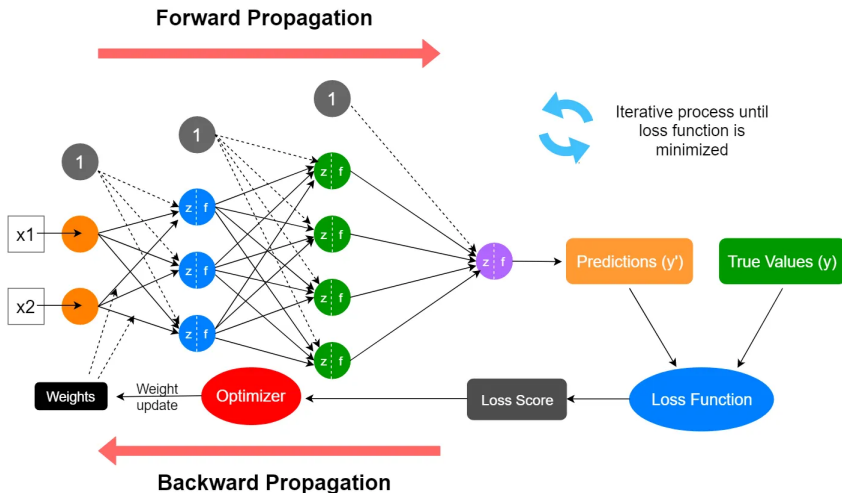


Introdução ao Aprendizado de Máquina

Cesar de Oliveira F. Silva

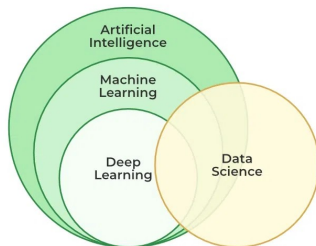
Setembro / 2026

Introdução ao Aprendizado de Máquina



IA, Aprendizado de Máquina e Deep Learning

- **Inteligência Artificial (IA):**
sistemas que executam tarefas que exigem inteligência humana.
- **Aprendizado de Máquina (ML):**
métodos estatísticos que aprendem a partir de dados.
- **Aprendizado Profundo (DL):**
subcampo de ML baseado em **redes neurais profundas**.
- **Ciência de Dados:** ciclo completo de coleta, preparação, modelagem e comunicação.



Subconjuntos: $DL \subset ML \subset IA$; interseção com Ciência de Dados.

Onde entra o Aprendizado de Máquina?

- **Inteligência Artificial (IA):** área que busca criar sistemas capazes de executar tarefas que, se feitas por humanos, dependeriam de inteligência.
- **Aprendizado de Máquina (Machine Learning, ML):**
 - Subárea da IA;
 - Foca em algoritmos que **aprendem padrões a partir de dados**.
- Ideia central (Mitchell): um programa *aprende* se seu desempenho em uma tarefa melhora com a experiência.

É tipo assim...

Pense em ML como treinar um estagiário: você mostra exemplos (**dados**), corrige seus erros (**função de perda**) e, com prática, ele melhora (**generalização**).

Exemplos de problemas de ML

- Classificar e-mails em **spam** ou **não spam**.
- Prever a **vazão de um rio** amanhã a partir da série histórica.
- Agrupar pixels de uma imagem de satélite em tipos de uso da terra.
- Decidir **quando abrir comportas** de um reservatório para minimizar risco de cheias.

Padrão comum

Em todos os casos, temos:

- Dados de entrada $\Rightarrow$ **modelo** $\Rightarrow$ decisão ou previsão.

Paradigmas de Aprendizado de Máquina

- **Supervisionado**

- Temos pares entrada-saída ($\mathbf{x}, y$).
- Ex.: prever turbidez a partir de chuva, vazão, temperatura.

- **Não supervisionado**

- Só temos as entradas $\mathbf{x}$ (sem rótulos y).
- Ex.: agrupar estações com perfis de qualidade de água semelhantes.

- **Semi-supervisionado**

- Poucos dados rotulados, muitos sem rótulo.
- Ex.: poucas amostras de laboratório + muitas imagens de satélite.

- **Aprendizado por reforço**

- Um agente aprende por **tentativa e erro**, recebendo recompensas.
- Ex.: controlar operação de um reservatório para maximizar segurança hídrica.

Analogias

- **Supervisionado:**
 - Professor corrige todas as listas de exercício;
 - Você aprende com gabarito completo.
- **Não supervisionado:**
 - Você entra numa festa sem conhecer ninguém e tenta descobrir grupos de afinidade “pelo papo”.
- **Semi-supervisionado:**
 - O professor resolve poucos exemplos na lousa e pede que você se vire com o resto.
- **Reforço:**
 - Adestrar um cachorro: comportamentos bons ganham petisco, comportamentos ruins, não.

