

Inteligência Artificial Aplicada à Física

Fundamentos e Aplicações com Redes Neurais

AULA 3

Alexandre Suaide

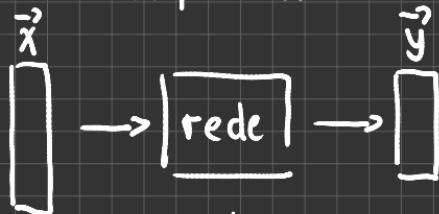
Escola Jayme Tiomno
IFUSP 2025

Objetivos para hoje

- Explorar mais a fundo algumas questões sobre o treinamento de redes
 - Overfitting e como lidar com isso
- Redes convolucionais
- Exemplo na física: modelo de Ising

Treinamento de uma rede

- Obter pesos e biases, w, b que transformem input em output minimizando custo com dados treino

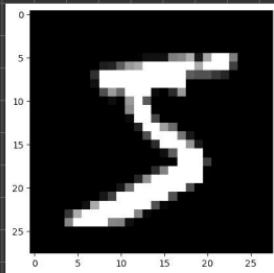


↳ Função de custo

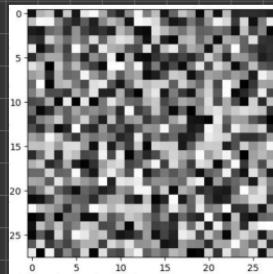
Em geral uma rede apenas transforma $\vec{x} \rightarrow \vec{y}$

- A rede treinada é capaz de generalizar seus resultados?
dado um \vec{x} fora do conjunto de treino, a rede é capaz de obter \vec{y} que faça sentido?

Ex: reconhecimento de números (base MNIST)



→ Rede → 5



treino
→ Rede →

0
1
2
3
4
5
6
7
8
9

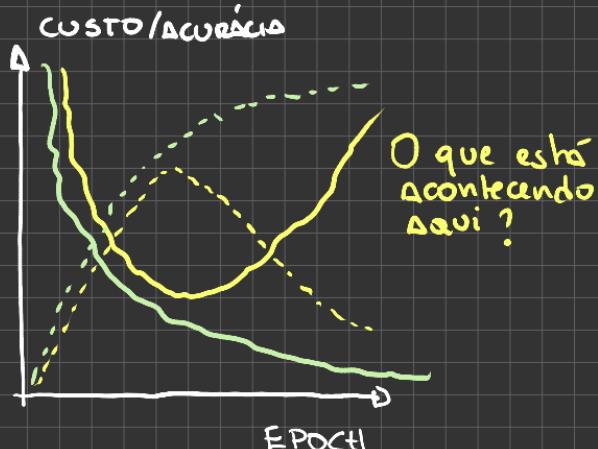
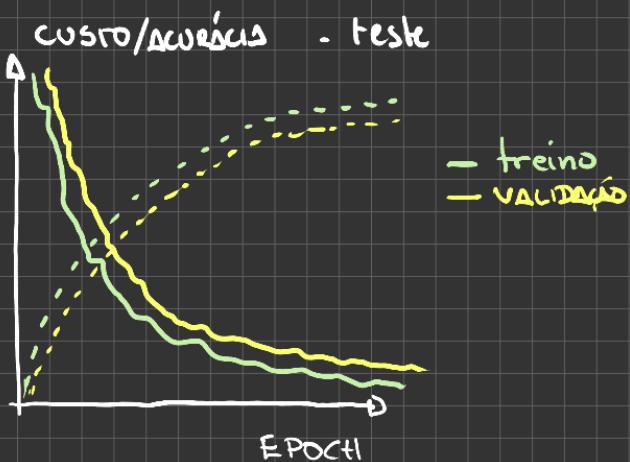
Como? △ rede
nunca foi exposta
à essa informação

→ Rede → 5

Uma rede está apenas transformando um vetor de dados em outro.
Elas não estão necessariamente generalizando ideias

