad \frac{\partial f}{\partial z} = 5 \qquad \frac{\partial q}{\partial x} = 1$$

Nosso objetivo é calcular as **derivadas parciais** de f com relação a x, y e z

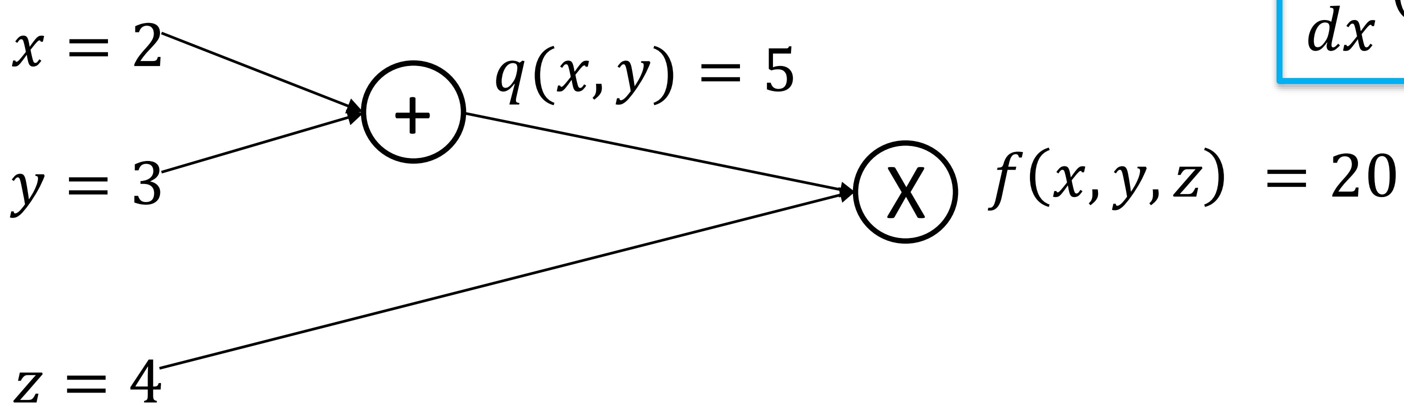
$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} 1 \qquad \frac{\partial f}{\partial q} = 4 \qquad \frac{\partial q}{\partial y} = 1$$

Regra da Cadeia

$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$

Colinha

$$\frac{d}{dx}(2ax) = 2a$$
$$\frac{d}{dx}(x + a) = 1$$



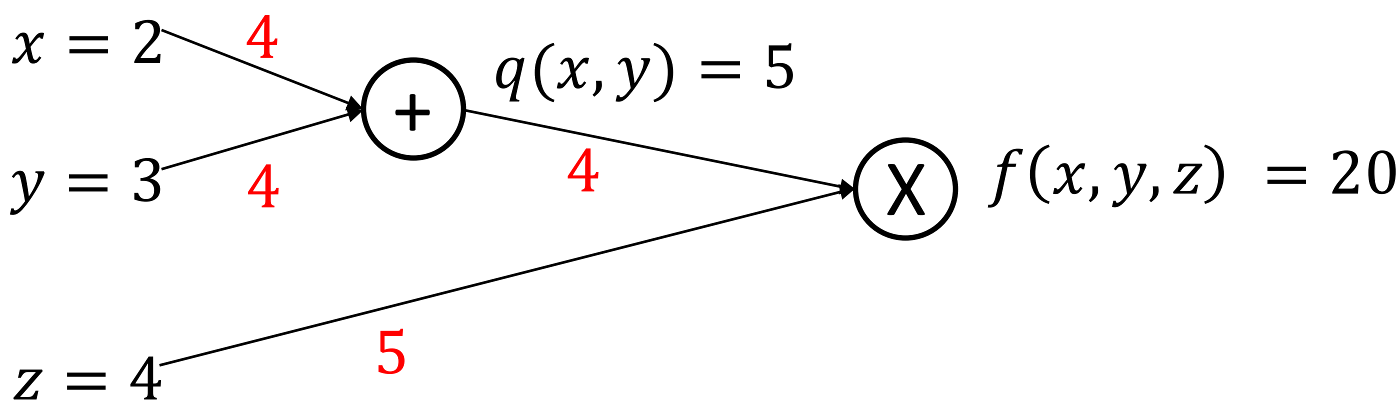
$$\frac{\partial f}{\partial x} = 4 * 1 \qquad \frac{\partial f}{\partial z} = 5 \qquad \frac{\partial q}{\partial x} = 1$$

Nosso objetivo é calcular as **derivadas parciais** de f com relação a x, y e z

$$\frac{\partial f}{\partial y} = 4 * 1 \qquad \frac{\partial f}{\partial q} = 4 \qquad \frac{\partial q}{\partial y} = 1$$

Regra da Cadeia

$$f(x, y, z) = q(x, y) * z$$
$$q(x, y) = x + y$$



$$\frac{\partial f}{\partial x} = 4 \qquad \frac{\partial f}{\partial z} = 5$$

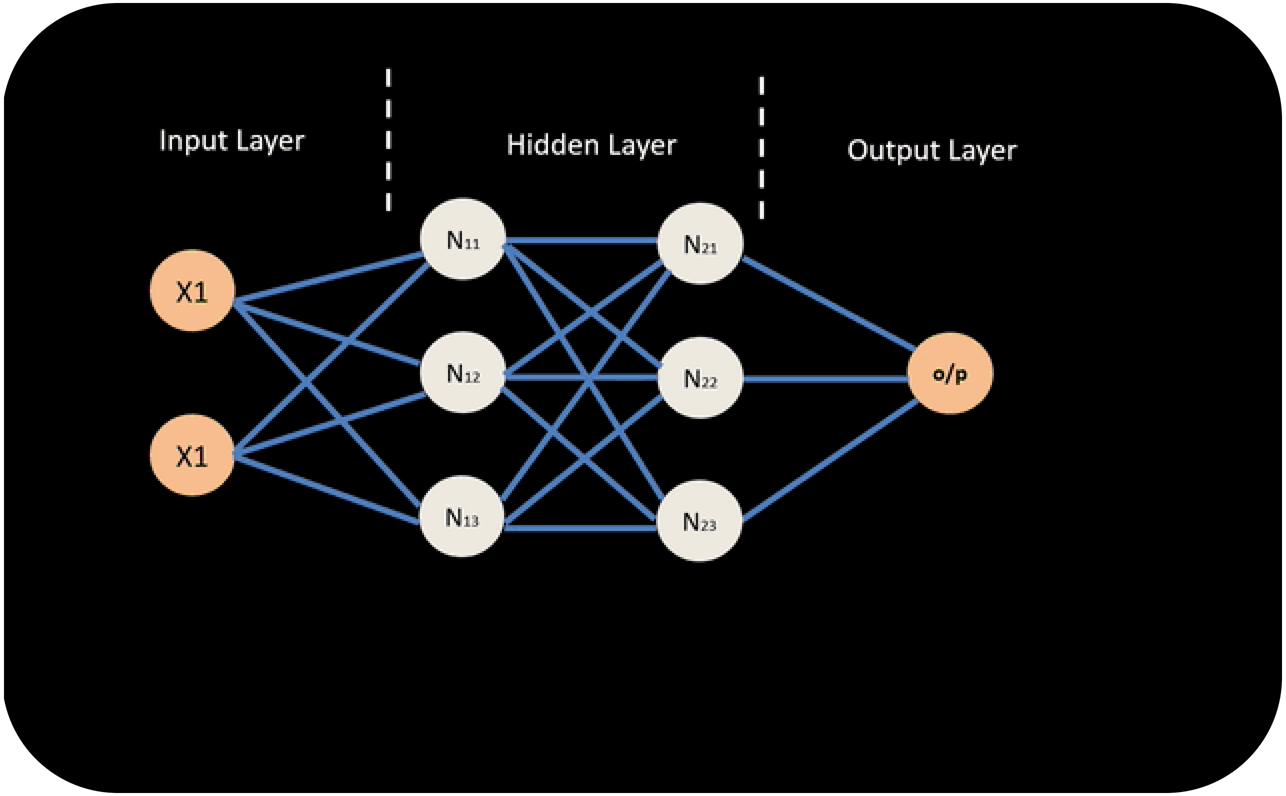
Descobrimos as **derivadas parciais** de f com relação a x, y e z !

$$\frac{\partial f}{\partial y} = 4$$

Etapas no Treinamento

- O treinamento de uma rede neural é dividido em três partes:
  - Forward
    - Realiza a predição
  - Cálculo da Loss
    - Avalia o resultado obtido
  - Backward
    - Ajusta os pesos da rede

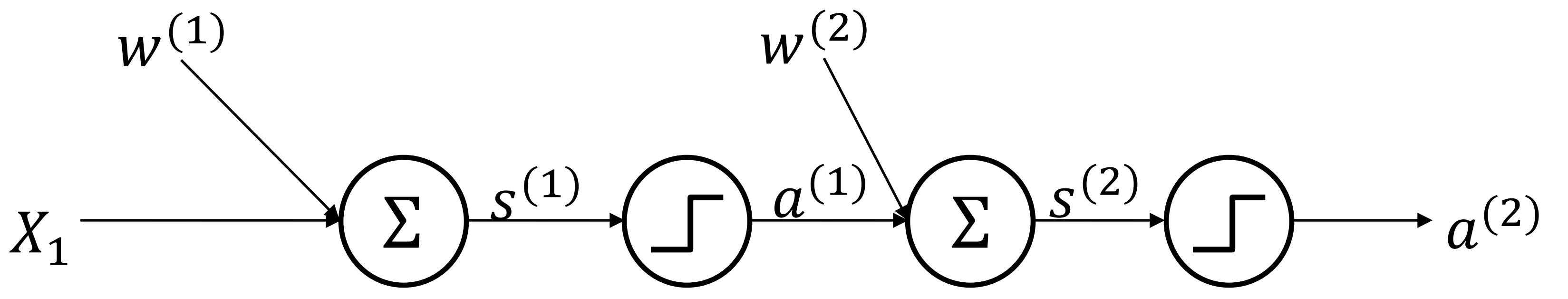
Etapas no Treinamento



Conceitos Importantes

- Backpropagation
  – Calcular os gradientes usando a regra da cadeia
- Época
  – Todos os dados de treinamento passaram pela rede, tanto forward quanto backward
- Batch
  – Uma pequena porção dos dados
- Total de Iterações
  – $\frac{N\ inst\hat{a}ncias}{Tam.Batch}$

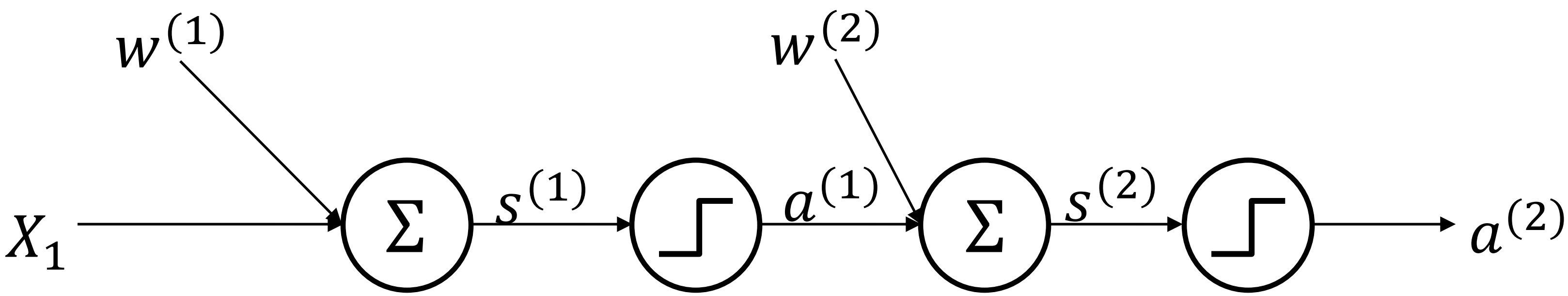
Treinando uma MLP Simples



Ideia: Treinar rede neural usando SGD

- 2 camadas
- 1 neurônio em cada camada
- Vamos tentar aprender a função identidade $o^{(2)} = X_1$

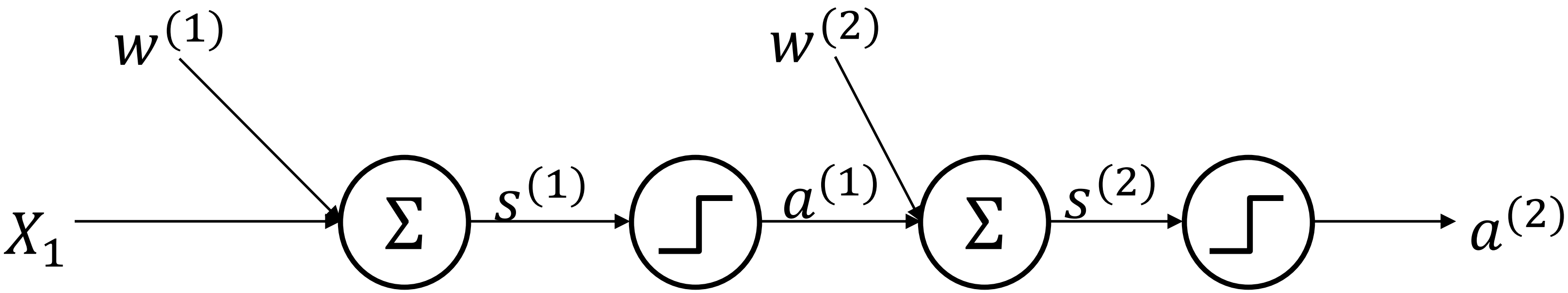
Treinando uma MLP Simples



Problema: Função Sinal não é diferenciável em 0 (e a derivada é 0 onde ela é diferenciável)

