



# Classificação de Casos de Doenças Reumáticas

Fernando Concatto





# Introdução

# Introdução

- Doenças reumáticas são caracterizadas por dores e rigidez em articulações, tendões e músculos.
- Grande parte das doenças são crônicas e algumas não possuem cura.
- Segundo a Organização Mundial da Saúde, os tipos mais comuns de doenças reumáticas são a **Artrite Reumatoide** e a **Osteoartrite**, além de traumas, inflamações e dores na coluna vertebral.

# Objetivos e Materiais

- Buscou-se implementar uma Rede Neural Artificial alimentada adiante (*feedforward*) para realizar a classificação de casos de artrite, identificando a categoria da patologia.
- A base de dados contém 49 casos reais, especificando sintomas, exames laboratoriais e o diagnóstico médico (17 entradas e 3 saídas).
- Os dados foram cedidos pela Clínica Mont Godinne, da Université Catholique de Louvain, na Bélgica.



# Variáveis

# Sintomas

Sintoma	Descrição	Não	Sim
Artrite	Dor e inflamação nas articulações, independente de tipo	44,9%	55,1%
Rigidez Matinal	Dores e dificuldade na movimentação após acordar	65,3%	34,7%
Bursite	Inflamação das bolsas nas ligações entre ossos e tendões	87,8%	12,2%
Tofo	Presença de cristais de ácido úrico nas articulações	81,6%	18,4%
Sinovite	Inflamação na membrana que reveste as articulações	55,1%	44,9%
Artralgia	Presença de dor nas articulações, porém sem inflamação	26,5%	73,5%
Nódulos Reumatóides	Presença de nódulos inchados subcutâneos	95,9%	4,1%

# Sintomas

Sintoma	Descrição	A / Não	B / Sim	C
HLA-B27	Presença do antígeno leucócito B27 no sangue	61,2%	12,2%	26,6%
Deformação nas Juntas	Alterações estruturais anormais nas juntas	95,9%	4,1%	0%
Dor Lombar	Presença de dor na região lombar do paciente	59,2%	40,8%	0%
Rigidez na Coluna	Dificuldade na movimentação da coluna vertebral	77,6%	22,4%	0%
Deformação na Coluna	Desvios ou defeitos na região da coluna vertebral	87,8%	12,2%	0%
Mobilidade	Situação da capacidade de movimento	51%	49%	0%
Dor no Sacroilíaco	Dor ao toque nas articulações da região pélvica	83,7%	16,3%	0%

# Exames Laboratoriais

Exame	Descrição	A	B	C	D	E
Inflamação Laboratorial	Identificação de regiões com atividade inflamatória	42,8%	18,4%	28,6%	4,1%	6,1%
Evidências Radiológicas	Identificação de anormalidades nos corpo por meio de radiografias	38,8%	4,1%	30,6%	16,3%	10,2%
Tomografia Computadorizada	Identificação de anormalidades via imagens em diversos cortes da estrutura anatômica	65,3%	4,1%	20,4%	10,2%	0%



## Diagnóstico

Espondilite Anquilosante

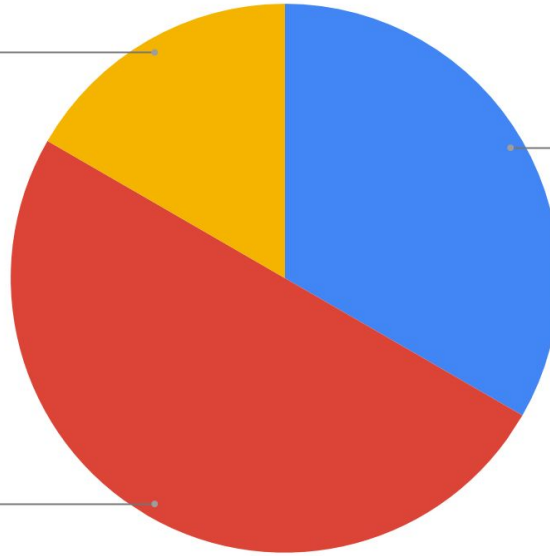
16,7%

Artrite de Gota

33,3%

Artrite Reumatoide

50,0%



Distribuição das patologias diagnosticadas nos casos clínicos.



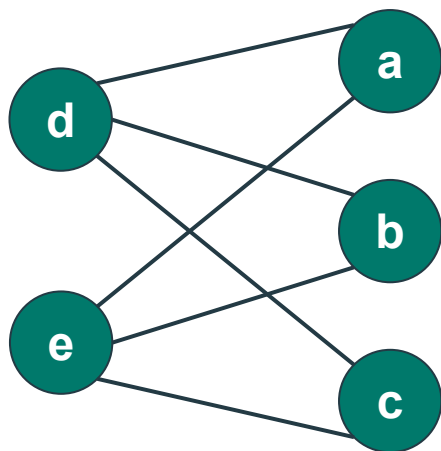
# Metodologia

# Implementação

- A implementação da rede neural foi efetuada com a linguagem Python, utilizando a biblioteca NumPy para manipulação numérica.
- Adotou-se uma abordagem inteiramente matricial para os parâmetros treináveis e valores intermediários da rede.
- Como técnica de treinamento, o algoritmo *Gradient Descent* foi utilizado, buscando minimizar a função do **Erro Quadrático Médio**.

