

IA e ML Aplicados a Finanças

Prof. Leandro Maciel

AULA 6: Redes Neurais Artificiais

- 1 Redes Neurais Artificiais
- 2 Topologia e Treinamento
- 3 Previsão com RNAs
- 4 Redes LSTM
- 5 Bibliografia

1. Redes Neurais Artificiais

- Modelos de previsão de séries temporais estado-da-arte:

Econometria de séries temporais (ARMA);

Métodos de suavização exponencial (decomposição).

- Limitações:

Pressupostos acerca da dinâmica das séries (estacionariedade);

Estrutura linear para problemas complexos.

Redes Neurais Artificiais (RNAs):

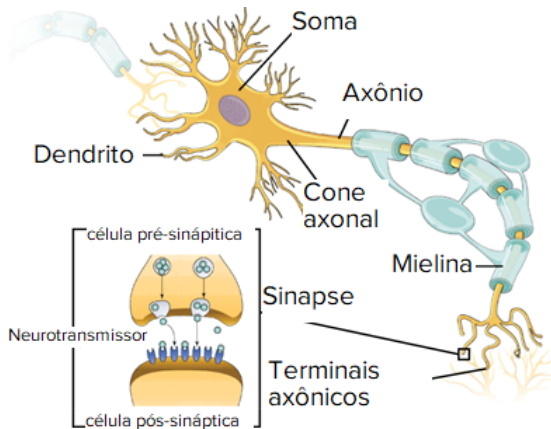
- McCulloch e Pitts (1943), Hebb (1949), e Roseblatt (1958);
- Modelo inspirado na estrutura neural de organismos inteligentes;
- *Machine Learning* → aprendizado com base em dados;
- Rede neural biológica → bilhões de unidades de processamento;
- Aquisição de conhecimento por experiência (aprendizado supervisionado).

Rede neural como modelo *black-box* do tipo entrada-saída:

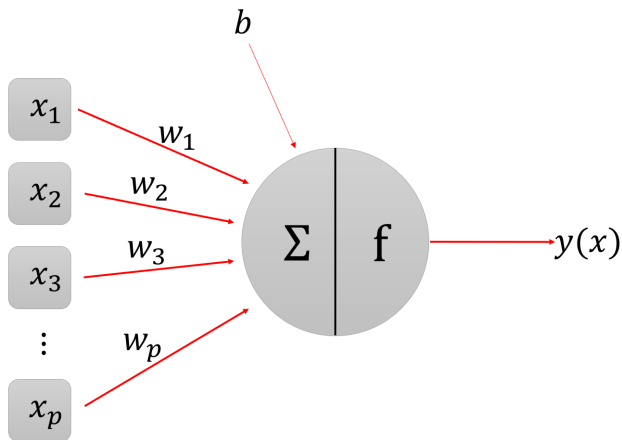
- Entrada → sinais de estímulo;
- Saída → resposta ao estímulo;
- Rede de neurônios → sistema complexo de processamento de estímulos;
- Interconexão permite grande nível de aprendizagem (especialização).

1. Redes Neurais Artificiais

■ Neurônio biológico:



- Neurônio artificial (modelo matemático):



- Entradas (x) \rightarrow dendritos - estímulos do ambiente;
- Pesos (w) \rightarrow contribuição/relevância de cada entrada;
- Neurônio soma sinais: $b + w_1x_1 + w_2x_2 + \dots + w_px_p = b + \sum_{i=1}^p w_ix_i$;
- b é o bias (*proxy* de um 'intercepto');
- Sinal de saída $y(x)$ processado por uma função de ativação $f(\cdot)$:

$$y(x) = f\left(b + \sum_{i=1}^p w_ix_i\right)$$

- União de vários neurônios artificiais → RNA:
- Inúmeros são os modelos de RNAs;
- Especialista deve determinar:

Topologia (estrutura) e **treinamento** (definição dos pesos);

- Método **não-linear** para processamento dados → **previsão**.

2. Topologia e Treinamento

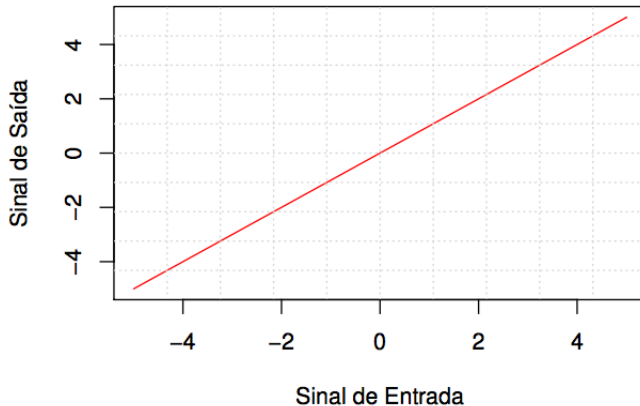
- Principais componentes de um modelo de RNA:
 1. Função de ativação;
 2. Topologia (arquitetura) - número camadas, número neurônios;
 3. Algoritmo de treinamento (aprendizagem) - como definir os pesos?