- 2 camadas
- 1 neurônio em cada camada
- Vamos tentar aprender a função identidade $o^{(2)} = X_1$

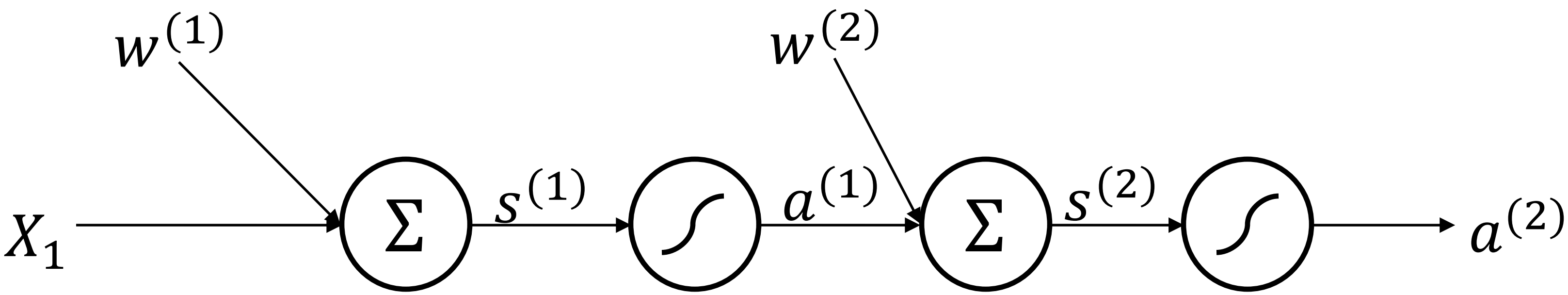
Treinando uma MLP Simples



Problema: Função Sinal não é diferenciável em 0 (e a derivada é 0 onde ela é diferenciável)
Solução: Trocar por outra função de ativação não-linear

- 2 camadas
- 1 neurônio em cada camada
- Vamos tentar aprender a função identidade $o^{(2)} = X_1$

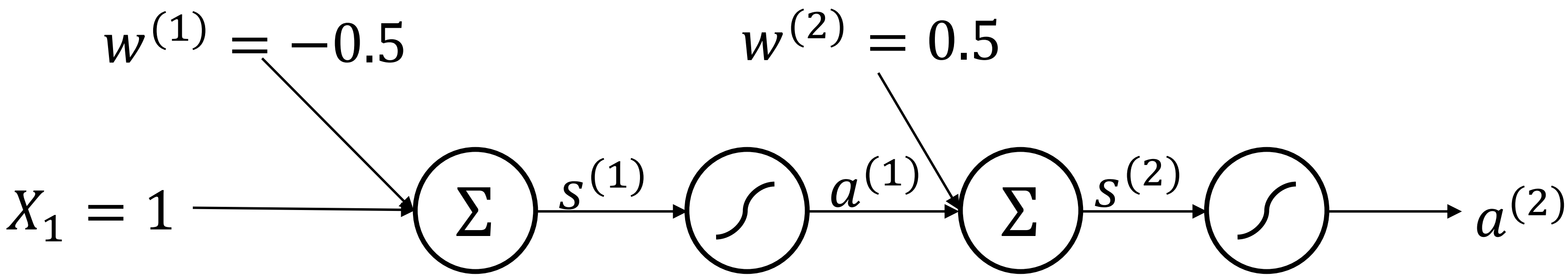
Treinando uma MLP Simples



Problema: Função Sinal não é diferenciável em 0 (e a derivada é 0 onde ela é diferenciável)
Solução: Trocar por outra função de ativação não-linear -> Sigmoide!

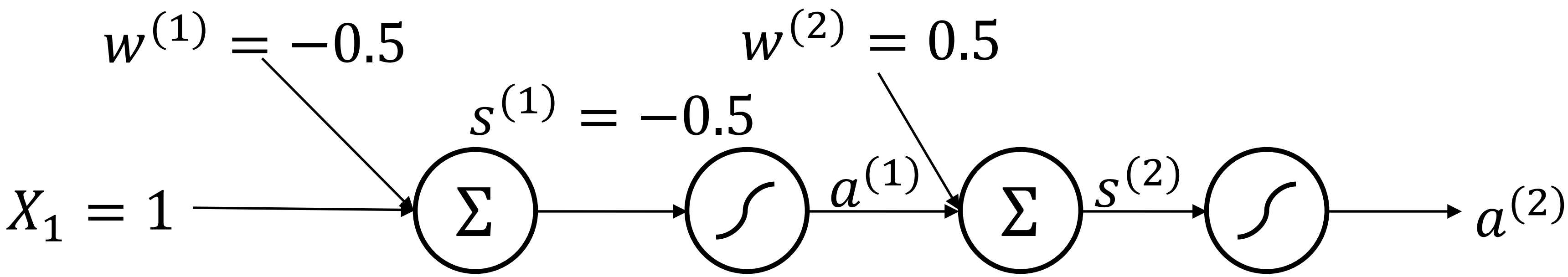
- 2 camadas
- 1 neurônio em cada camada
- Vamos tentar aprender a função identidade $o^{(2)} = X_1$

Inicializando os Valores



- $w^{(1)} = -0.5$
- $w^{(2)} = 0.5$
- $X_1 = 1$
- $y \quad = 1$
- $MSELoss = \frac{1}{2}(\hat{y} - y)^2$

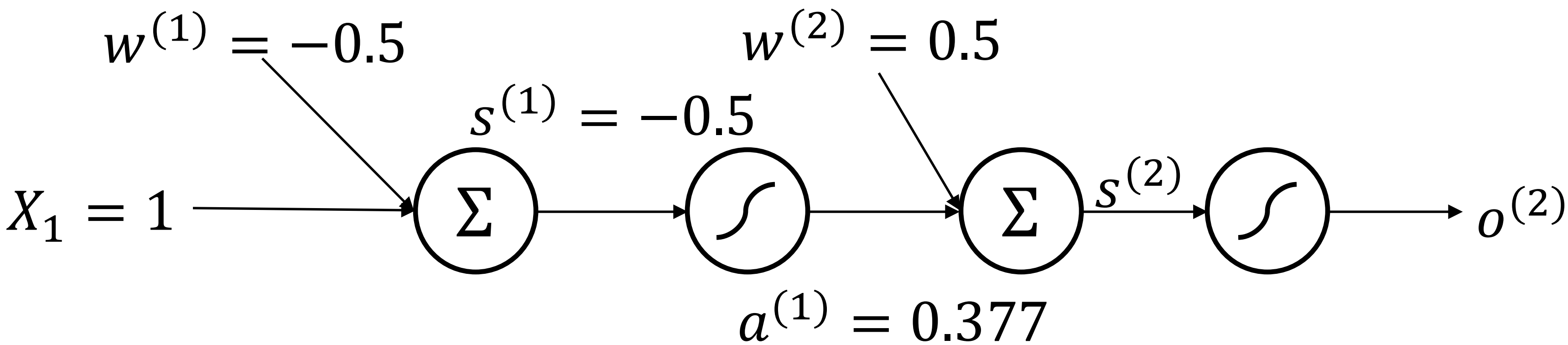
Forward Pass



$$s^{(1)} = 1 * -0.5$$

- $w^{(1)} = -0.5$
- $w^{(2)} = 0.5$
- $X_1 = 1$
- $y \quad = 1$
- $MSELoss = \frac{1}{2}(\hat{y} - y)^2$

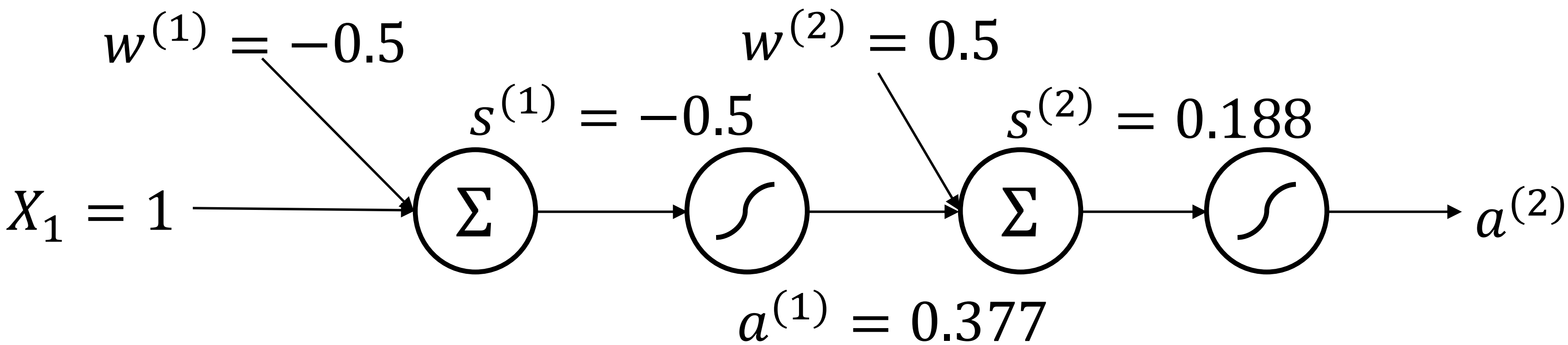
Forward Pass



$$a^{(1)} = sigmoid(-0.5)$$

- $w^{(1)} = -0.5$
- $w^{(2)} = 0.5$
- $X_1 = 1$
- $y \quad = 1$
- $MSELoss = \frac{1}{2}(\hat{y} - y)^2$

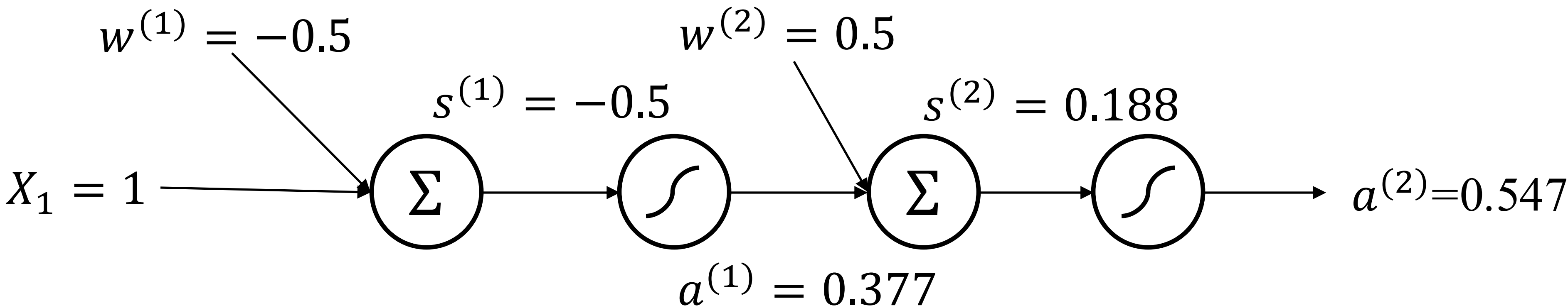
Forward Pass



$$s^{(2)} = 0.377 * 0.5$$

- $w^{(1)} = -0.5$
- $w^{(2)} = 0.5$
- $X_1 = 1$
- $y \quad = 1$
- $MSELoss = \frac{1}{2}(\hat{y} - y)^2$

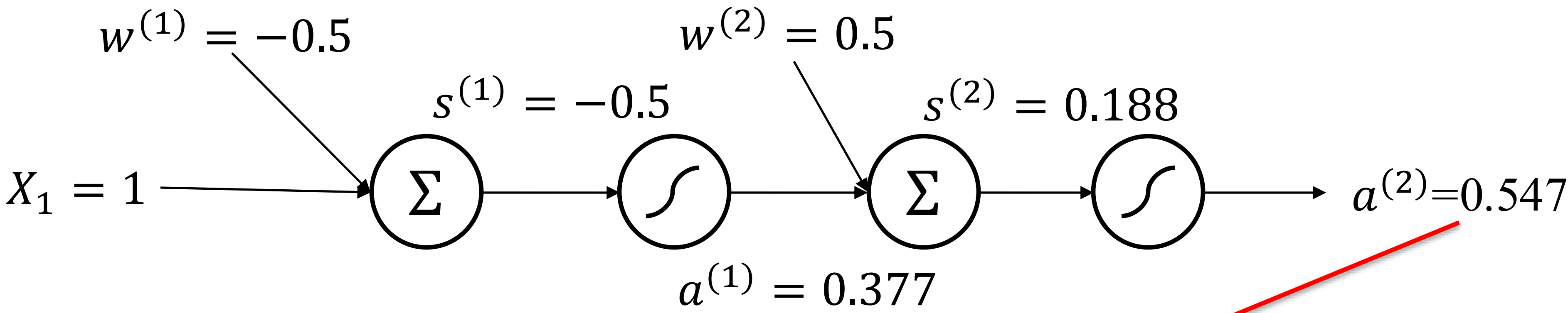
Forward Pass



$$a^{(2)} = sigmoid(0.188)$$

- $w^{(1)} = -0.5$
- $w^{(2)} = 0.5$
- $X_1 = 1$
- $y \quad = 1$
- $MSELoss = \frac{1}{2}(\hat{y} - y)^2$