# Arquitetura

- Cada camada possui uma matriz de pesos, uma matriz de *biases* e três para valores intermediários (gradientes, campos locais induzidos e saída).
- A matriz de pesos possui dimensionalidade  $\mathbf{M} \times \mathbf{N}$ , onde M denota o número de neurônios da camada anterior e N o número de neurônios da camada atual.
- A matriz de *biases* possui dimensões de  $\mathbf{1} \times \mathbf{N}$ , e as intermediárias possuem  $\mathbf{K} \times \mathbf{N}$ , onde K representa o número de exemplos da base.



	<b>a</b>	<b>b</b>	<b>c</b>
<b>d</b>	1.42	-1.08	1.79
<b>e</b>	0.74	0.39	1.44

Modelagem da matriz de pesos das camadas da rede neural.

# Propagação para frente

- A propagação é realizada de camada em camada, onde a principal operação é a multiplicação de matrizes.

$$\mathbf{Z}_i = \mathbf{A}_{i-1} \mathbf{W}_i + \boldsymbol{\beta}_i$$

$$\mathbf{A}_i = \varphi(\mathbf{Z}_i)$$

## Propagação para frente

- A propagação ocorre de camada em camada, e a operação é a multiplicação de matrizes.

**Pesos**

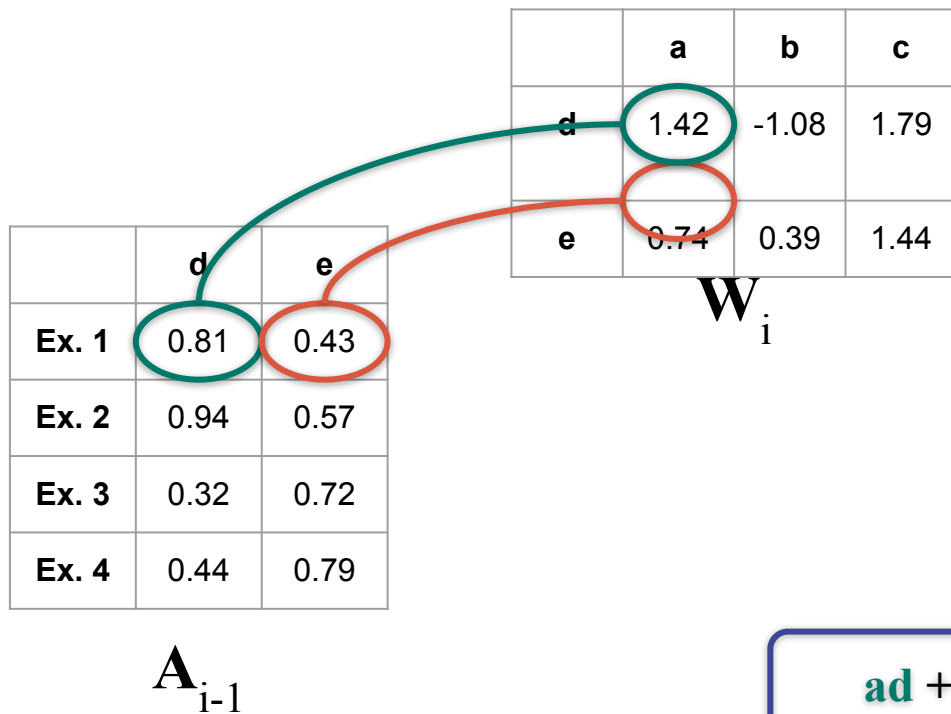
**Saída anterior**

**Biases**

**Função de ativação**

$$\mathbf{Z}_i = \mathbf{A}_{i-1} \mathbf{W}_i + \boldsymbol{\beta}_i$$

$$\mathbf{A}_i = \varphi(\mathbf{Z}_i)$$



$$Z_i - \beta_i$$

	a	b	c
Ex. 1	1.47	-0.71	2.07
Ex. 2	1.76	-0.79	2.5
Ex. 3	0.99	-0.06	1.61
Ex. 4	1.21	-0.17	1.93

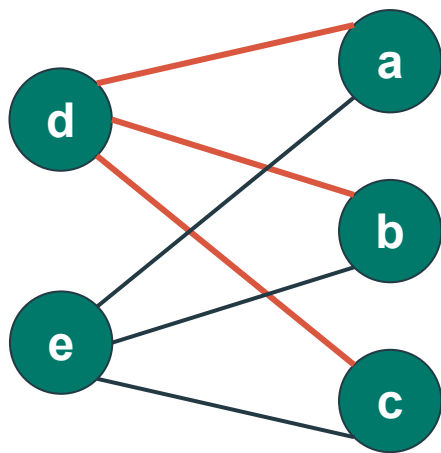
$$ad + ae$$

Formulação matricial da propagação adiante, exceto *bias* e ativação.



