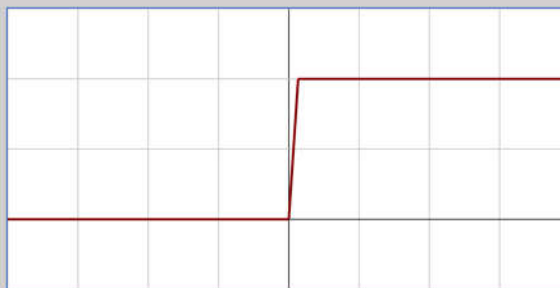


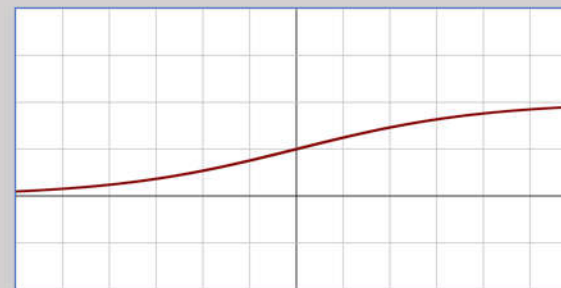
Funções de Ativação

Principais Funções de Ativação

Threshold (Binary Step)



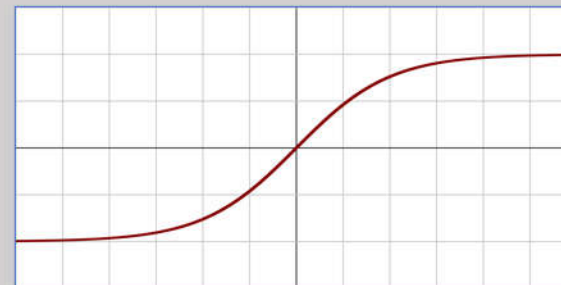
Sigmoid



Relu (Rectified Linear Unit)



Hyperbolic Tangent (tanh)

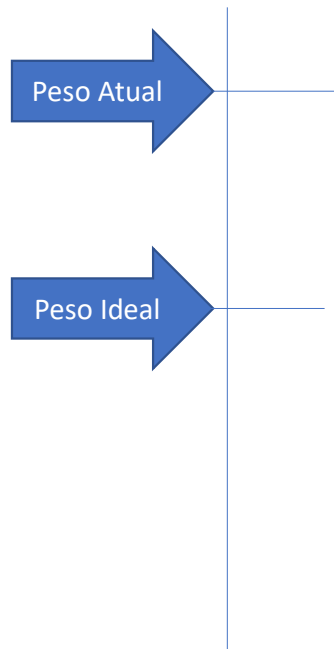




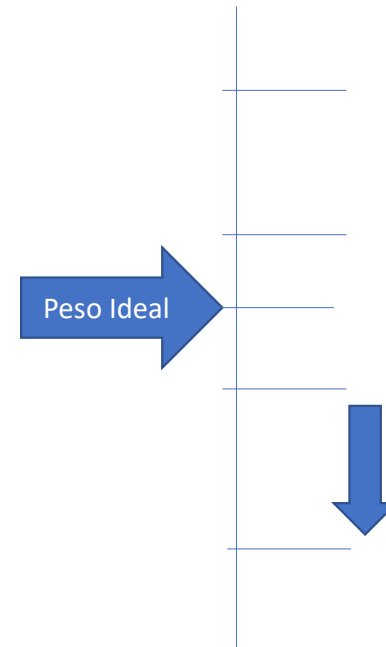
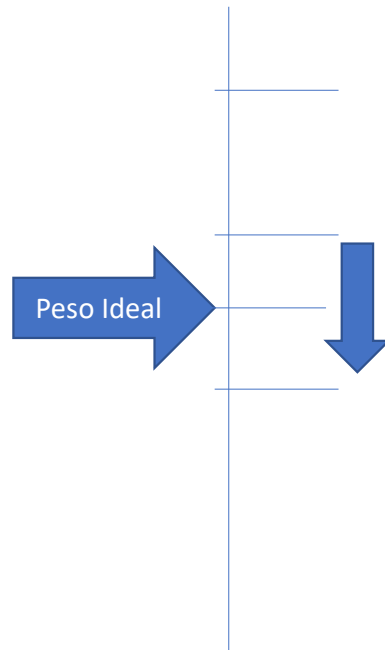
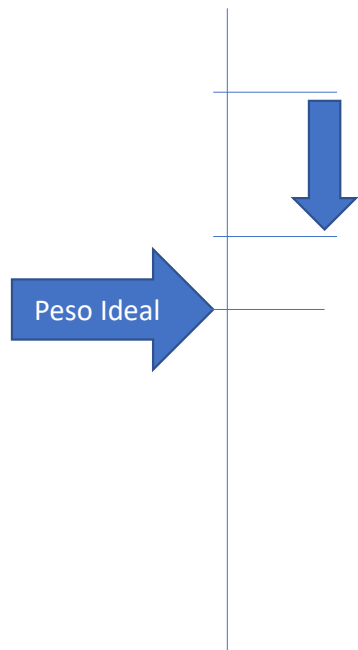
Custo Computacional!