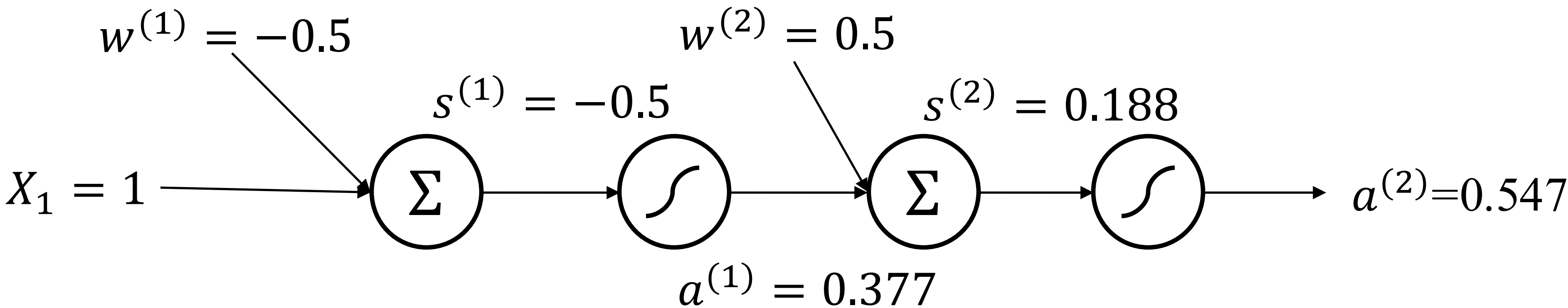
Loss



$$Loss = \frac{1}{2}(0.547 - 1)^2$$

- $w^{(1)} = -0.5$
- $w^{(2)} = 0.5$
- $X_1 = 1$
- $y \quad = 1$
- $MSELoss = \frac{1}{2}(\hat{y} - y)^2$

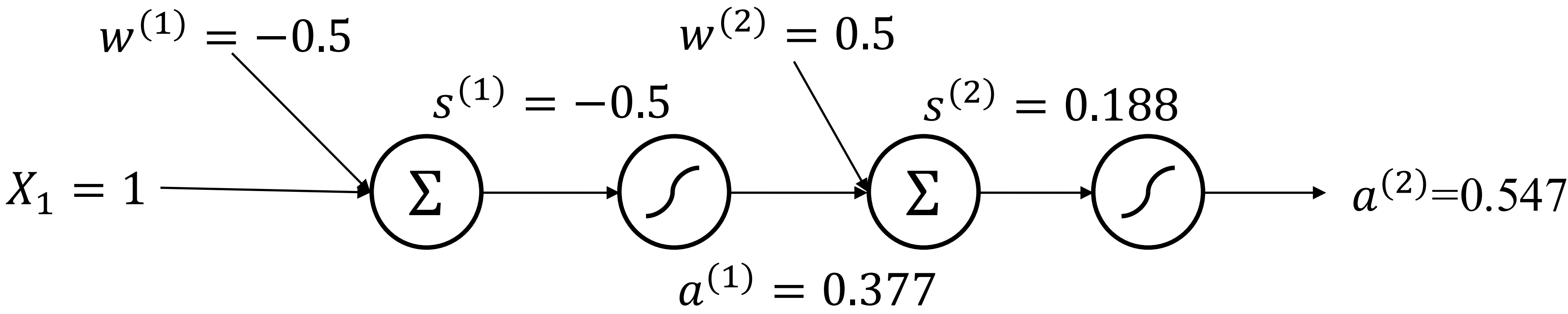
Loss



$$Loss = \frac{1}{2}(0.547 - 1)^2 = 0.102$$

- $w^{(1)} = -0.5$
- $w^{(2)} = 0.5$
- $X_1 = 1$
- $y \quad = 1$
- $MSELoss = \frac{1}{2}(\hat{y} - y)^2$

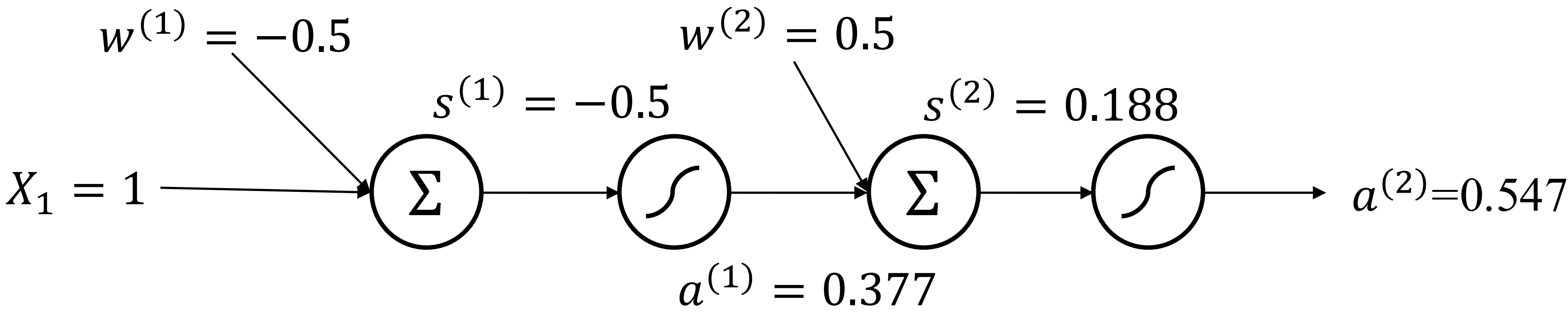
Backward



$$\textbf{SGD: } \underbrace{w_t}_{\text{Novos pesos}} = \underbrace{w_{t-1}}_{\text{Pesos anteriores}} - \underbrace{\alpha \nabla_w Loss}_{\text{Gradientes da Loss com relação a w}}$$

$$Loss = \frac{1}{2}(0.547 - 1)^2 = 0.102$$

Backward



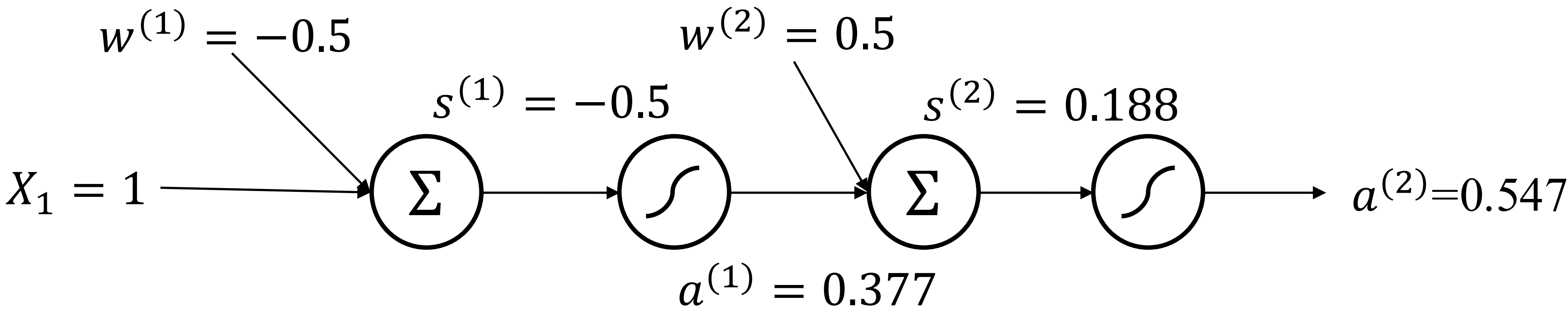
$$\hat{y} = sig(sig\left(X_1 * w^{(1)}\right) * w^{(2)})$$

Problema 2: Como computar $\frac{\partial Loss}{\partial w^{(2)}}$ e $\frac{\partial Loss}{\partial w^{(1)}}$ se o erro é uma composição de funções?

$$Loss = \frac{1}{2}(0.547 - 1)^2 = 0.102$$

$$w_t = w_{t-1} - \alpha \nabla_w Loss$$

Backward



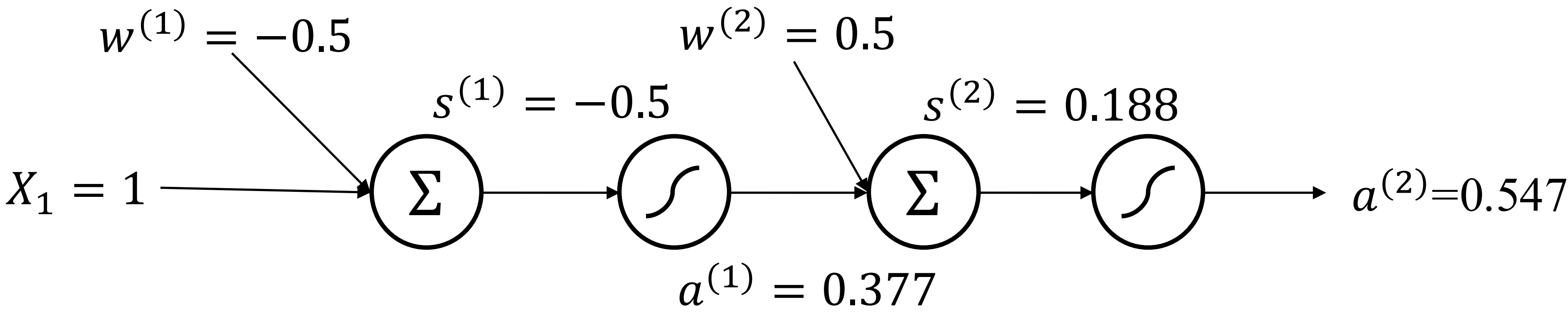
REGRA DA CADEIA

$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}} \qquad \frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

$$Loss = \frac{1}{2}(0.547 - 1)^2 = 0.102$$

$$w_t = w_{t-1} - \alpha \nabla_w Loss$$

Backward

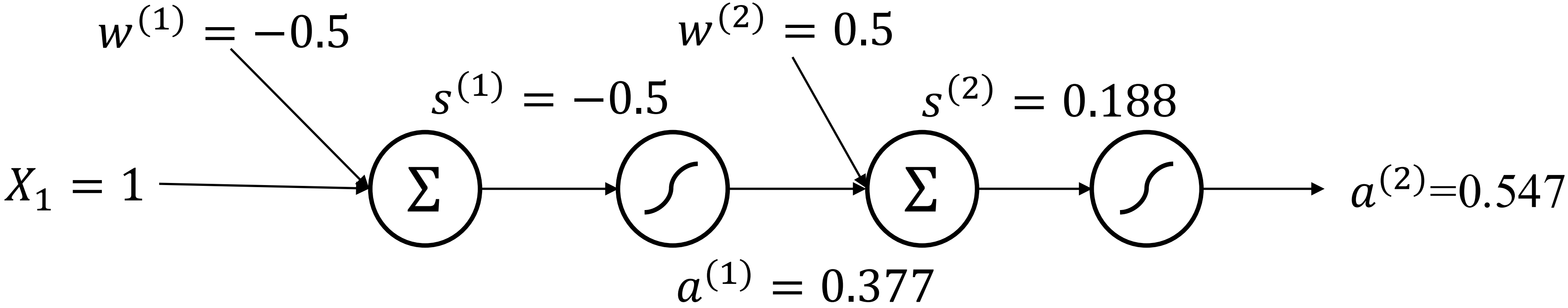


$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

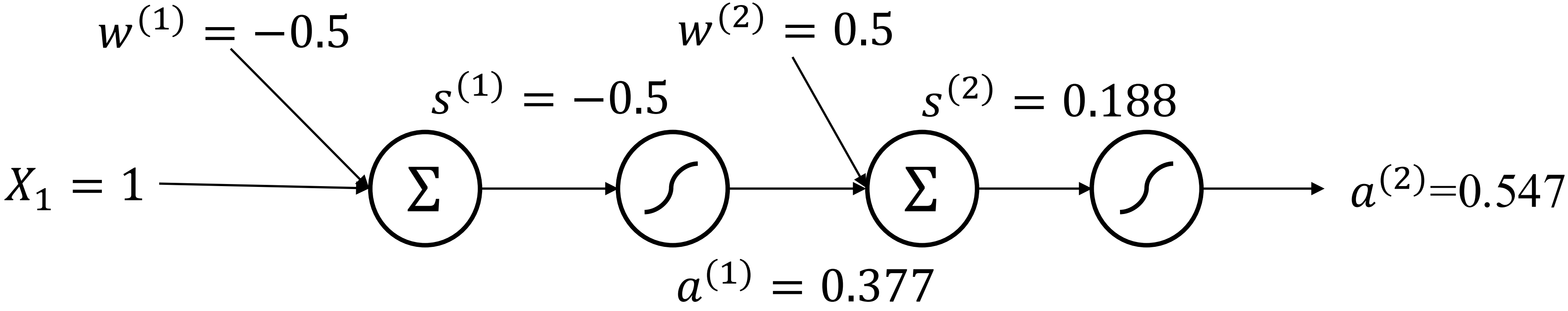
$$Loss = \frac{1}{2}(0.547 - 1)^2 = 0.102$$