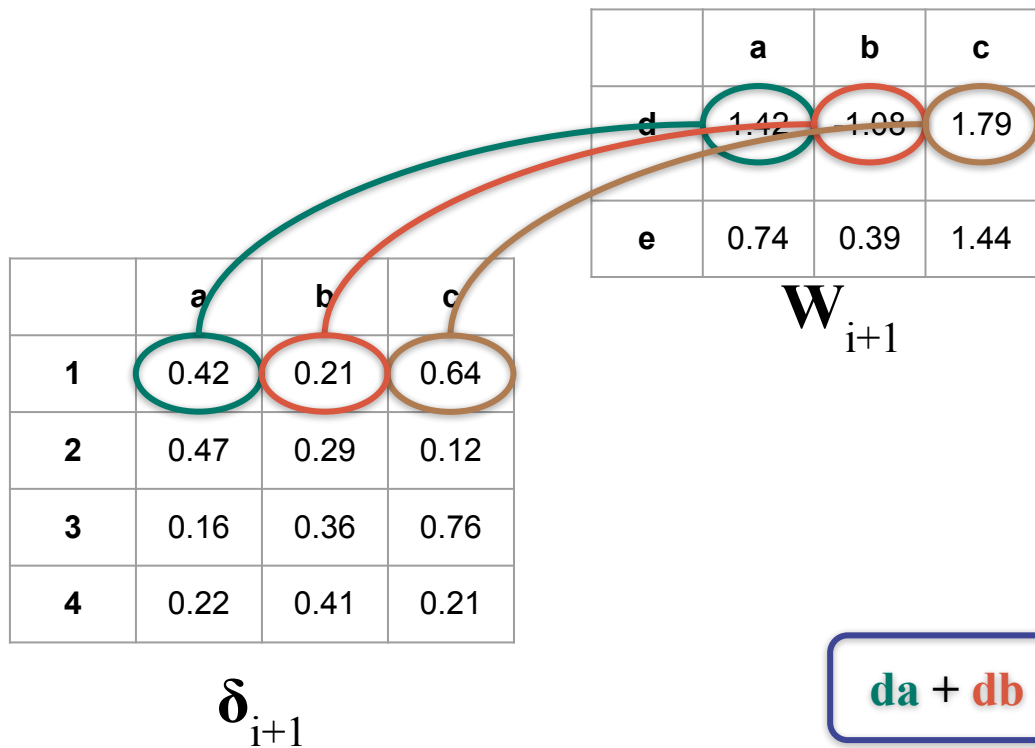
# Propagação para trás

- Segundo Haykin, o **gradiente** ( $\delta$ ) de um neurônio é dado pela aplicação da derivada da função de ativação sobre seu campo local induzido multiplicado por seu sinal de erro ( $e$ ).
- O sinal de erro na camada de saída é dado pela diferença entre a saída e o valor esperado, enquanto as demais envolvem o somatório dos produtos dos pesos sinápticos com os gradientes da camada à frente.

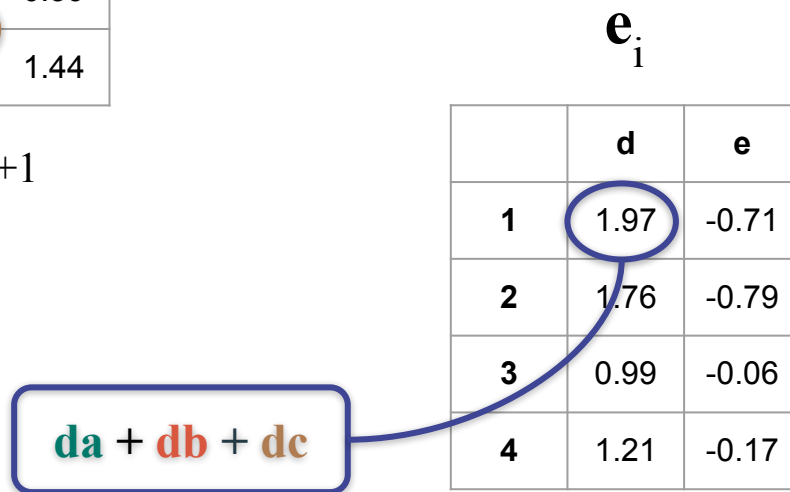
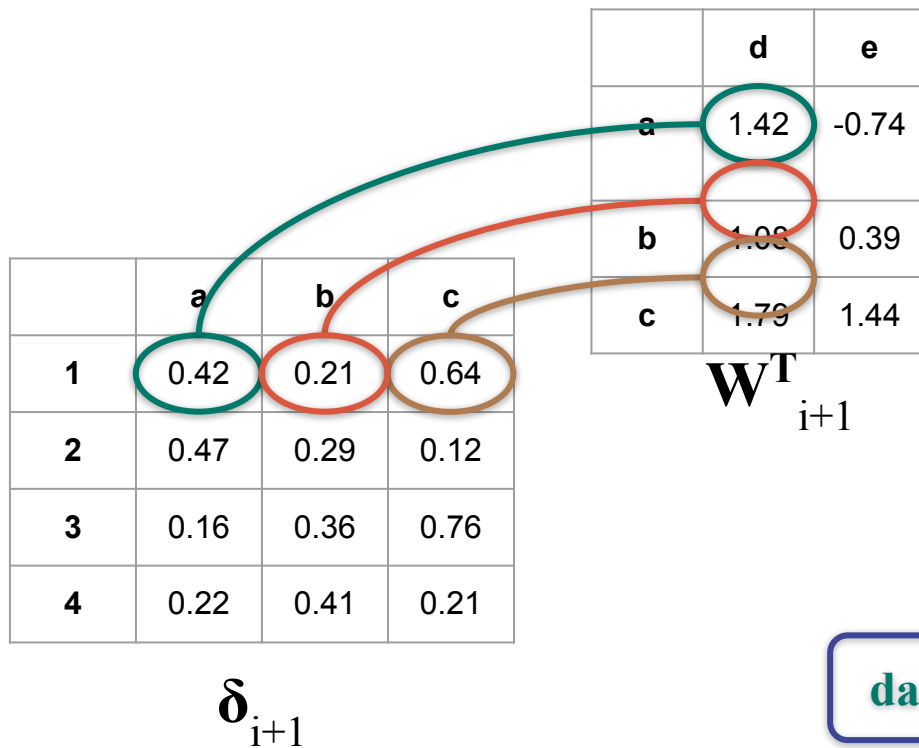