- Uma rede com 25 nós e 1000 pesos para cada sinapse
- $1.000 \times 1.000 \times \dots \times 1.000 = 1.000^{25} = 10^{75}$ combinações
- Tempo aproximado:
 - $3,42 \times 10^{50}$ anos

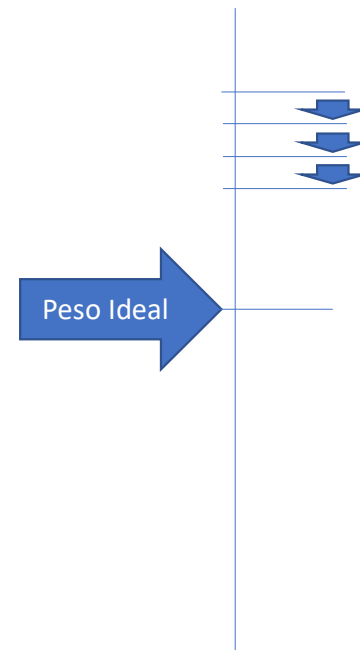
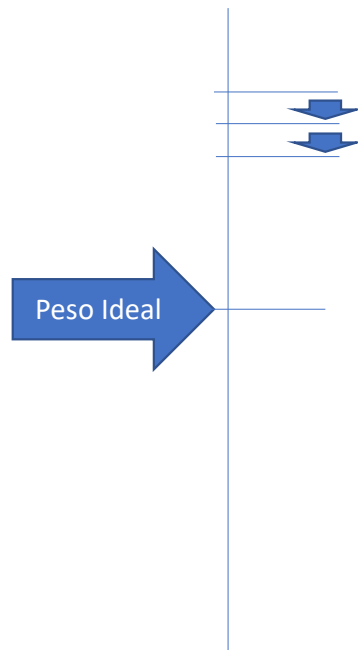
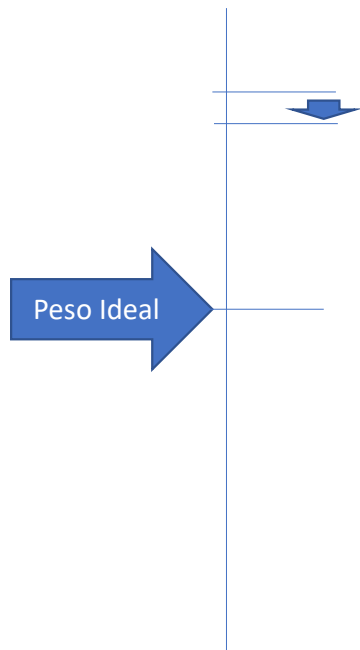
Learning Rate



Learning Rate "Grande"