$$w_t = w_{t-1} - \alpha \nabla_w Loss$$



$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

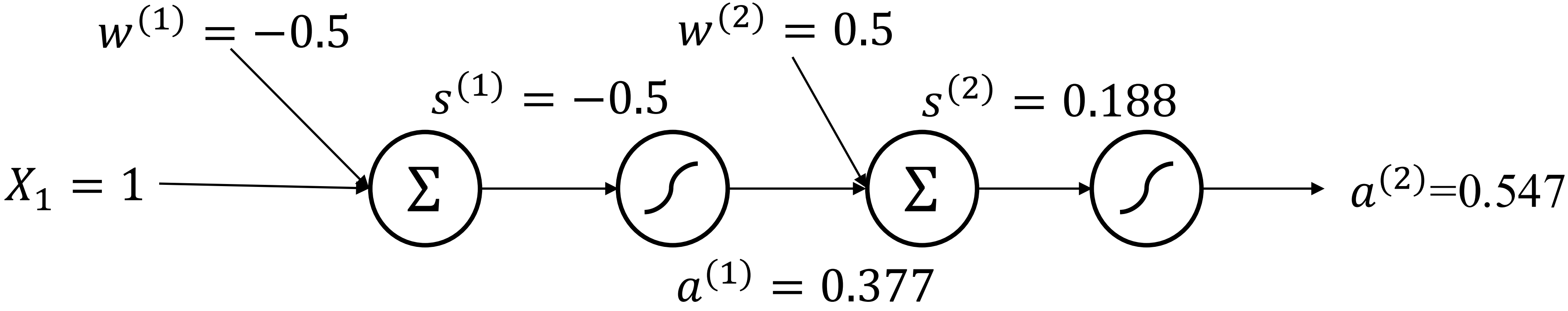


$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

Calculando as Derivadas

$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$



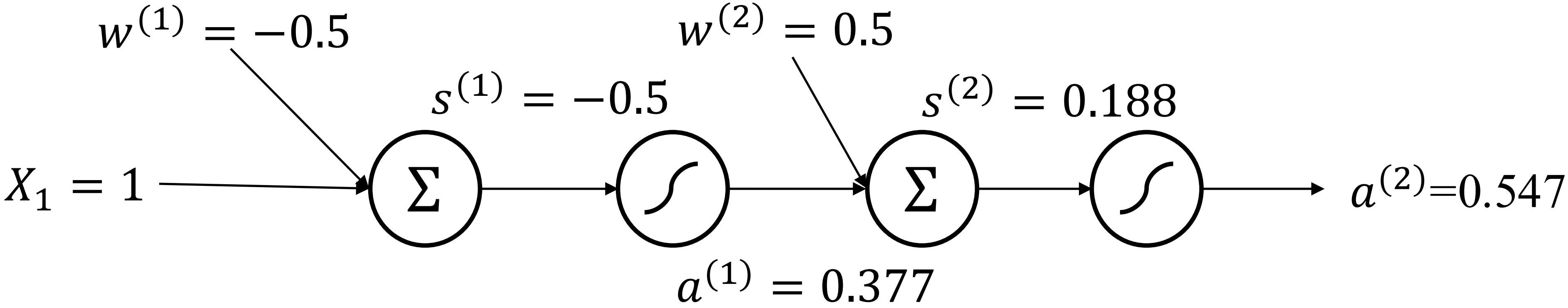
$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

Calculando as Derivadas

$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$

$$\frac{\partial a^{(2)}}{\partial s^{(2)}} = a^{(2)}(1 - a^{(2)})$$



$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

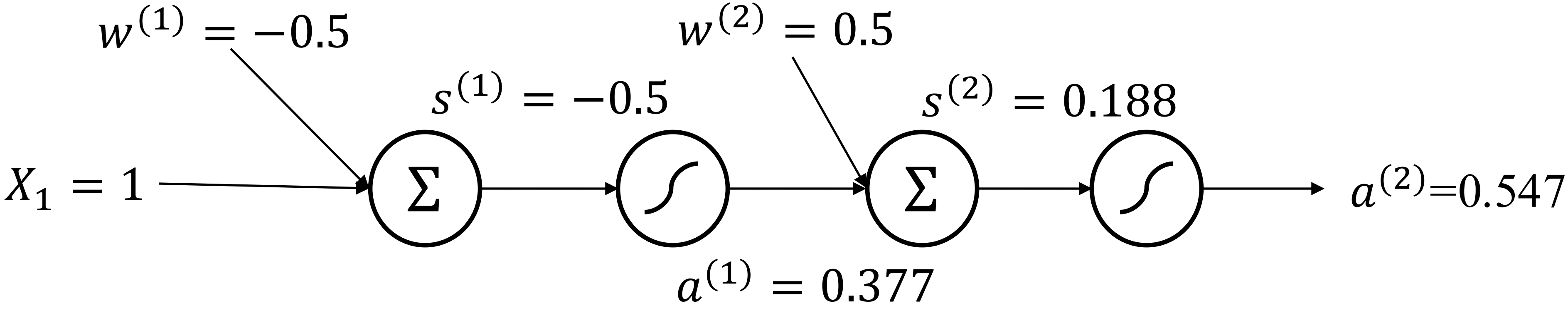
$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

Calculando as Derivadas

$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$

$$\frac{\partial a^{(2)}}{\partial s^{(2)}} = a^{(2)}(1 - a^{(2)})$$

$$\frac{\partial s^{(2)}}{\partial w^{(2)}} = a^{(1)}$$



$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial o^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

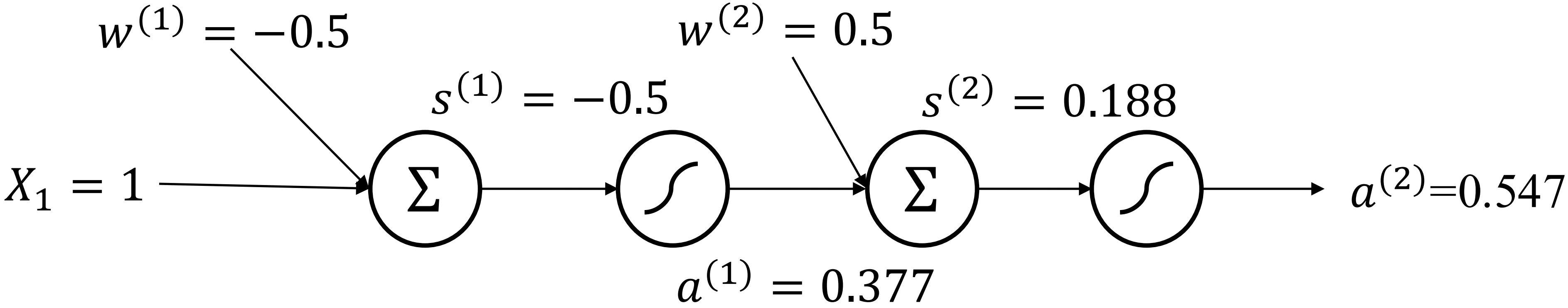
Calculando as Derivadas

$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$

$$\frac{\partial s^{(2)}}{\partial a^{(1)}} = w^{(2)}$$

$$\frac{\partial a^{(2)}}{\partial s^{(2)}} = a^{(2)}(1 - a^{(2)})$$

$$\frac{\partial s^{(2)}}{\partial w^{(2)}} = a^{(1)}$$



$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial o^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

Calculando as Derivadas

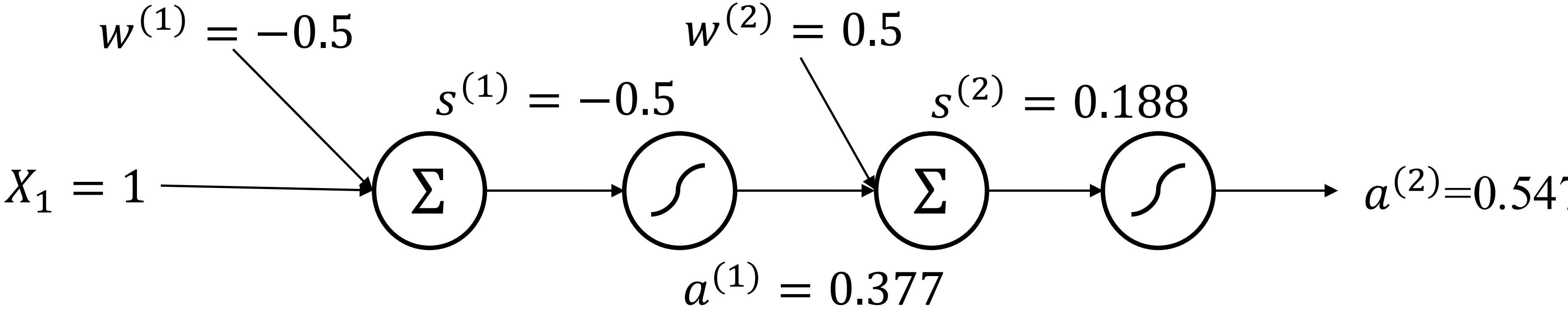
$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$

$$\frac{\partial s^{(2)}}{\partial a^{(1)}} = w^{(2)}$$

$$\frac{\partial a^{(2)}}{\partial s^{(2)}} = o^{(2)}(1 - a^{(2)})$$

$$\frac{\partial a^{(1)}}{\partial s^{(1)}} = a^{(1)}(1 - a^{(1)})$$

$$\frac{\partial s^{(2)}}{\partial w^{(2)}} = a^{(1)}$$



$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial o^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

Calculando as Derivadas

$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$

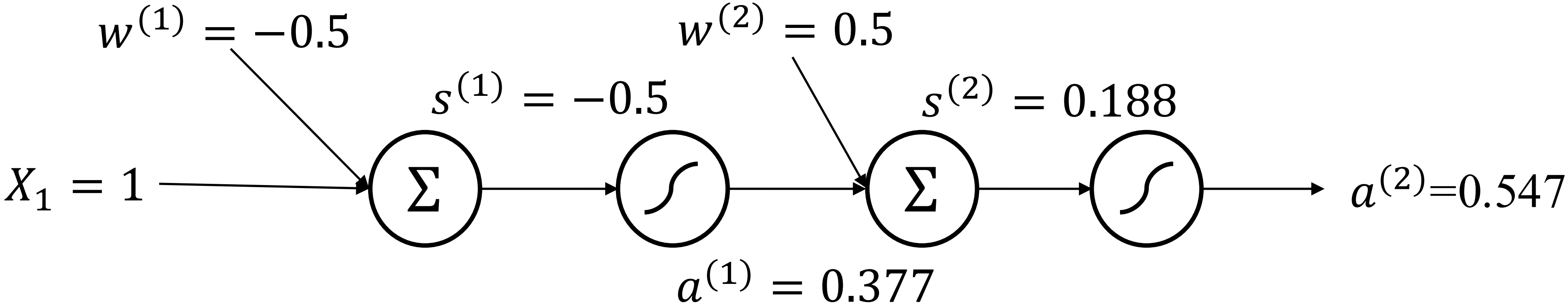
$$\frac{\partial s^{(2)}}{\partial a^{(1)}} = w^{(2)}$$

$$\frac{\partial a^{(2)}}{\partial s^{(2)}} = a^{(2)}(1 - a^{(2)})$$

$$\frac{\partial a^{(1)}}{\partial s^{(1)}} = a^{(1)}(1 - a^{(1)})$$

$$\frac{\partial s^{(2)}}{\partial w^{(2)}} = a^{(1)}$$

$$\frac{\partial s^{(1)}}{\partial w^{(1)}} = X_1$$



$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial o^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}}$$

Calculando as Derivadas

$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$

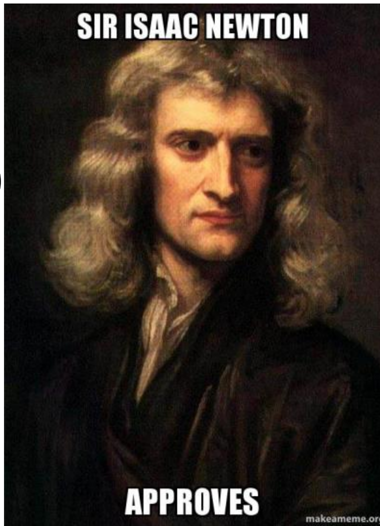
$$\frac{\partial a^{(2)}}{\partial s^{(2)}} = a^{(2)}(1 - a^{(2)})$$