$$\mathbf{e}_d = \delta_a * \mathbf{W}_{da} + \delta_b * \mathbf{W}_{db} + \delta_c * \mathbf{W}_{dc}$$

Modelagem da matriz de pesos das camadas da rede neural.



Formulação matricial do cômputo dos gradientes.



Formulação matricial do cômputo dos gradientes, corrigida.

# Atualização de pesos

- A mudança ( $\Delta$ ) de um peso sináptico para um exemplo é dado pela saída do neurônio à esquerda multiplicado pelo gradiente do neurônio à direita, ponderado pela taxa de aprendizagem.
- Para os *biases*, considera-se que saída à esquerda é igual a 1.
- A mudança total é dada pela soma das mudanças através de todos os exemplos de treinamento.

	d	e
1	0.42	0.21
2	0.47	0.29
3	0.16	0.36
4	0.22	0.41

	a	b	c
1	1.42	-0.74	0.21
2	1.08	0.39	0.63
3	1.78	1.44	0.75
4	0.53	1.24	0.42

$\Delta_{i-1}$

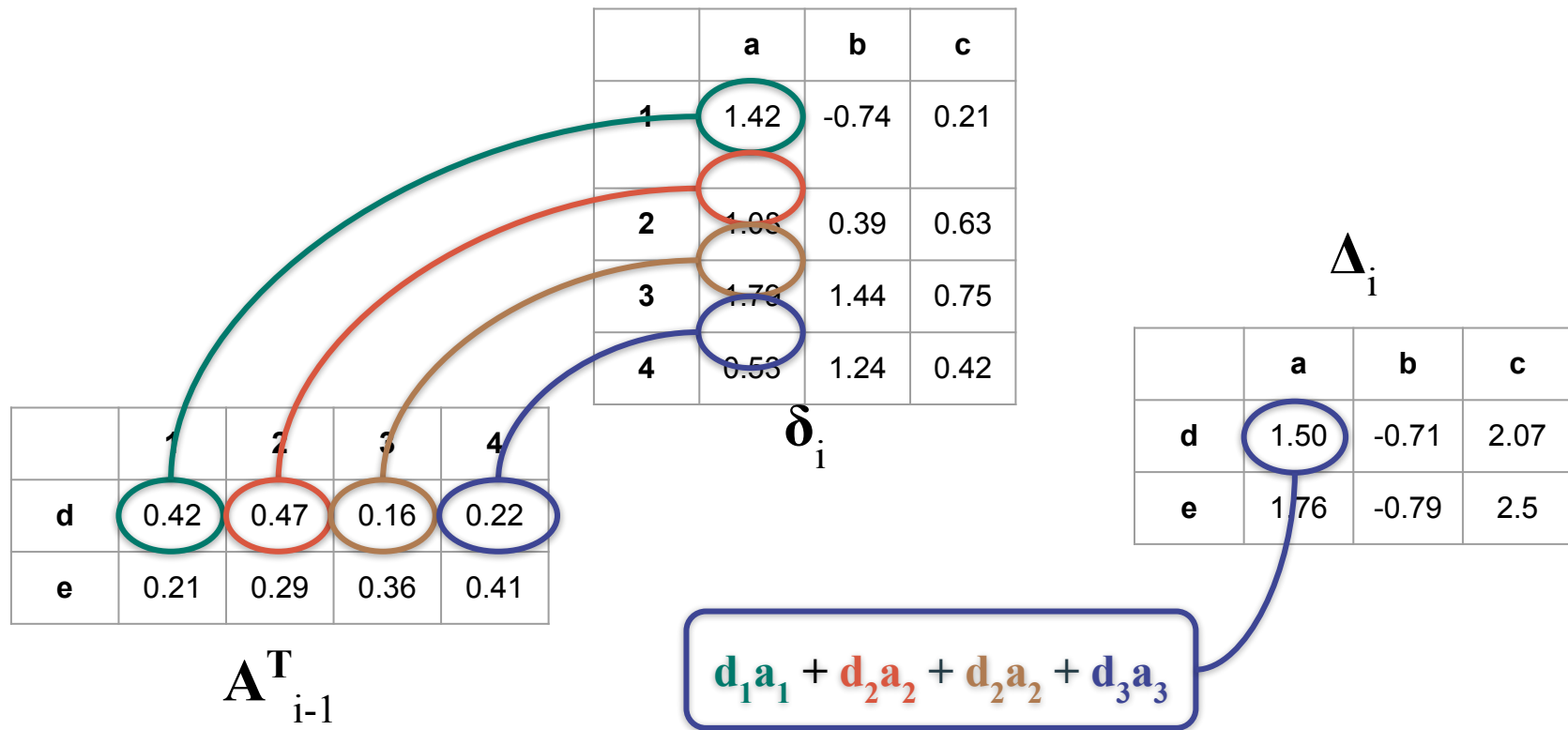
$\delta_i$

$\Delta_i$

	a	b	c
d	1.50	-0.71	2.07
e	1.76	-0.79	2.5

$$d_1 a_1 + d_2 a_2 + d_2 a_2 + d_3 a_3$$

Formulação matricial do cômputo das mudanças de pesos.



Formulação matricial do cômputo das mudanças de pesos, corrigido.



# Experimentos

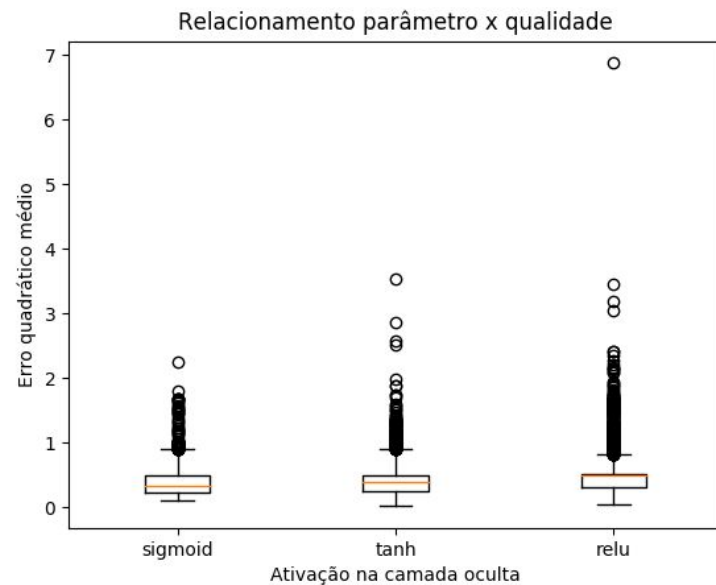
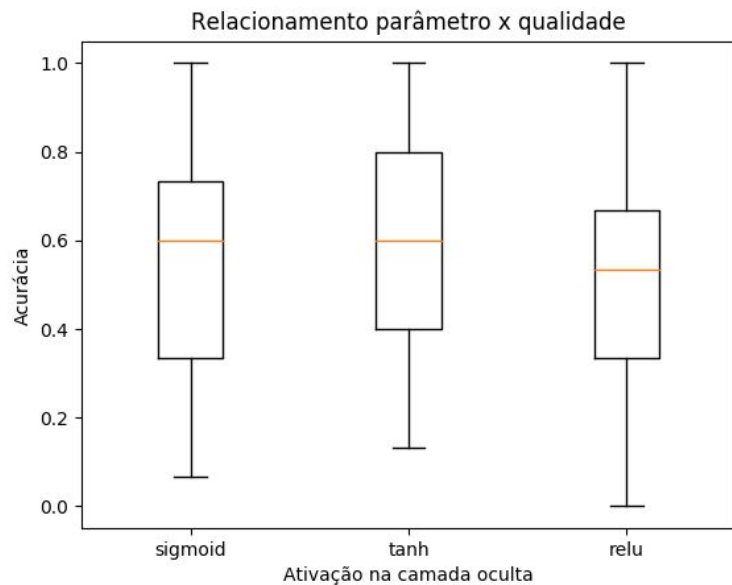