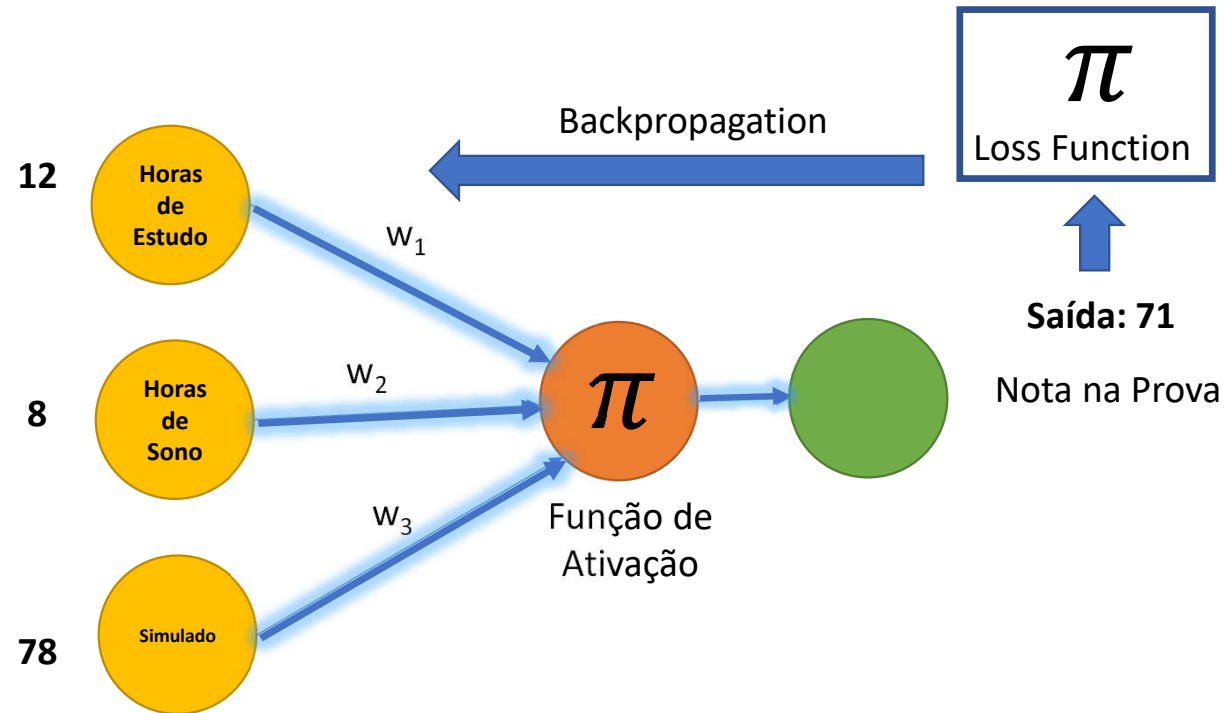


Learning Rate "Pequeno"



Gradient Descent

ID	HoarsEstudo	HorasSono	Simulado	Prova
1	12	6	78%	93%
2	22	6,5	24%	68%
3	115	4	100%	95%
4	31	9	67%	75%
5	0	10	58%	51%
6	5	8	78%	60%
7	92	6	82%	89%
8	57	8	91%	97%



Gradient Descent

- É uma Função de Otimização
- Ela diz em que direção em que proporção atualizar os pesos
- Reduz o tempo de treinamento
- Encontra pesos otimizados

Topologia e Arquiteturas



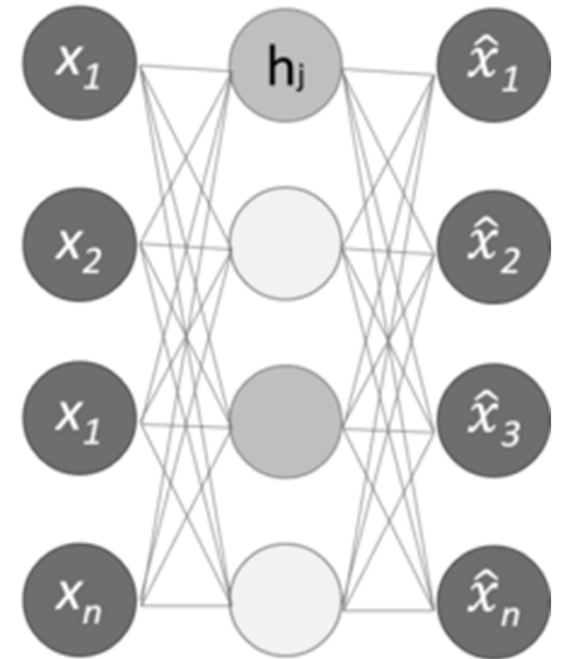
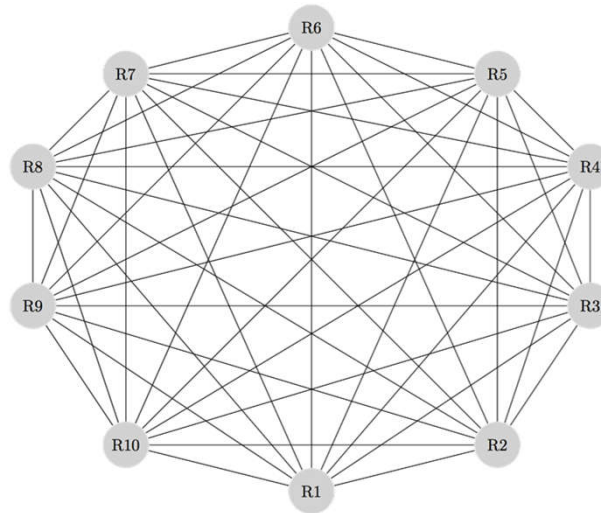
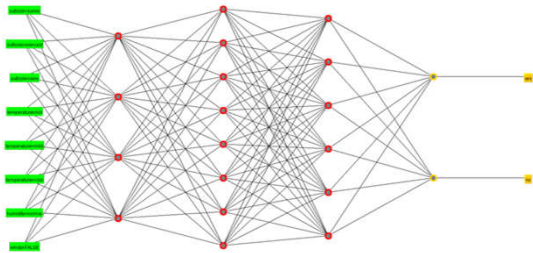
Topologia

Estrutura de nós e camadas



Arquitetura

Feed Forward
Recorrente - Retorno



Topologia

Topologia

- Quantas camadas?

➤ Entrada: 1

➤ Saída: 1

➤ Oculta: n

Topologia

- Quantos nós?
- Entrada: Um para cada atributo / valor
- Saída: Valores de Classes para Classificação ou 1 para Regressão
- Oculta: $\frac{a+c}{2}$



Regressão vs Classificação

- Regressão: Prever valor numérico
- Classificação: Prever valor categórico