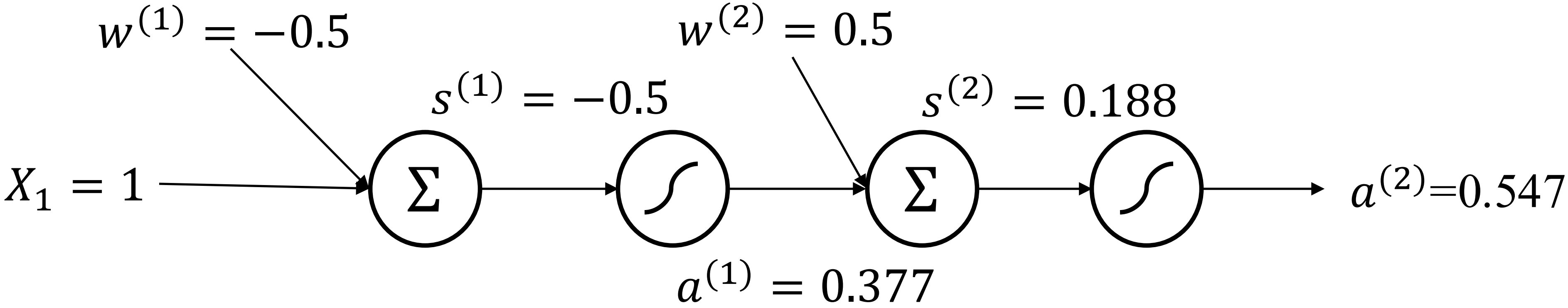
$$\frac{\partial s^{(2)}}{\partial w^{(2)}} = a^{(1)}$$

$$\frac{\partial s^{(2)}}{\partial a^{(1)}} = w^{(2)}$$

$$\frac{\partial a^{(1)}}{\partial s^{(1)}} = a^{(1)}(1 - a^{(1)})$$

$$\frac{\partial s^{(1)}}{\partial w^{(1)}} = X_1$$





$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$

$$\frac{\partial s^{(2)}}{\partial a^{(1)}} = w^{(2)}$$

$$\frac{\partial a^{(2)}}{\partial s^{(2)}} = a^{(2)}(1 - a^{(2)})$$

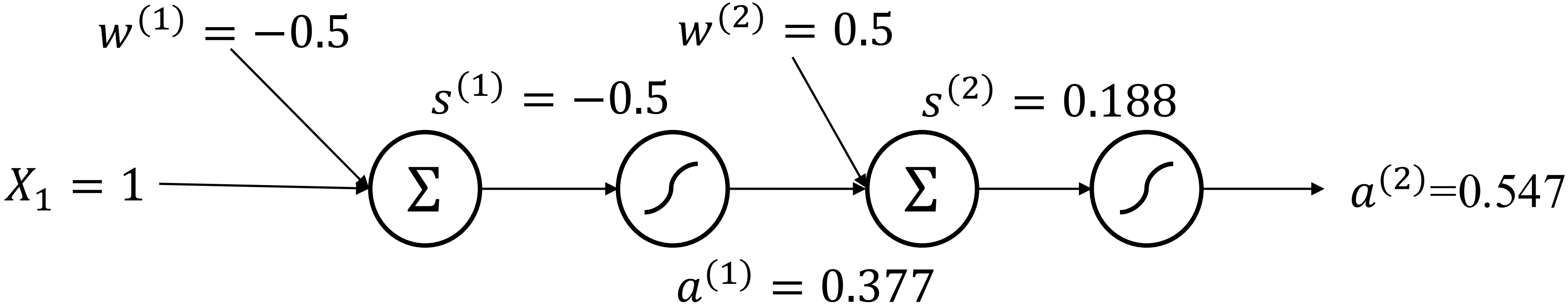
$$\frac{\partial a^{(1)}}{\partial s^{(1)}} = a^{(1)}(1 - a^{(1)})$$

$$\frac{\partial s^{(2)}}{\partial w^{(2)}} = a^{(1)}$$

$$\frac{\partial s^{(1)}}{\partial w^{(1)}} = X_1$$

$$\frac{\partial Loss}{\partial w^{(2)}} = (a^{(2)} - y)a^{(2)}(1 - a^{(2)})a^{(1)}$$

$$\frac{\partial Loss}{\partial w^{(1)}} = (a^{(2)} - y)a^{(2)}(1 - a^{(2)})w^{(2)}a^{(1)}(1 - a^{(1)})X_1$$



$$\frac{\partial Loss}{\partial a^{(2)}} = \frac{\partial \frac{1}{2}(a^{(2)} - y)^2}{\partial a^{(2)}} = (a^{(2)} - y)$$

$$\frac{\partial s^{(2)}}{\partial a^{(1)}} = w^{(2)}$$

$$\frac{\partial a^{(2)}}{\partial s^{(2)}} = a^{(2)}(1 - a^{(2)})$$

$$\frac{\partial a^{(1)}}{\partial s^{(1)}} = a^{(1)}(1 - a^{(1)})$$

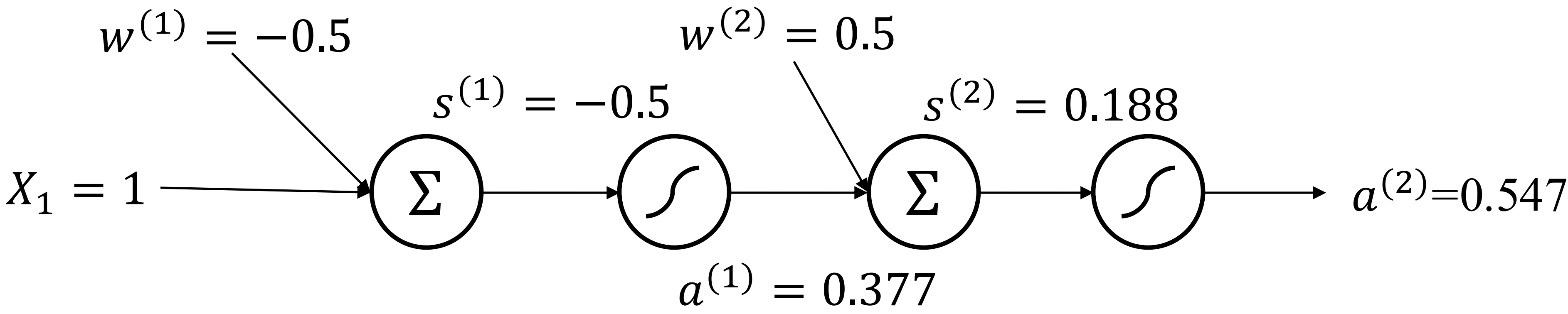
$$\frac{\partial s^{(2)}}{\partial w^{(2)}} = a^{(1)}$$

$$\frac{\partial s^{(1)}}{\partial w^{(1)}} = X_1$$

$$\frac{\partial Loss}{\partial w^{(2)}} = -0.0423$$

$$\frac{\partial Loss}{\partial w^{(1)}} = -0.0131$$

Backward



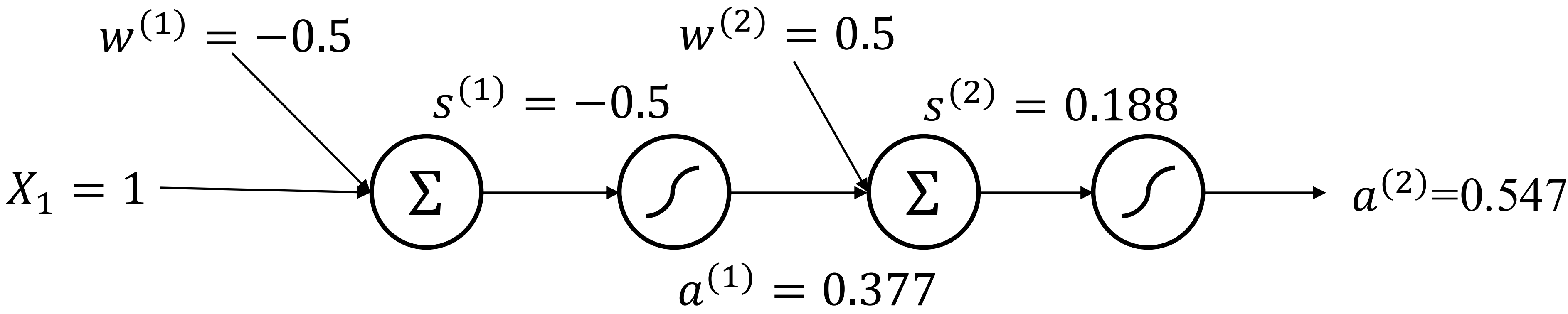
$$\frac{\partial Loss}{\partial w^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial w^{(2)}} = -0.0423$$

$$\frac{\partial Loss}{\partial w^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial w^{(1)}} = -0.0131$$

$$Loss = \frac{1}{2}(0.547 - 1)^2 = 0.102$$

$$w_t = w_{t-1} - \alpha \nabla_w Loss$$

Otimizando os Pesos



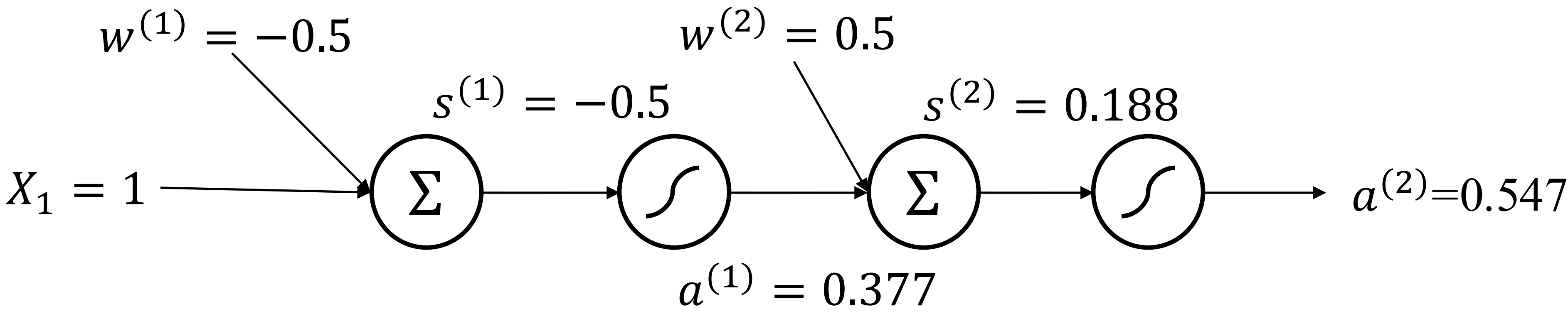
$$w_t^{(2)} = w_{t-1}^{(2)} - \alpha(-0.0423)$$

$$w_t^{(1)} = w_{t-1}^{(1)} - \alpha(-0.0131)$$

$$Loss = \frac{1}{2}(0.547 - 1)^2 = 0.102$$

$$w_t = w_{t-1} - \alpha \nabla_w Loss$$

Otimizando os Pesos

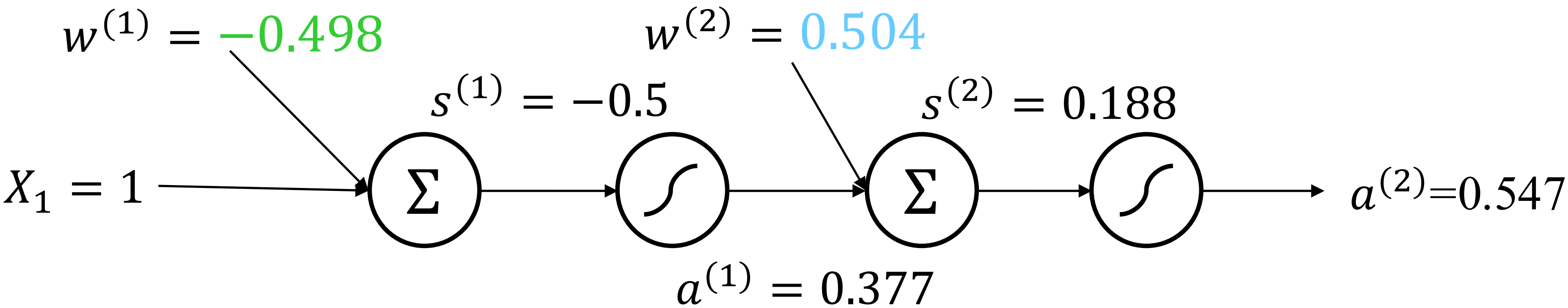


$$w_t^{(2)} = w_{t-1}^{(2)} - \alpha(-0.0423) = 0.5042374$$

$$w_t^{(1)} = w_{t-1}^{(1)} - \alpha(-0.0131) = -0.4986811$$

$Se\ considerarmos\ \alpha = 0.1$

Otimizando os Pesos

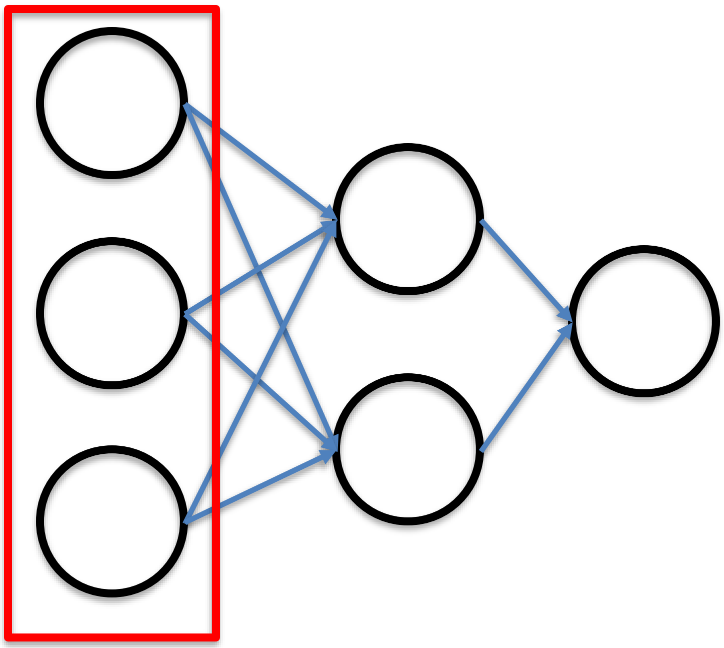


Na Prática

- Frameworks constroem automaticamente o grafo autodiferenciável
- Executam as operações em matrizes
- Utilizam dos conceitos de gradiente local e global