# Configuração dos experimentos

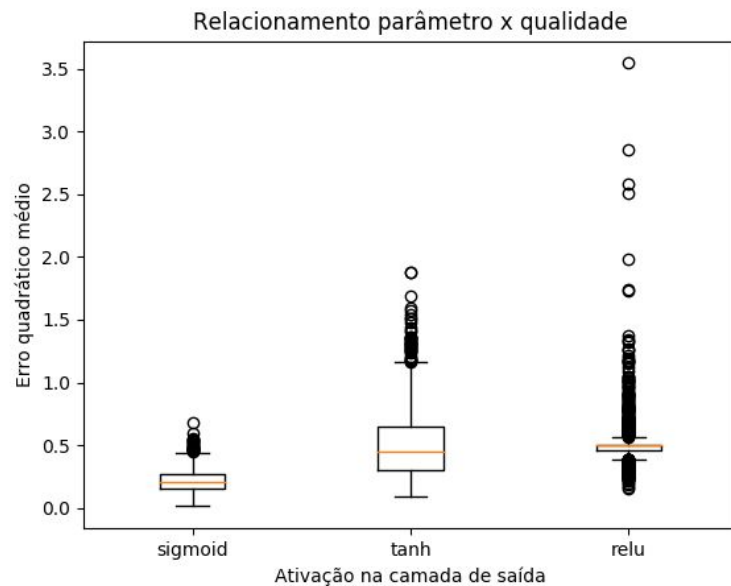
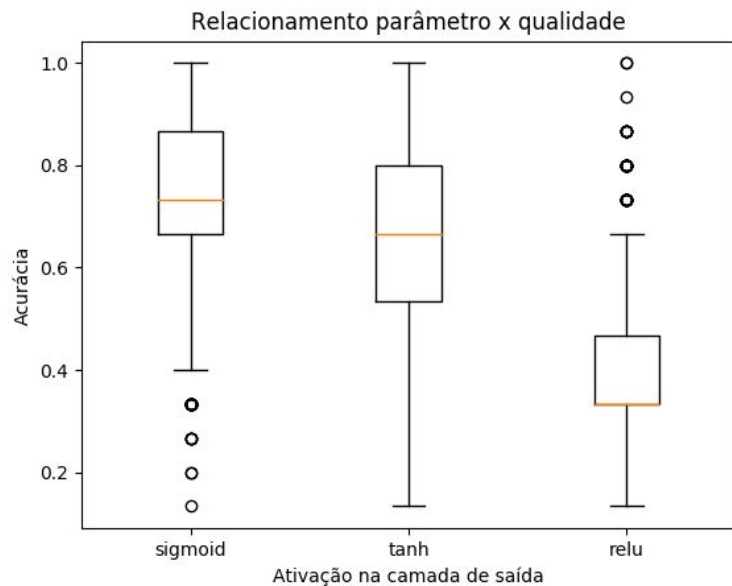
- Dividiu-se a base de 49 exemplos em treinamento (70%) e validação (30%), resultando em 34 e 15 exemplos, respectivamente.
- Coletou-se o erro quadrático médio e a acurácia de cada época.
- Considerou-se como saída da rede a classe com maior valor na camada de saída. Quando esta era igual à esperada, contabilizava-se um acerto.
- Para cada configuração de parâmetros, 10000 épocas foram executadas.

# Otimização de parâmetros

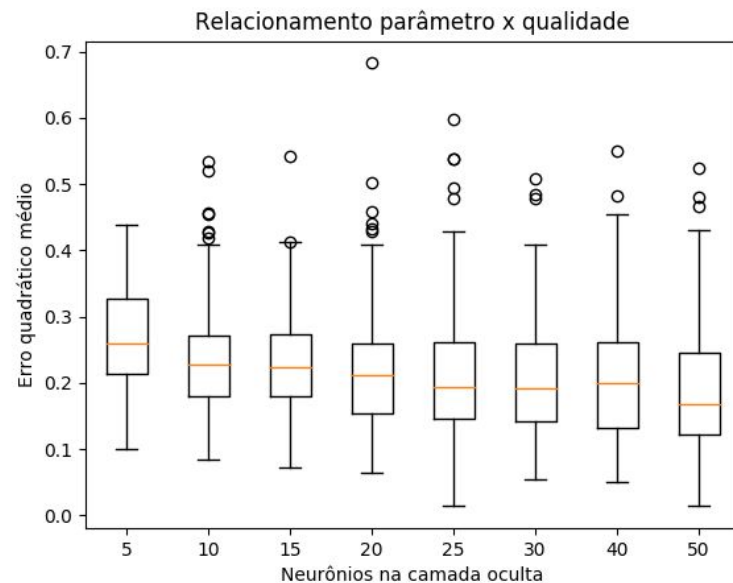
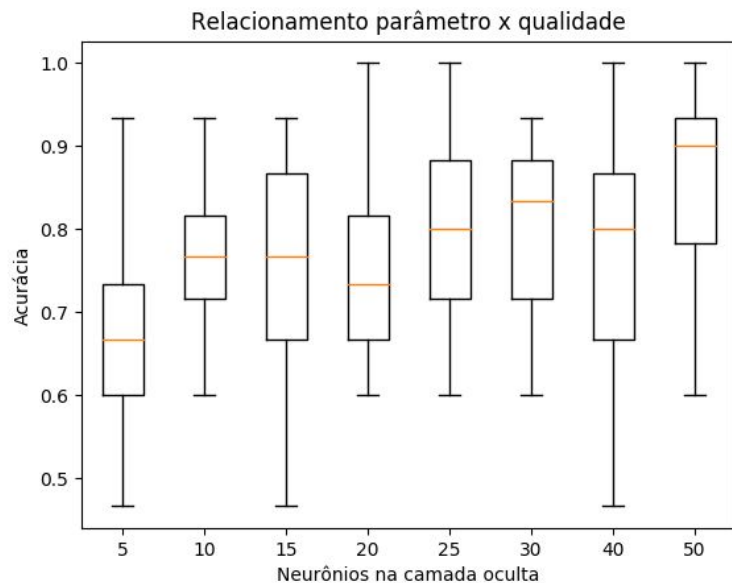
- Todas as combinações dos seguintes valores foram testados:
  - Taxa de Aprendizagem: 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4
  - Função de ativação: Logística, Tangente hiperbólica, ReLU
  - Neurônios: 5, 10, 15, 20, 25, 30, 40, 50
- Para cada configuração, 24 replicações foram executadas.
- Ao total, 13824 execuções foram efetuadas, coletando a melhor acurácia e erro quadrático médio no conjunto de validação.