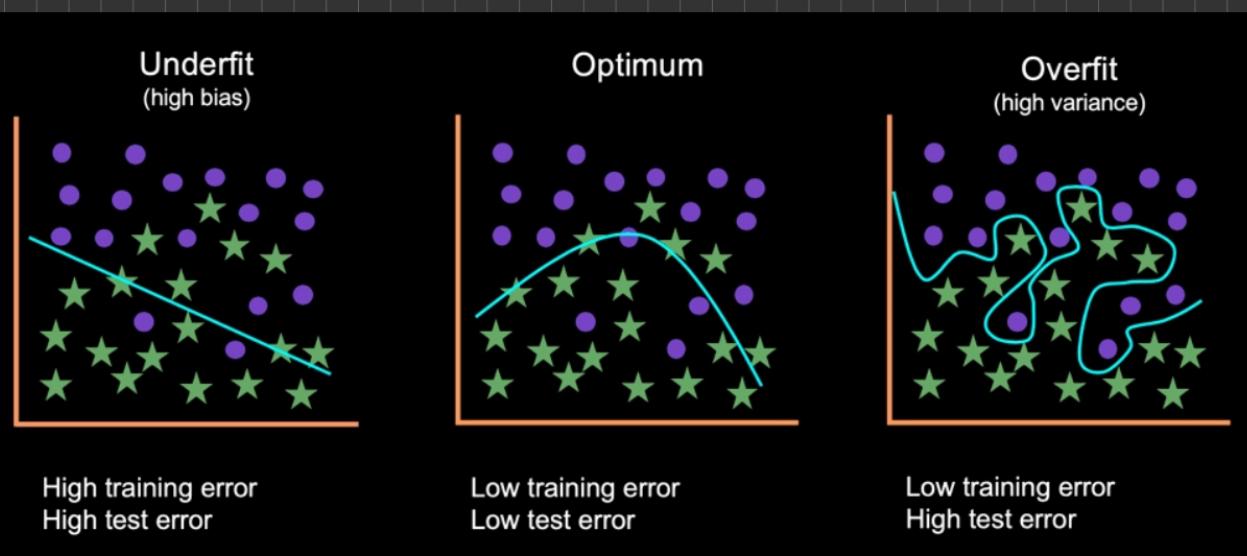
O processo de treinamento

- Optimização baseada em minimizar o custo
 - Mais lemos as métricas de acompanhamento
 - ↳ Acurácia, precisão etc
 - Dados são divididos em subgrupos
 - treino
 - VALIDAÇÃO
- Usados durante o treino



Overtaining

- Poucos dados de treino \rightarrow não representativos do todo
- Modelos muito complexos
- Excesso de treino \rightarrow especialização dos w, b

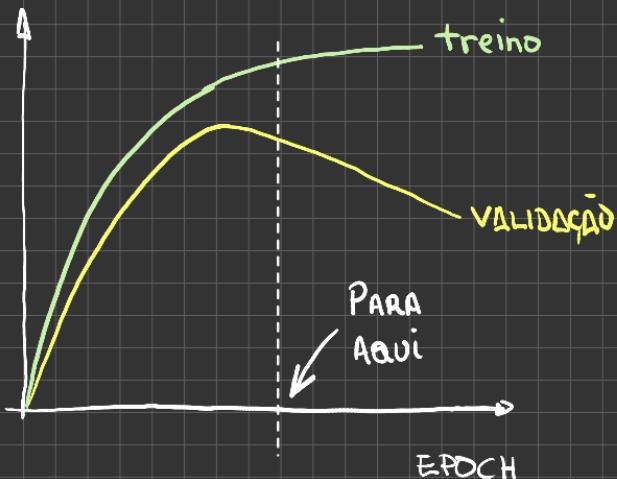


Há várias técnicas para evitar overtraining

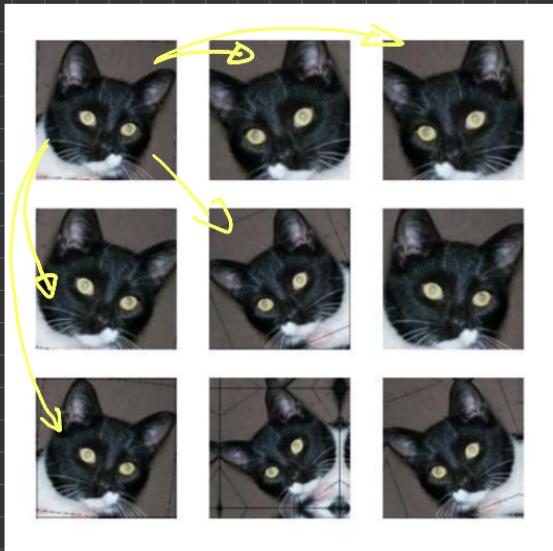
- Data augmentation
- Regularização
- Early stop
- Dropout