Na Prática

$x \quad = 1 \quad 2 \quad 3$

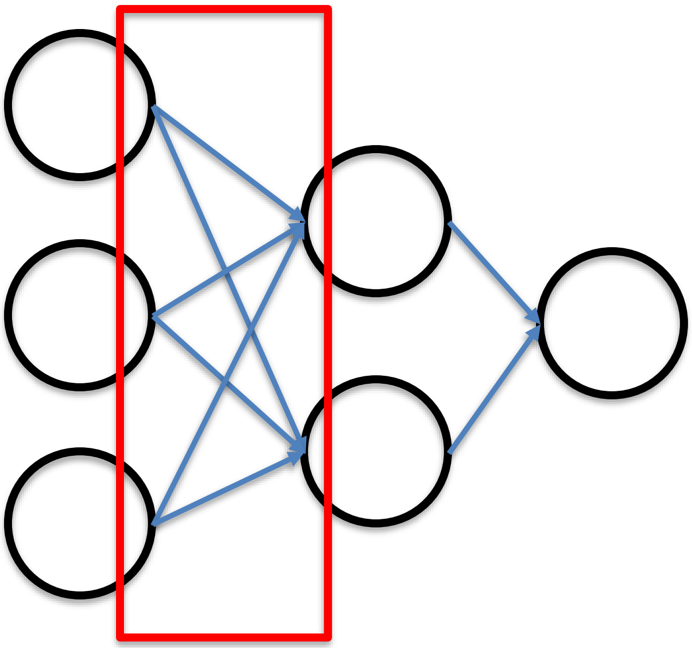


$y = 1$

Na Prática

$$x \quad = 1 \quad 2 \quad 3$$

$$W^{(1)} = \begin{matrix} 0.5 & 0.2 \\ 0.1 & 0.4 \\ 1 & -0.5 \end{matrix}$$



$$y = 1$$

Na Prática

$$x \;\; = 1 \quad 2 \quad 3$$

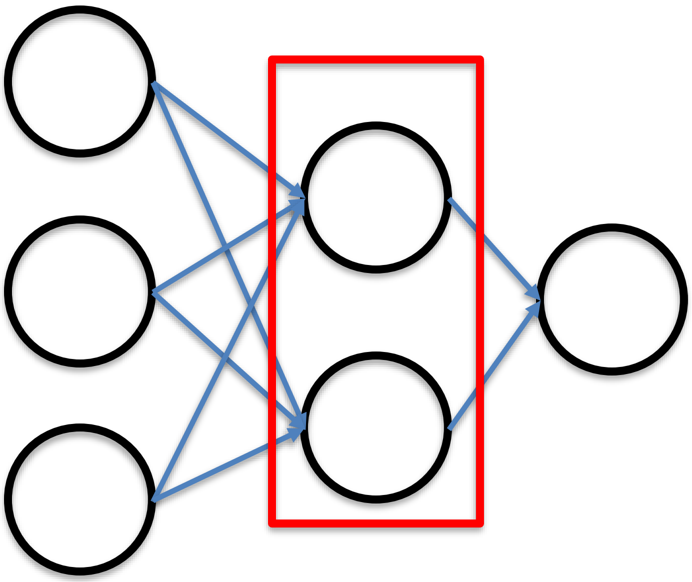
$$W^{(1)} = \begin{matrix} 0.5 & 0.2 \\ 0.1 & 0.4 \\ 1 & -0.5 \end{matrix}$$

$$s^{(1)} = xW^{(1)} = 3.7 \quad -0.5$$

$$a^{(1)} = \sigma\left(s^{(1)}\right) = \sigma(3.7) \quad \sigma(-0.5)$$

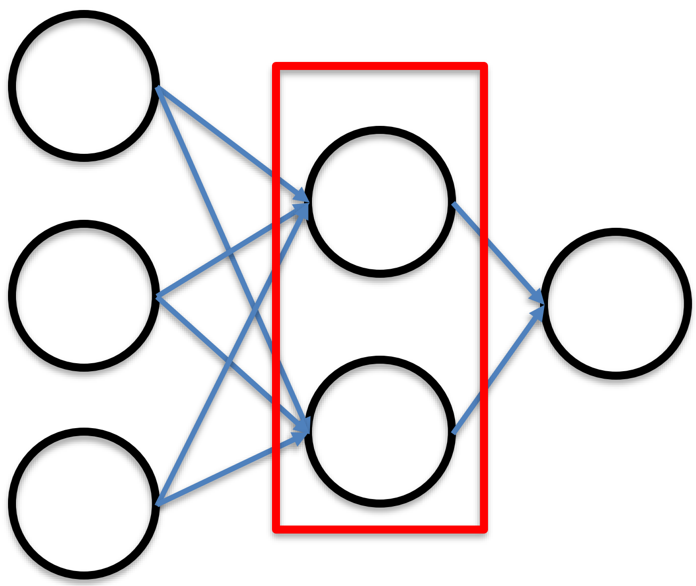
$$a^{(1)} = 0.976 \quad 0.377$$

$$y = 1$$



Na Prática

$a^{(1)} = 0.976 \quad 0.377$

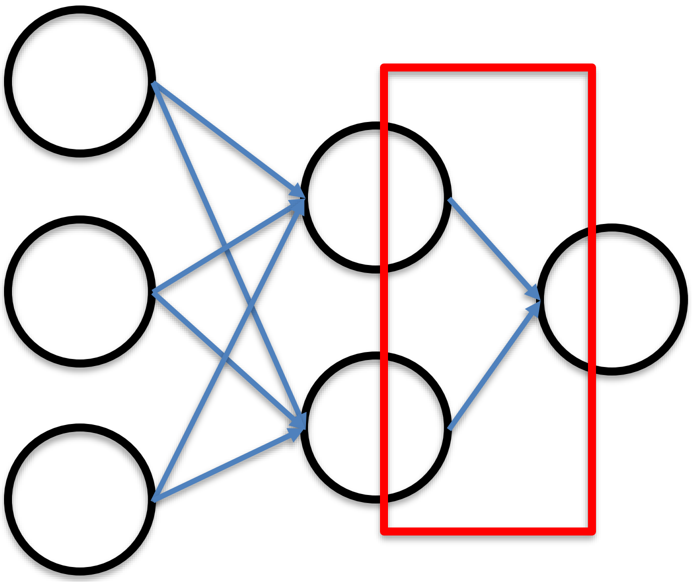


$y = 1$

Na Prática

$$a^{(1)} = 0.976 \quad 0.377$$

$$W^{(2)} = \begin{matrix} 0.3 \\ 0.5 \end{matrix}$$



$$y = 1$$

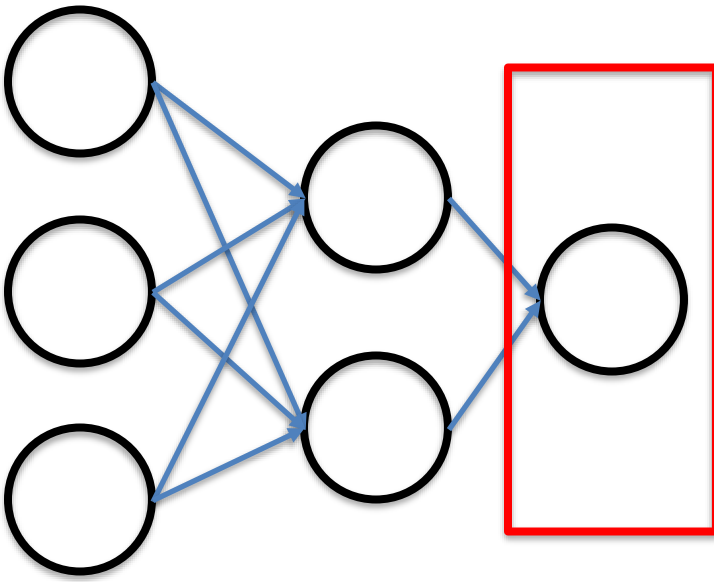
Na Prática

$a^{(1)} = 0.976 \quad 0.377$

$W^{(2)} = \begin{matrix} 0.3 \\ 0.5 \end{matrix}$

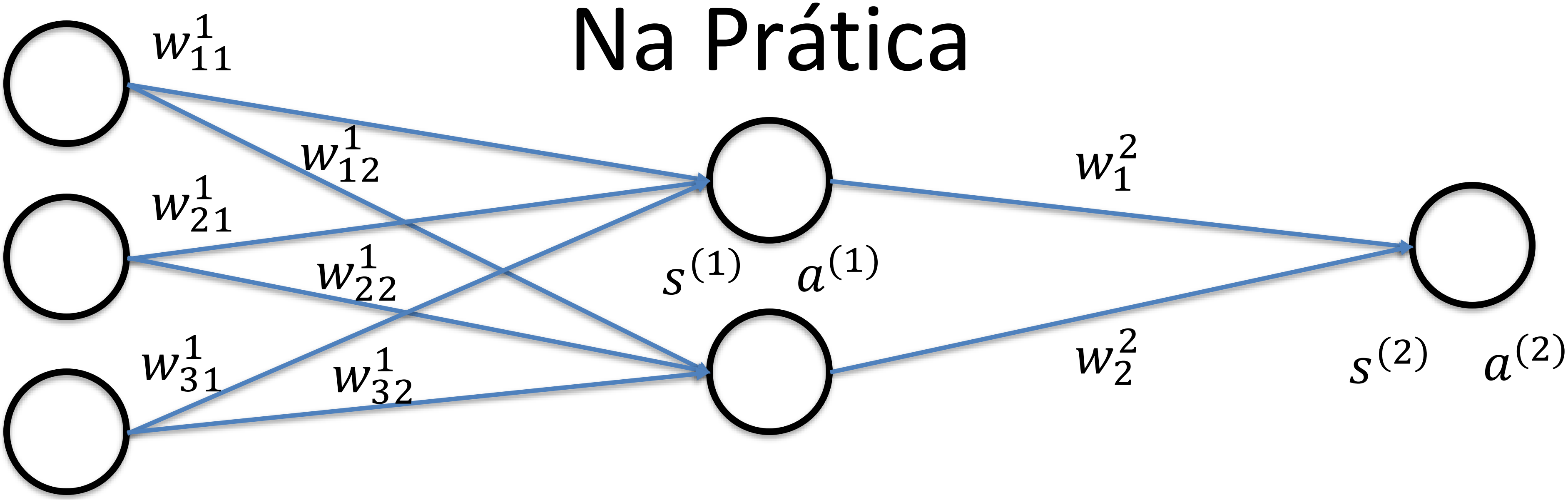
$s^{(2)} = 0.481$

$a^{(2)} = 0.618$



$y = 1$

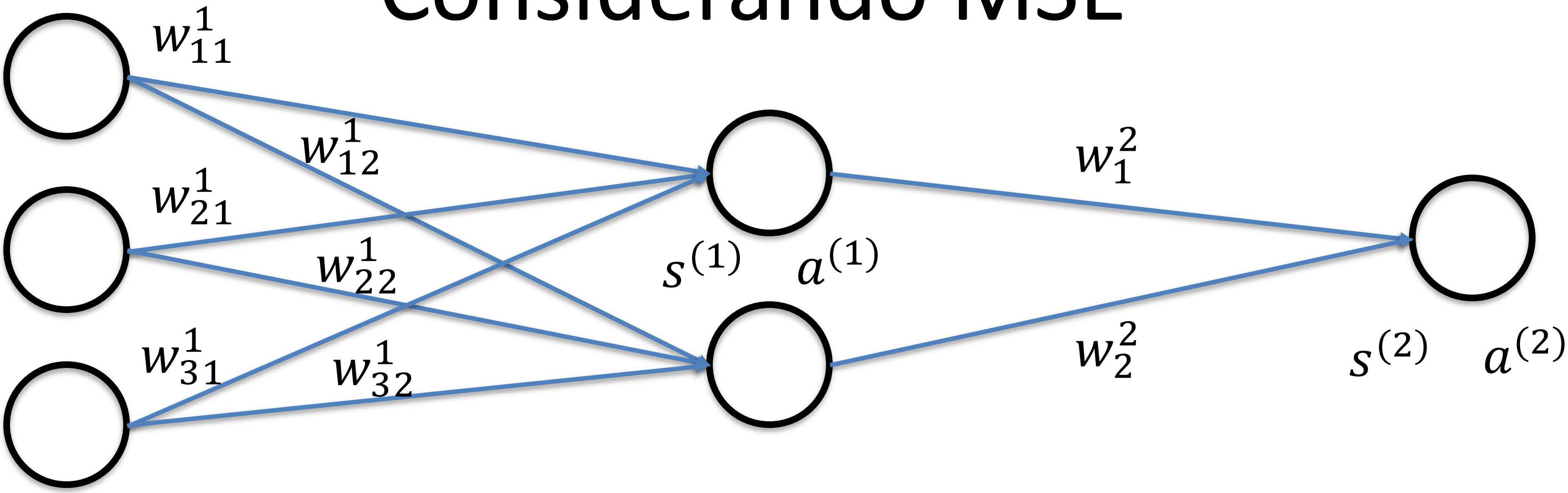
Na Prática



$$\frac{\partial Loss}{\partial W^{(2)}} = \underbrace{\frac{\partial Loss}{\partial a^{(2)}}\frac{\partial a^{(2)}}{\partial s^{(2)}}}_{\text{Gradiente Global}}\underbrace{\frac{\partial s^{(2)}}{\partial W^{(2)}}}_{\text{Gradiente Local}}$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \underbrace{\frac{\partial Loss}{\partial a^{(2)}}\frac{\partial a^{(2)}}{\partial s^{(2)}}\frac{\partial s^{(2)}}{\partial a^{(1)}}\frac{\partial a^{(1)}}{\partial s^{(1)}}}_{\text{Gradiente Global}}\underbrace{\frac{\partial s^{(1)}}{\partial W^{(1)}}}_{\text{Gradiente Local}}$$