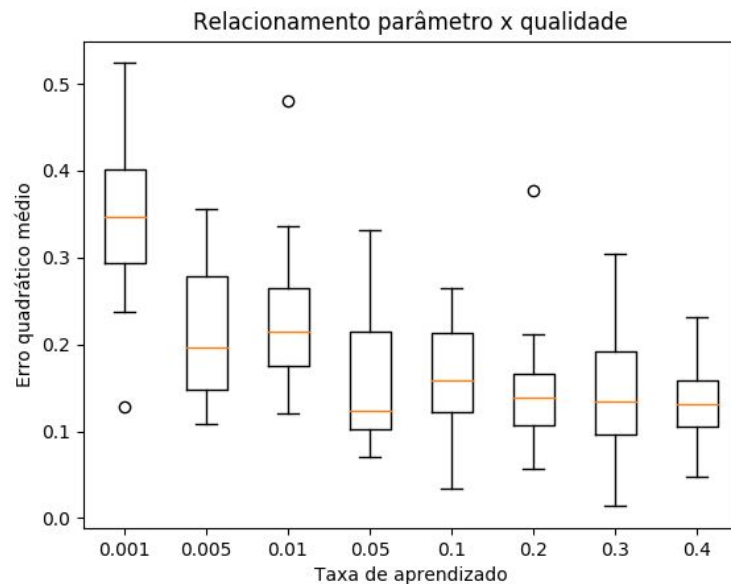
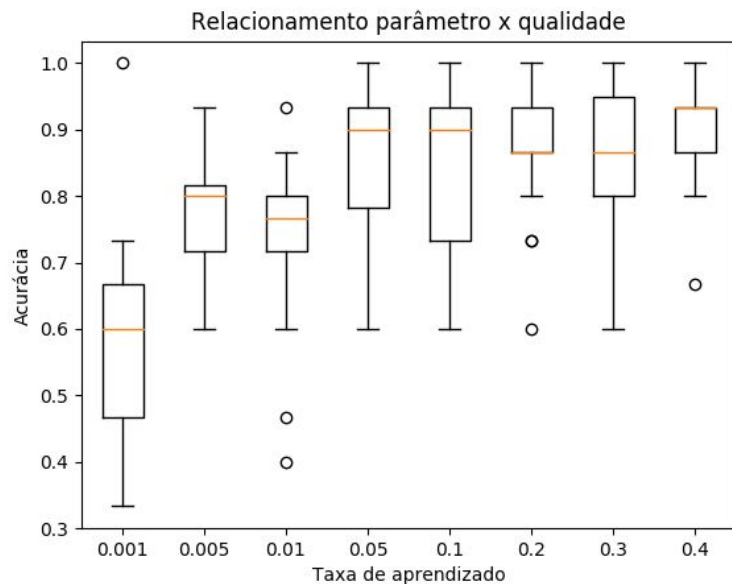
Resultados dos experimentos em relação à **função de ativação na camada oculta**.



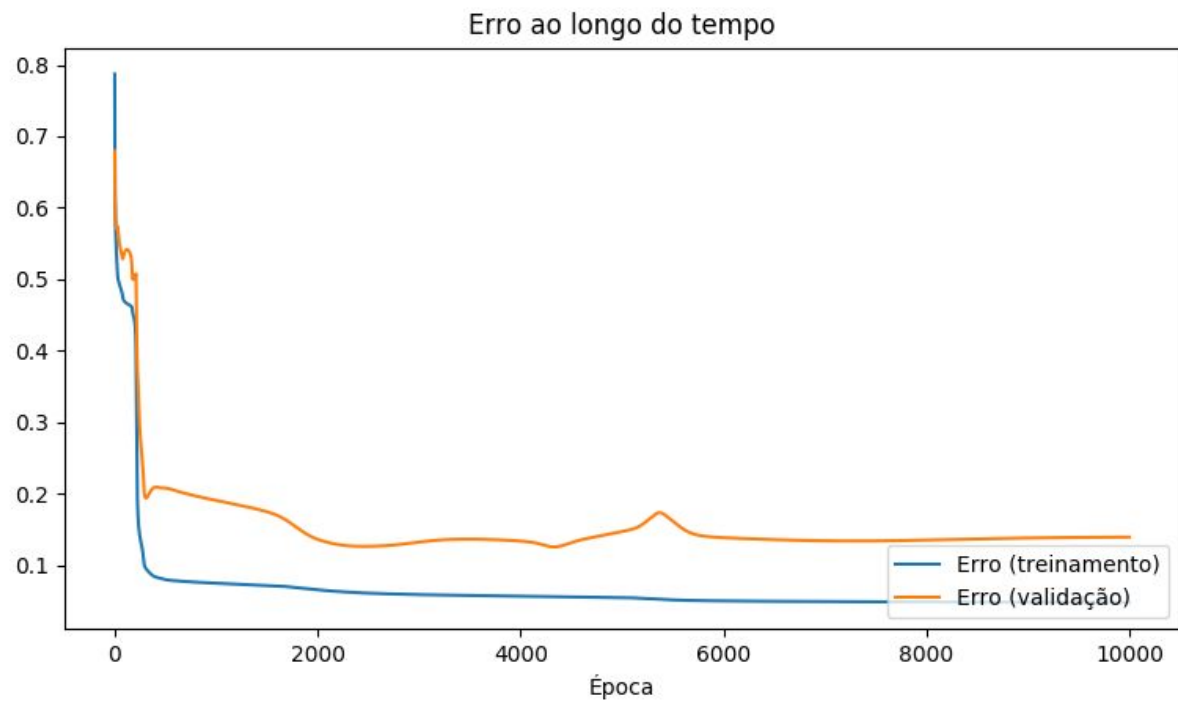
Resultados dos experimentos em relação à **função de ativação na camada de saída**, com a função de tangente hiperbólica na camada oculta.