Tarefas típicas de ML

- **Regressão:**
 - Saída numérica contínua.
 - Ex.: prever concentração de nitrato (mg/L).
- **Classificação:**
 - Saída categórica (classes).
 - Ex.: água “própria” ou “imprópria” para banho.
- **Clusterização:**
 - Descobrir grupos em dados sem rótulo.
 - Ex.: tipologias de bacias hidrográficas.
- **Redução de dimensionalidade:**
 - Comprimir muitos atributos em poucos fatores.
 - Ex.: sintetizar dezenas de bandas espectrais em 2–3 componentes.
- **Tomada de decisão:**
 - Escolher ações (abrir/fechar válvula, ligar/desligar bomba) a partir das previsões.

Exemplo visual: regressão vs classificação

Regressão

- Queremos aproximar uma função:

$$y \approx f(\mathbf{x})$$

- Ex.: curva que relaciona chuva acumulada e vazão.

Classificação

- Queremos estimar probabilidades de classe:

$$\hat{p}(y = k \mid \mathbf{x})$$

- Ex.: probabilidade de ocorrência de bloom de algas.

Metáfora

Regressão é *desenhar uma curva* que melhor passa pelos pontos; classificação é *desenhar fronteiras* que separam tipos de pontos.

Redução de dimensionalidade

- Muitos problemas ambientais têm **dezenas ou centenas de variáveis**: bandas de satélite, índices, variáveis climáticas, etc.
- Redução de dimensionalidade busca um mapeamento

$$\mathbf{z} = g(\mathbf{x}), \quad \mathbf{z} \in \mathbb{R}^p, \quad p \ll d,$$

preservando ao máximo a informação relevante.

- Exemplos:
 - PCA, autoencoders.

Metáfora

Como comprimir um relatório de 50 páginas em 1 página de resumo: menos dimensões, mas as ideias principais continuam lá.

Aprendizado de conceitos

- Em ML, um **conceito** é uma regra que separa exemplos positivos e negativos.
 - Ex.: “água imprópria para banho” = concentração de E. coli > 800.
- Formalmente, um conceito pode ser visto como uma função:

$$c : \mathcal{X} \rightarrow \{0, 1\},$$

que diz se um exemplo $\mathbf{x}$ pertence (1) ou não (0) ao conceito.

- O objetivo do algoritmo é **descobrir** uma boa aproximação h desse conceito verdadeiro, usando exemplos rotulados.

Metáfora

Conceito = “regra do jogo” que você quer descobrir observando poucas partidas.

Espaço de hipóteses

- O algoritmo não pode inventar qualquer regra; ele escolhe h dentro de um **espaço de hipóteses** $\mathcal{H}$.
- $\mathcal{H}$ é o conjunto de funções candidatas:

$$\mathcal{H} = \{h_1, h_2, \dots\}, \quad h: \mathcal{X} \rightarrow \mathcal{Y}.$$

- Ex.: todas as retas em 2D, todas as árvores de decisão até certa profundidade, todas as redes neurais com certa arquitetura.

Metáfora

- Espaço de hipóteses = **cardápio de modelos** que o algoritmo pode escolher.
- Os dados dizem qual item do cardápio funciona melhor para explicar o mundo.

Aprendizado como busca em $\mathcal{H}$

- Temos um conjunto de treino

$$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^N.$$

- Definimos uma **função de perda** $\mathcal{L}(h, D)$ que mede quão bem uma hipótese h se ajusta aos dados.
- O objetivo do algoritmo é encontrar

$$h^* = \arg \min_{h \in \mathcal{H}} \mathcal{L}(h, D),$$

ou algo próximo disso.

Intuição

É como procurar, em um armário cheio de modelos, qual deles “veste” melhor os seus dados sem ficar apertado demais (overfitting) nem largo demais (underfitting).

Pipeline típico de um projeto de ML

- 1 **Entender o problema** e definir a tarefa (regressão? classificação?).
- 2 **Coletar e preparar** dados (limpeza, integração, seleção de atributos).
- 3 Escolher um **espaço de hipóteses** (tipo de modelo).
- 4 **Treinar** o modelo nos dados rotulados.
- 5 Avaliar com dados de teste / validação.
- 6 Implantar o modelo e monitorar.

Conexão com a próxima aula

A partir desse ponto, vamos mergulhar em um tipo específico de modelo de ML: **redes neurais artificiais**, que são a base do Aprendizado Profundo.