Considerando MSE



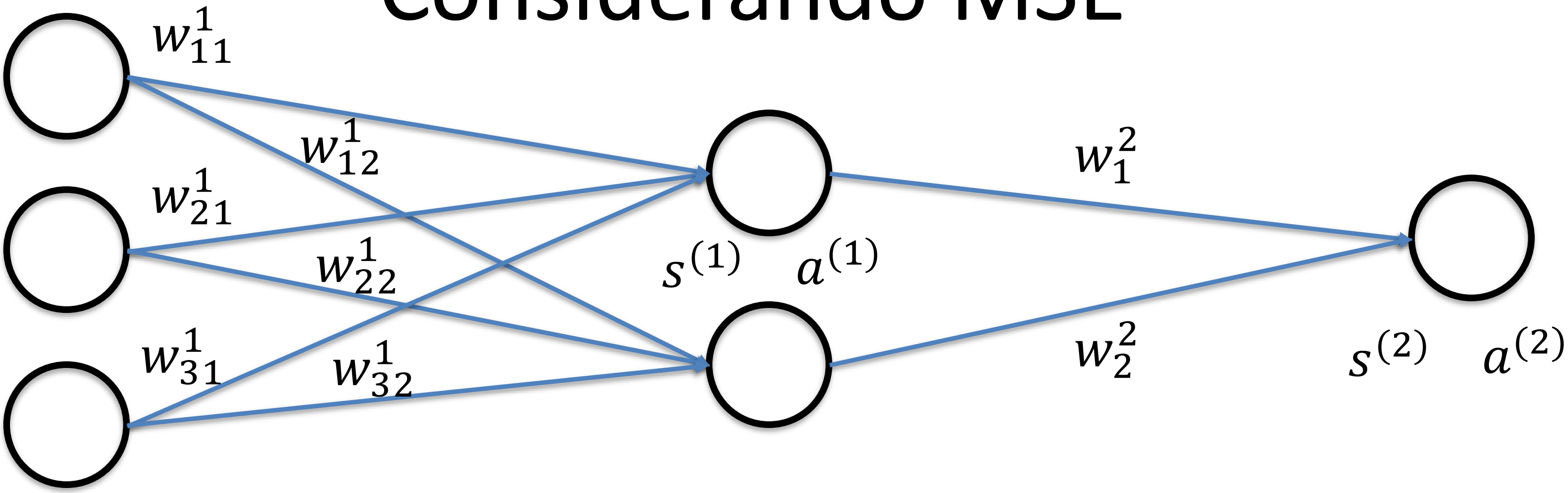
$$\frac{\partial Loss}{\partial W^{(2)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial W^{(2)}}$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial W^{(1)}}$$

$$\frac{\partial Loss}{\partial W^{(2)}} = (1 - 0.618) * 0.618(1 - 0.618) \cdot [0.976 \quad 0.377]$$

$$\frac{\partial Loss}{\partial W^{(2)}} = 0.090 \cdot [0.976 \quad 0.377]$$

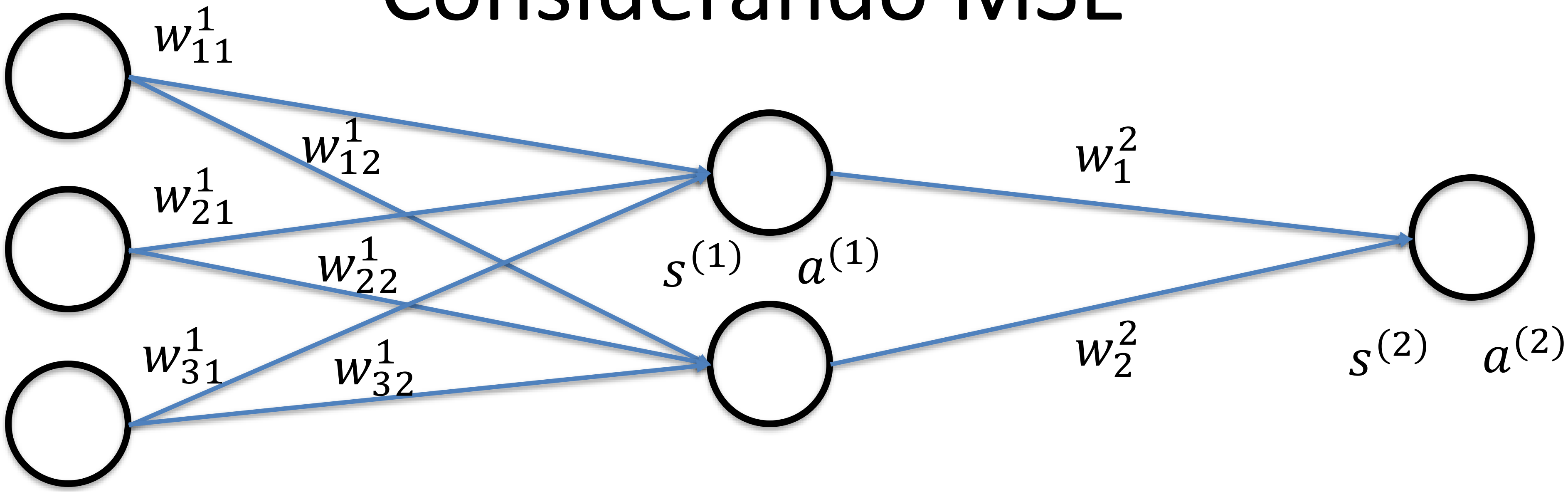
Considerando MSE



$$\frac{\partial Loss}{\partial W^{(2)}} = 0.087 \quad 0.034$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial W^{(1)}}$$

Considerando MSE

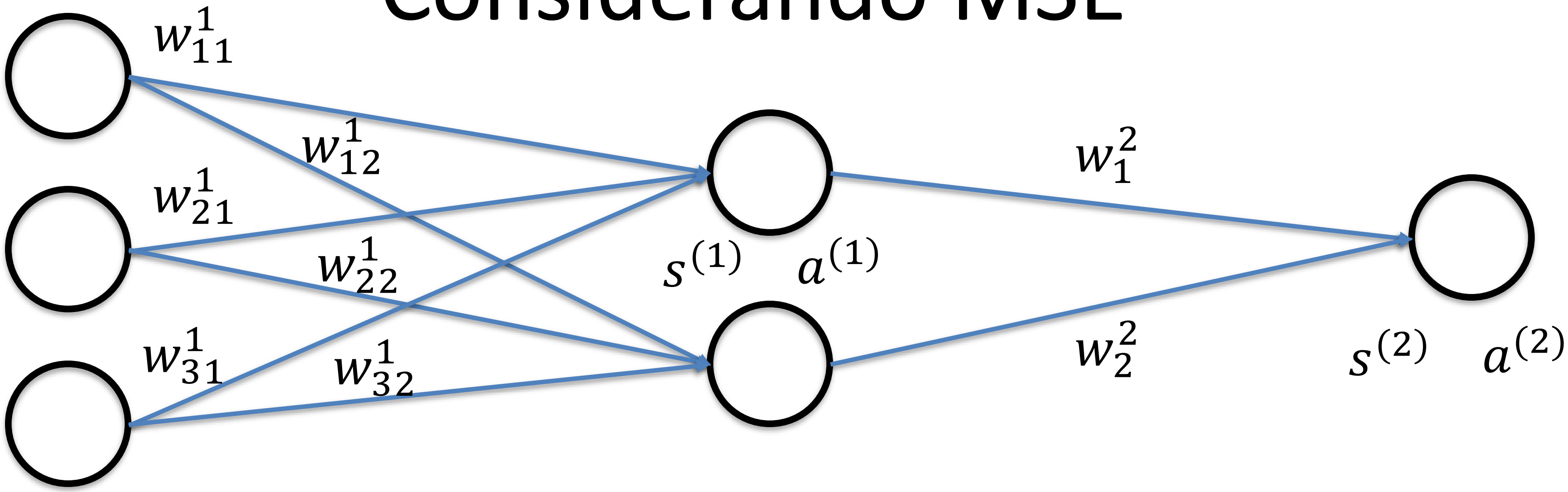


$$\frac{\partial Loss}{\partial W^{(2)}} = 0.087 \quad 0.034$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial W^{(1)}}$$

$$\frac{\partial Loss}{\partial W^{(1)}} = 0.090$$

Considerando MSE

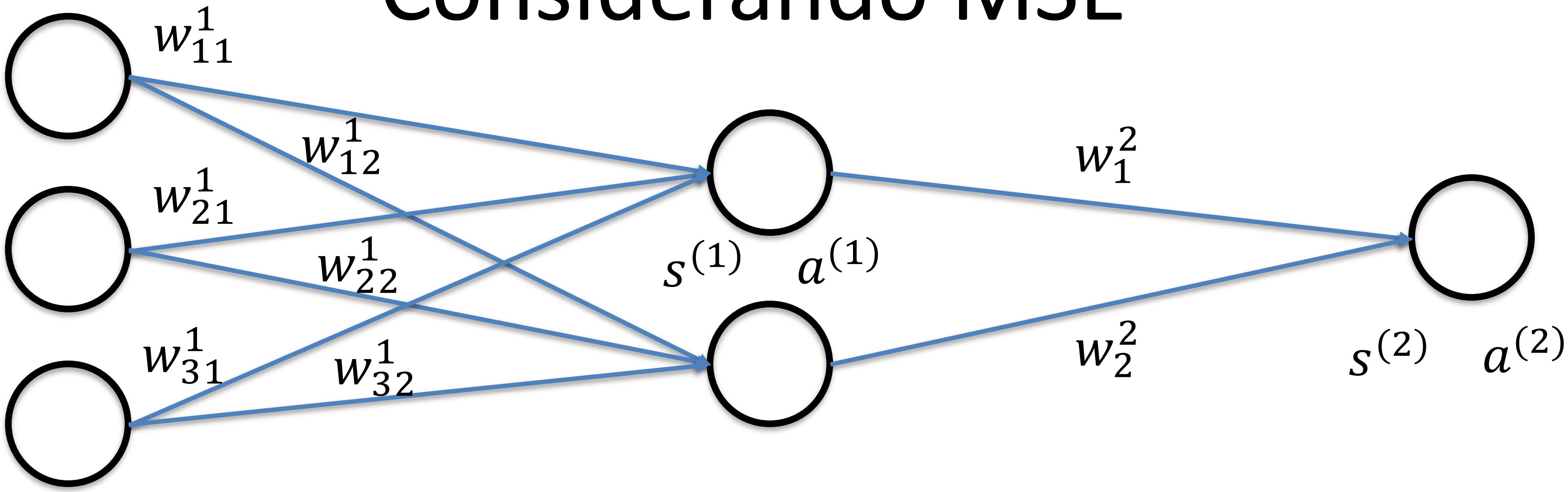


$$\frac{\partial Loss}{\partial W^{(2)}} = 0.087 \quad 0.034$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial W^{(1)}}$$

$$\frac{\partial Loss}{\partial W^{(1)}} = 0.090 \cdot [0.3\ 0.5] \cdot [0.976 \quad 0.377] * (1 - [0.976 \quad 0.377]) * x$$