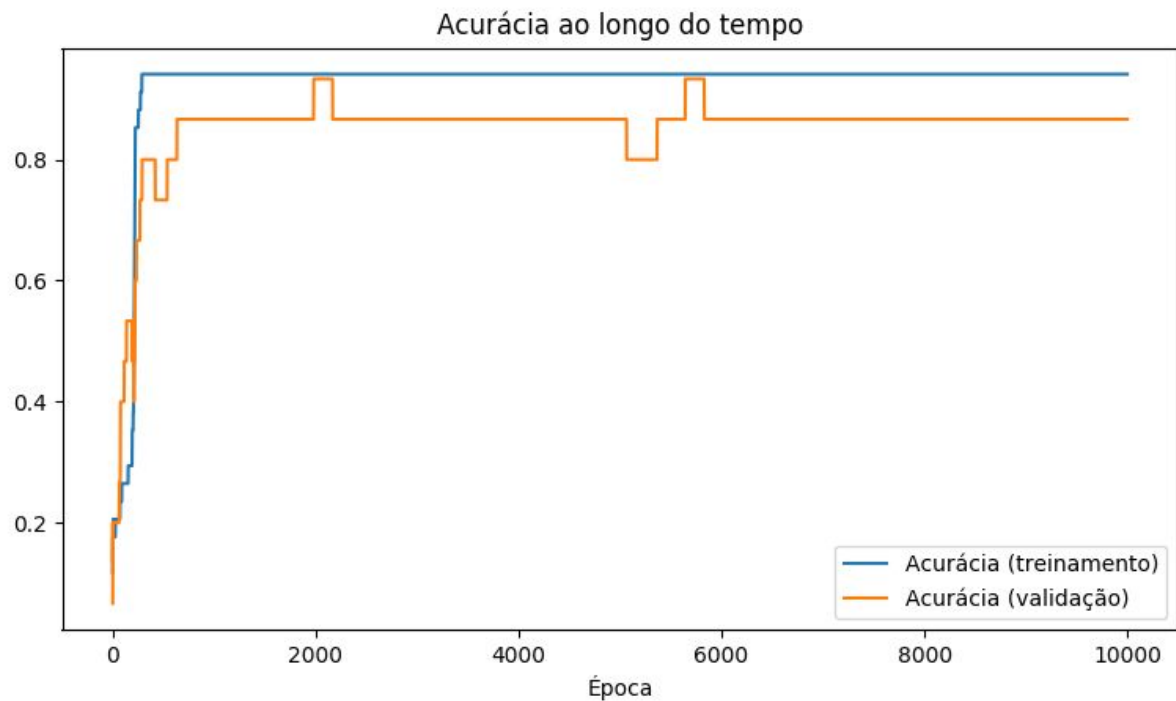
Resultados dos experimentos em relação ao **número de neurônios na camada oculta**, com a função de tangente hiperbólica na camada oculta e sigmóide na camada de saída.



Resultados dos experimentos em relação à **taxa de aprendizado**, com a função de tangente hiperbólica na camada oculta, sigmóide na camada de saída e 50 neurônios na camada oculta.



Evolução do erro quadrático médio ao longo das épocas de treinamento, com a configuração ótima.



Evolução da acurácia ao longo das épocas de treinamento, com a configuração ótima.





# Conclusões

# Considerações finais

- Realizou-se um conjunto de experimentos para determinar a melhor configuração da rede neural.
- Verifica-se pouca diferença em acurácia entre as funções de ativação na camada oculta, porém a sigmóide se destaca na camada de saída.
- Observa-se que quanto maior a taxa de aprendizagem e número de neurônios, mais alta será a acurácia (entre os valores testados).
- Cerca de 2000 épocas foram necessárias para atingir a maior acurácia.