Função de ativação:

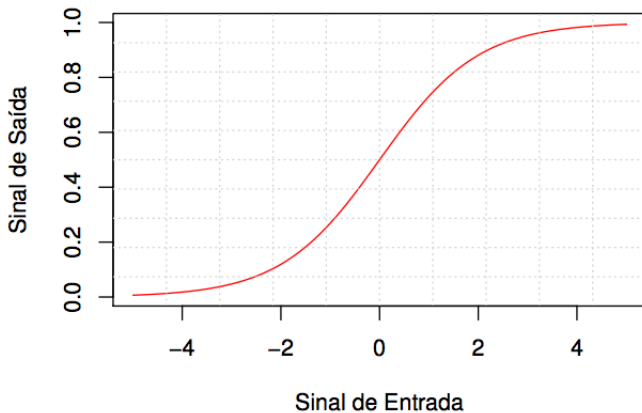
- Mecanismo de transmissão (filtragem) da informação na rede;
- Biologia: **função de ativação com limiar**;
- Algumas regiões da rede neural são ativadas e outras não;
- Regiões analíticas, emocionais, sensitivas, etc;
- Estímulos são repassados apenas para as regiões relevantes.

2. Topologia e Treinamento

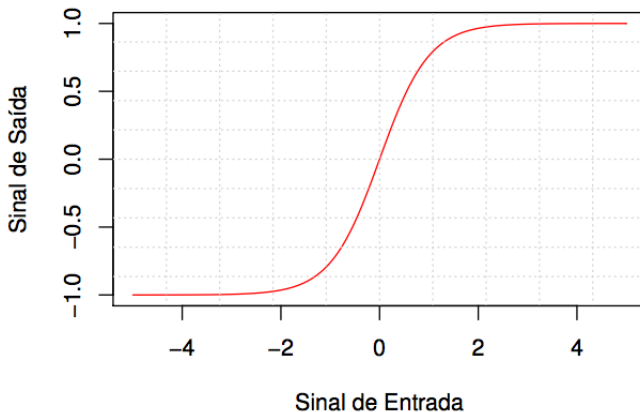
- Função de ativação linear: $f(x) = x$



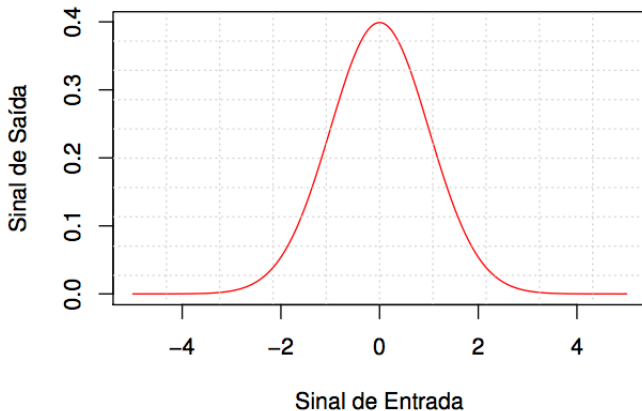
- Função de ativação logística: $f(x) = \frac{1}{1 + e^{-x}}$



- Função de ativação tangente hiperbólica: $f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$



- Função de ativação Gaussiana: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$

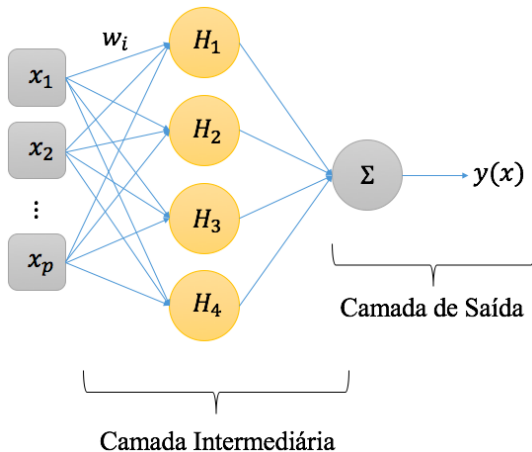


- Funções de ativação são sensíveis a escala dos dados;
- RNAs requerem a **normalização** dos dados de entrada/saída:

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad x_{norm} \in [0, 1] \quad \rightarrow \quad \text{min-max}$$

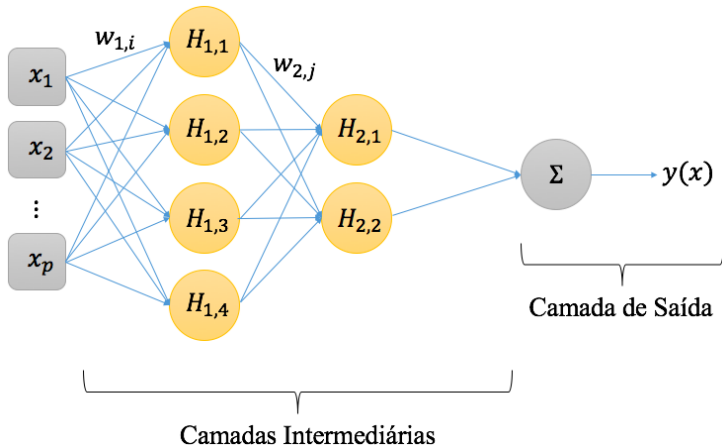
- Capacidade de aprendizagem depende da **topologia** da rede:
 - Número de camadas;
 - Número de neurônios em cada camada;
 - Presença ou não de retroalimentação da informação (sinais);
- Não há regras para determinação da topologia (*problem-based*);
- Topologia \times complexidade \times custo computacional \times desempenho.

- Rede com 1 camada com 4 neurônios:



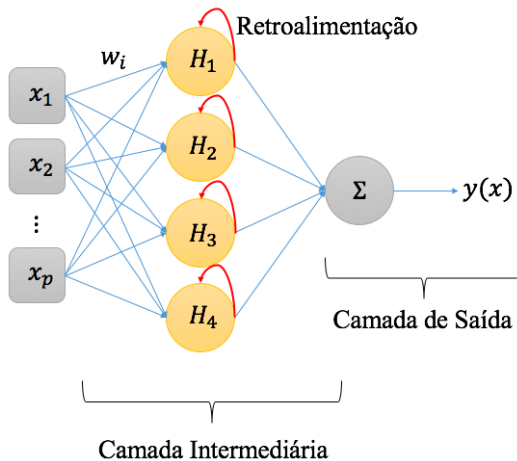
2. Topologia e Treinamento

- Rede com 2 camadas intermediárias com 4 e 2 neurônios:



2. Topologia e Treinamento

- Rede com 1 camada, 4 neurônios e retroalimentação (*delay*):



- **Treinamento** → determinar os pesos (parâmetros) da rede;
- Algoritmo **backpropagation** (otimização):

Método iterativo de ciclos (épocas) em dois processos;

Fase *forward* → saídas calculadas com inicialização aleatória pesos;

Fase *backward* → erro calculado e retro-propagado na rede;

Minimiza erro baseado no gradiente com uma **taxa de aprendizagem**.

- Vantagens e desvantagens do algoritmo backpropagation:

Elevado ajuste para problemas complexos e não lineares;

Não requer hipóteses acerca da dinâmica dos dados;

Computacionalmente intensivo e suscetível a **overfitting**;

Resultados (pesos) não interpretáveis (*black-box*);

Sensível a inicialização dos pesos (**Deep-Learning ANN**).

3. Previsão com RNAs

- IPCA mensal (Aula 5);

- Inicialização:

Normalizar dados (min-max);

Dividir amostras entre treinamento e teste.

- Definição do modelo de previsão:

$$\text{IPCA}_t = f_{\text{RNA}}(\text{IPCA}_{t-1}, \text{IPCA}_{t-2})$$

4. Redes LSTM

■ Redes neurais recorrentes:

- retroalimentação (memória);
- mais pesos, mais complexa identificação.

■ Redes profundas (*deep learning*):

- mais camadas e interconexões;
- maior capacidade processamento relações complexas;
- problemas de identificação.