Considerando MSE

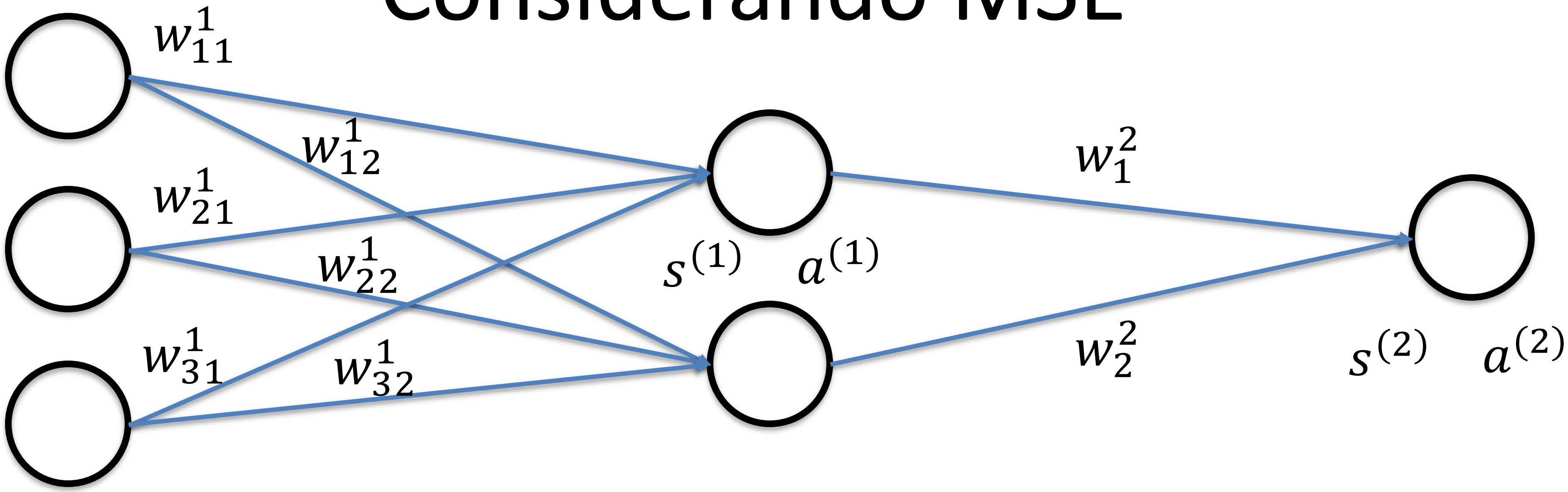


$$\frac{\partial Loss}{\partial W^{(2)}} = 0.087 \quad 0.034$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial W^{(1)}}$$

$$\frac{\partial Loss}{\partial W^{(1)}} = 0.090 \cdot [0.0006\ 0.0340] * [1\ 2\ 3]$$

Considerando MSE

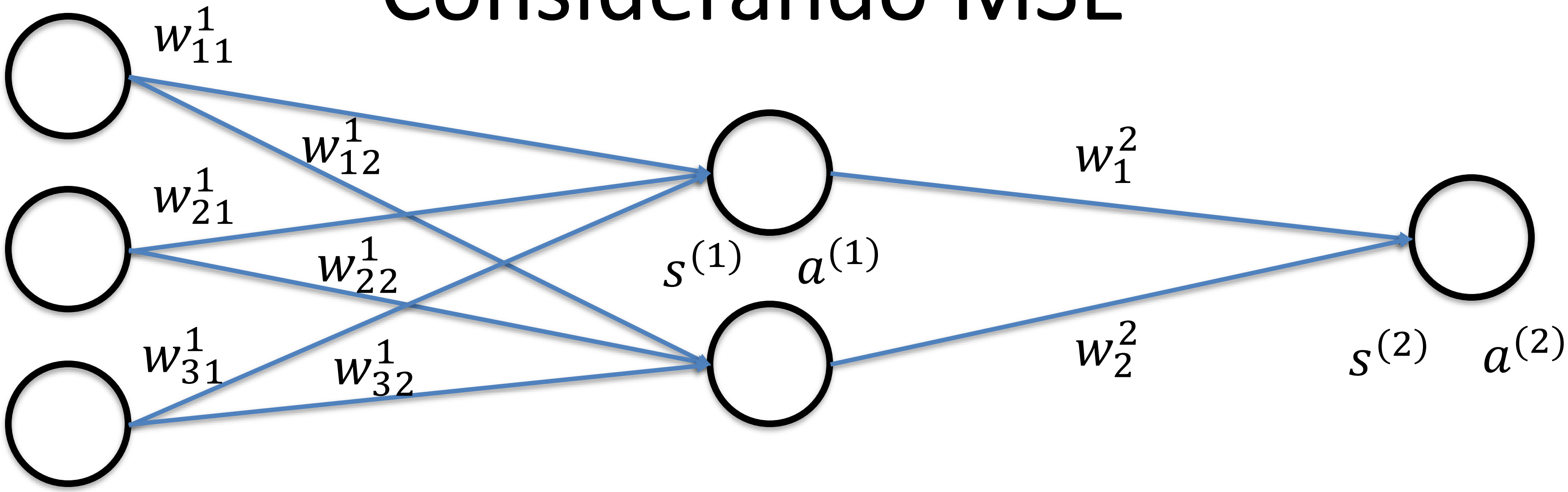


$$\frac{\partial Loss}{\partial W^{(2)}} = 0.087 \quad 0.034$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial W^{(1)}}$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \begin{matrix} 0.0006 & 0.0105 \\ 0.0012 & 0.0211 \\ 0.0019 & 0.0317 \end{matrix}$$

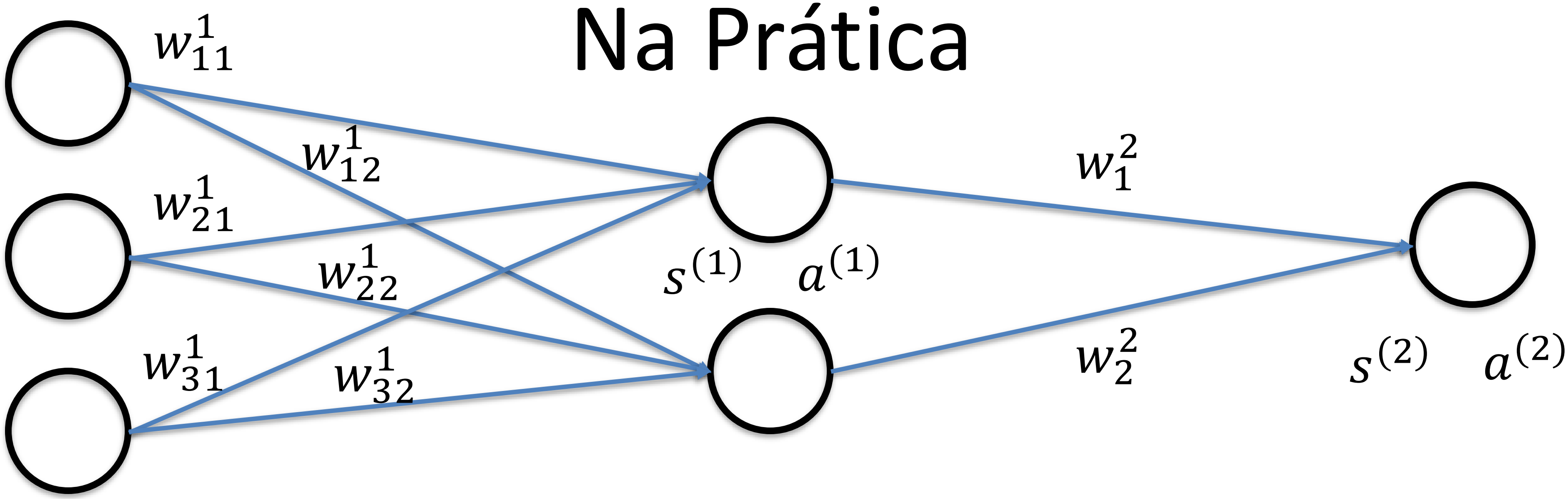
Considerando MSE



$$\frac{\partial Loss}{\partial W^{(2)}} = 0.087 \quad 0.034$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \begin{matrix} 0.0006 & 0.0105 \\ 0.0012 & 0.0211 \\ 0.0019 & 0.0317 \end{matrix}$$

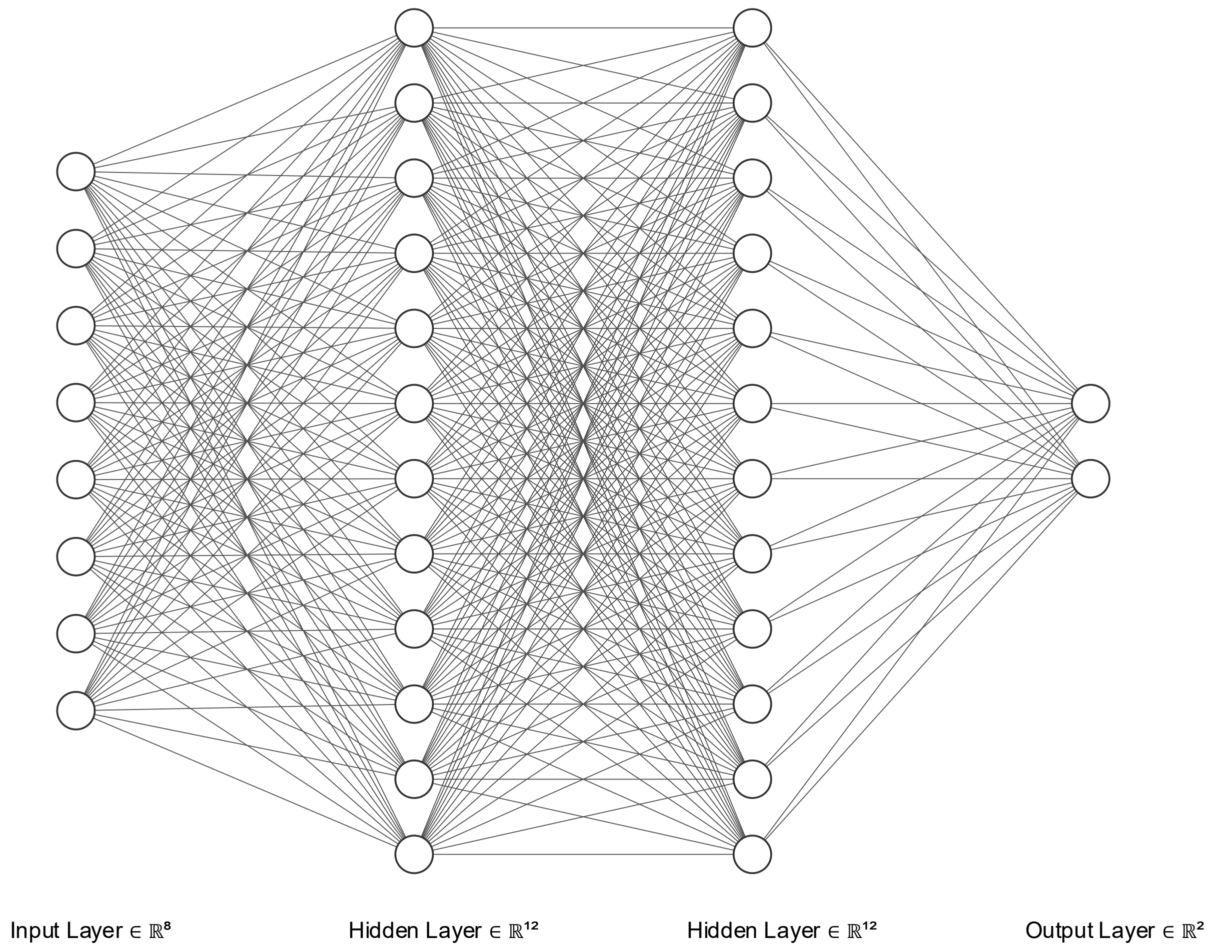
Na Prática



$$\frac{\partial Loss}{\partial W^{(2)}} = \underbrace{\frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}}}_{\text{Gradiente Global}} \underbrace{\frac{\partial s^{(2)}}{\partial W^{(2)}}}_{\text{Gradiente Local}}$$

$$\frac{\partial Loss}{\partial W^{(1)}} = \underbrace{\frac{\partial Loss}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}}}_{\text{Gradiente Global}} \underbrace{\frac{\partial s^{(1)}}{\partial W^{(1)}}}_{\text{Gradiente Local}}$$

Aumentando o tamanho da rede



$$\frac{\partial Loss}{\partial W^{(1)}} = \underbrace{\frac{\partial Loss}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial s^{(3)}} \frac{\partial s^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial a^{(1)}} \frac{\partial a^{(1)}}{\partial s^{(1)}}}_{\text{Gradiente Global}} \underbrace{\frac{\partial s^{(1)}}{\partial W^{(1)}}}_{\text{Gradiente Local}}$$

Referências

- HAYKIN, Simon. **Neural networks and learning machines, 3/E**. Pearson Education India, 2010.
- ABU-MOSTAFA, Yaser S.; MAGDON-ISMAIL, Malik; LIN, Hsuan-Tien. **Learning from data**. New York, NY, USA:: AMLBook, 2012.

- Slides adaptados dos originais dos profs. André Carvalho (ICMC-USP), Ricardo Campello (ICMC-USP), Andrew Ng (Stanford), Rodrigo C. Barros (PUCRS) e Lucas S. Kupssinskü (PUCRS)