Early stop

Acurácia ou custo



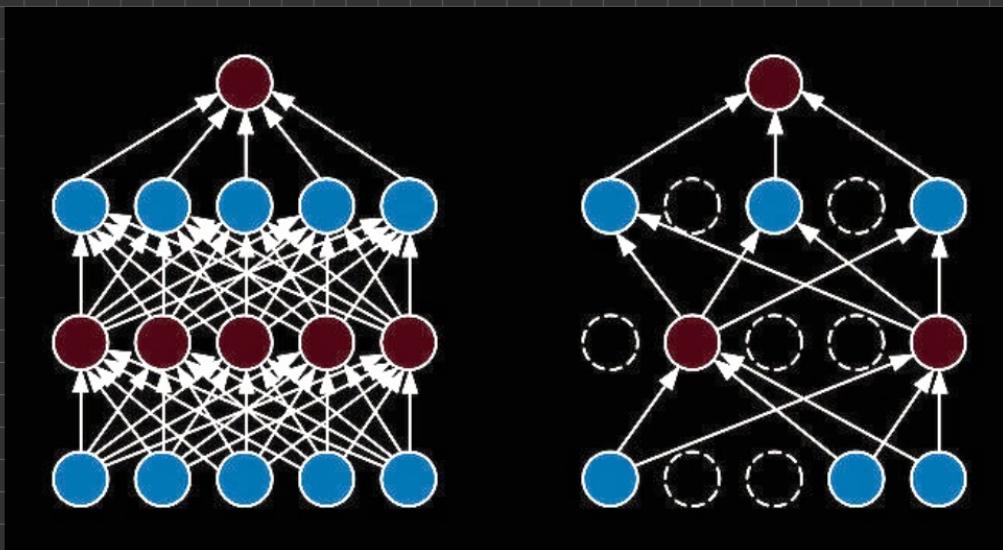
Data AUGMENTATION



Dropout

REMOVE, A CADA EPOCH, UMA FRACÃO ALEATÓRIA DE CONEXÕES
PARA O TREINAMENTO

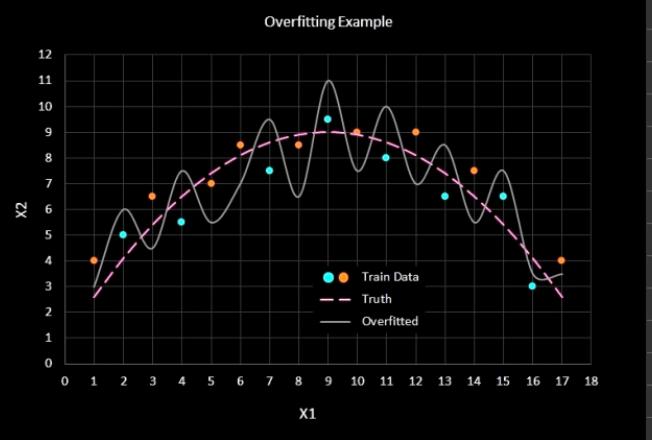
↳ AUMENTA A ROBUSTEZ DOS PESOS



REGULARIZAÇÃO

Adicionar à função de custo um termo que aumente o custo caso os valores dos pesos comecem a aumentar, evitando o favorecimento de algumas conexões sobre as demais. Isso diminui a tendenciosidade do treino

$$C(w) \rightarrow C(w) + \lambda R(w)$$



tipo L1 : $R(w) = \sum_{i,j} |w_{ij}|$

tipo L2 : $R(w) = \sum_{i,j} w_{ij}^2$

"EM relação à hipóteses que concorrem entre si, a mais simples é melhor"

William of Ockham 1285-1347

Há várias estruturas de redes

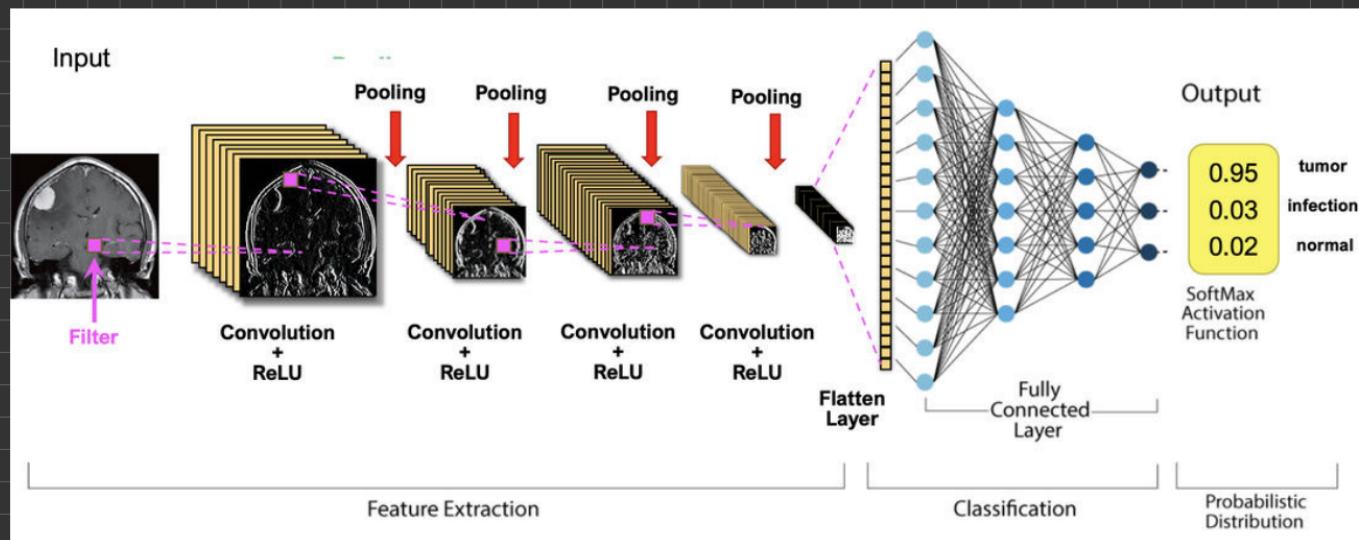
- Redes de MLP (MULTI-LAYER PERCEPTRON) são limitadas
 - Como tratar imagens, linguagem, séries temporais, etc
- Há VÁRIAS ESTRUTURAS DE REDES
 - Redes convolucionais
 - Redes recorrentes com MEMÓRIA
 - Mecanismos de ATENÇÃO (transformers, GNN)
 - Redes informadas por física
 - Redes fluidas, etc