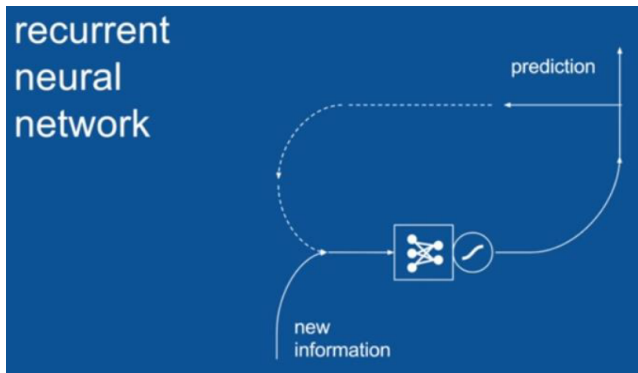
■ Séries temporais → sistemas dinâmicos:

- estados do sistema variam no tempo;
- trajetória define mudanças do estado;
- dependente no tempo (**memória**).

■ Redes *Long Short Term Memory* (LSTM):

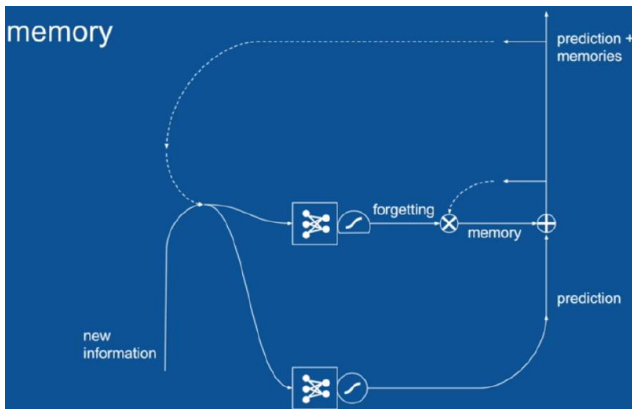
- HOCHREITER & SCHMIDHUBER em 1997;
- memórias de curto e longo prazos em redes recorrentes;
- memórias associadas ao estado interno da rede.

- Ideia básica das redes recorrentes:



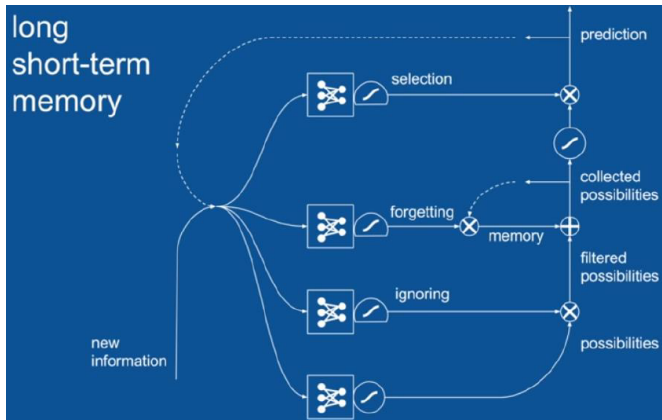
4. Redes LSTM

- Incorporação de memória:

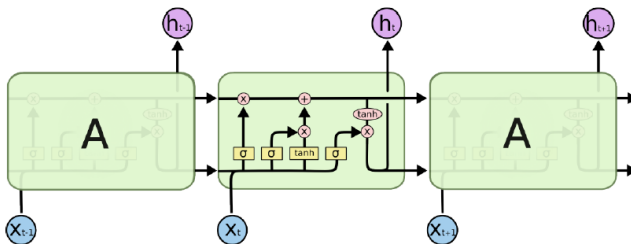


4. Redes LSTM

■ Redes LSTM:



■ Representação LSTM:



- Linha superior é o vetor de estados;
- Porta 1 define informação a ser preservada;
- Porta 2 processa nova informação e atualiza estado;
- Porta 3 calcula saída atualizada.

- Propriedades redes LSTM:
 - estrutura sofisticada;
 - mapeamentos multidimensionais não-lineares;
 - gates implementam filtros e memórias de prazo arbitrário;
 - adequado para processar informações sequenciais;
 - resultados promissores em previsão de séries temporais.
- Pacotes LSTM em R: keras e tensorflow, e TSLSTM.

■ Próxima aula...

- problemas de classificação;
- modelo de regressão logística...

HASTIE, Trevor, TIBSHIRANI, R., & FRIEDMAN, Jerome. **The Elements of Statistical Learning**. Data Mining, Inference, and Prediction. 2 Ed. Springer, 2008. **Capítulo 11**.

LANTZ, Brett. **Machine Learning with R**. Expert techniques for predictive modeling. 3^a Ed. Birmingham, UK: Packt Publishing Ltd. **Capítulo 7**.

HOCHREITER, S., SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, vol. 9, no. 8, pp. 1735-1780, 1997.

Prof. Leandro Maciel

leandromaciel@usp.br