Modelos se tornam complexos → CUIDADO COM OVERTRAINING!

Redes CONVOLUCIONAIS - CNN

CNN é um tipo de rede na qual há camadas de convolução em relação a filtros (kernel) que podem ser usados para buscar por estruturas (bordas, padrões geométricos, etc)



Estrutura de uma CNN

1. CAMADAS DE CONVOLUÇÃO

Usadas para detectar características, padrões nos dados

2 - Pooling

REDUÇÃO DO NÚMERO DE PARÂMETROS

3. Fully CONNECTED LAYERS

Realiza a regressão, classificação dos dados

Convolução

Sejam duas funções $f(x)$ e $g(x)$. A convolução entre elas é definida como:

$$h(t) = (f \otimes g)(t) = \int f(x) g(t-x) dx$$

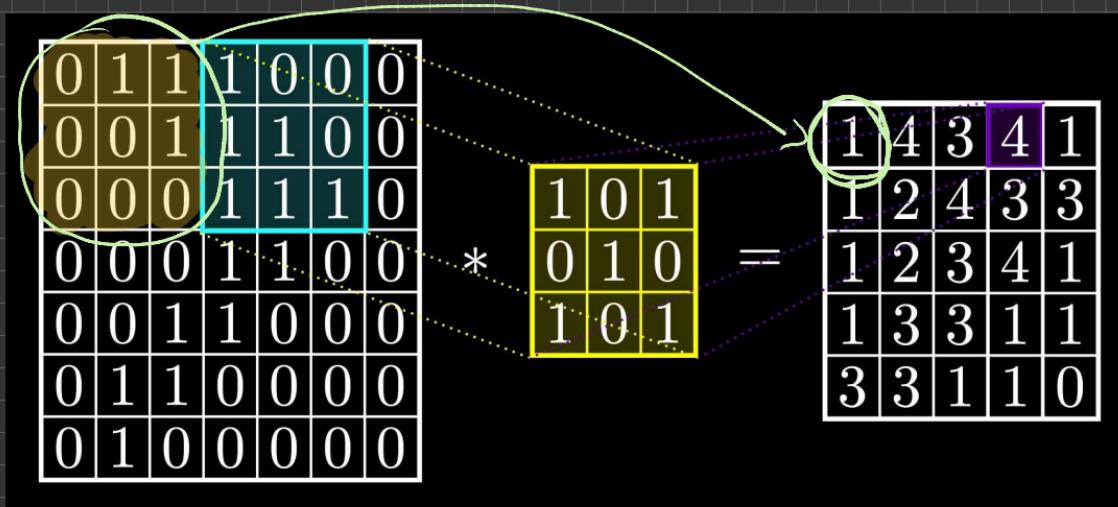
Ou seja, para cada t faz-se a soma sobre todos valores possíveis de x em $f(x)$ e $g(t-x)$

Na prática? \rightarrow Vamos ver como fazemos em matrizes
(x discreto)

Convolução em Matrizes

Seja uma imagem representada por $f(x,y)$ e um kernel, também na forma de matriz $K(i,j)$. A convolução entre essas duas matrizes é definida por

$$h(x,y) = \sum_i \sum_j K(i,j) f(x-i, y-j)$$



KERNELS

SÃO PEQUENAS MATRIZES
QUE EVIDENCIAM/TRANSFORMAM
ESTRUTURAS



Original	Gaussian Blur	Sharpen	Edge Detection
$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$
			

Pooling

CAMADAS DE POOLING SÃO USADAS PARA REDUZIR DIMENSÃO DA DADOS

- REDUZIR tempo de processamento
- EVIDENCIAR ESTRUTURAS DAS CAMADAS DE CONVOLUÇÃO



Atividade de hoje

Construir CNN p/ classificar transições de fase
ferromagnético \leftrightarrow paramagnético
 \uparrow
MODELO DE ISING

