**Machine Learning & CHATBOT**

1 – Modelos de Aprendizado de Máquina

Uma imagem contendo Texto

O conteúdo gerado por IA pode estar incorreto.

**O Paradigma Fundamental: A Inversão do Aprendizado**

* **Programação Tradicional:** Você, o programador, estuda um problema, cria regras e lógicas (usando if, else, for, etc.) e as escreve em um programa. Você alimenta o programa com **dados** e ele, seguindo suas **regras**, lhe dá uma **resposta**.
  + *Fluxo:* Dados + Regras = Respostas
* **Machine Learning:** inverte a lógica. Em vez de criar as regras, você fornece ao computador os dados e as respostas correspondentes. O objetivo do algoritmo de ML é analisar tudo isso e descobrir as regras que conectam os dados às respostas. O resultado desse processo é um "modelo" treinado.
  + *Fluxo:* Dados + Respostas = Regras (O Modelo)

**Machine Learning é o processo de ensinar um computador a encontrar padrões em dados para que ele possa tomar decisões ou fazer previsões sobre novos dados que nunca viu antes.**

**Os Três Pilares do Machine Learning**

Quase todos os problemas de ML se encaixam em uma de três categorias principais. A escolha depende do tipo de dado que você tem e do problema que quer resolver.

**1. Aprendizado Supervisionado (Supervised Learning)**

É o tipo mais comum e intuitivo de ML.

* **O Conceito:** O algoritmo aprende a partir de um conjunto de dados que foi previamente rotulado por um humano. Cada exemplo nos dados tem uma "resposta correta" ou "gabarito" associado a ele.
* **A Analogia:** É como um estudante aprendendo com um professor. O professor mostra exemplos **(dados**) e dá as respostas corretas **(rótulos).** Após ver exemplos suficientes, o estudante **(o modelo)** consegue responder a novas perguntas sozinho.
* **O "Idioma":**
  + **Features (Características):** As variáveis de entrada, as "perguntas" (ex: peso de uma fruta, área de uma casa).
  + **Labels (Rótulos):** A variável de saída, a "resposta correta" (ex: "maçã", "preço da casa").
* "Com base nos dados passados, preveja um resultado para novos dados."
* **Divide-se em duas tarefas principais:**
  + **Classificação:** Prever uma categoria. A resposta é uma classe discreta.
    - *Exemplos:* Este e-mail é *spam* ou *não spam*? Esta transação é *fraudulenta* ou *legítima*? Este paciente tem a *doença X* ou *não*?
  + **Regressão:** Prever um valor numérico contínuo. A resposta é um número.
    - *Exemplos:* Qual será o *preço* desta casa? Quantos *clientes* entrarão na loja amanhã? Qual o tempo de entrega de uma pizza?

**2. Aprendizado Não Supervisionado (Unsupervised Learning)**

Aqui, o desafio é maior, pois não temos um gabarito.

* **O Conceito:** O algoritmo recebe um conjunto de dados não rotulados e sua tarefa é encontrar a estrutura, os padrões ou as relações ocultas dentro deles, por conta própria.
* **A Analogia:** detetive que recebe uma caixa cheia de evidências misturadas. Ele não sabe o que procurar, mas começa a agrupar itens similares: fotos com fotos, documentos com documentos, impressões digitais com impressões digitais. Ele cria a ordem a partir do caos.
* **O "Idioma":**
  + **Clusterização (Agrupamento):** A tarefa mais comum. O algoritmo agrupa os dados em "clusters", onde os itens dentro de um cluster são mais parecidos entre si do que com os itens de outros clusters.
  + **Associação:** Encontrar regras que descrevem grandes porções dos dados. O exemplo clássico é "clientes que compram pão também tendem a comprar leite".
  + **Redução de Dimensionalidade:** Simplificar os dados, reduzindo o número de variáveis (features) sem perder informações cruciais. Útil para visualização e para otimizar outros algoritmos de ML.
* **Exemplos:**
  + **Segmentação de Clientes:** Agrupar clientes com comportamentos de compra semelhantes para campanhas de marketing.
  + **Sistemas de Recomendação:** Agrupar usuários com gostos parecidos (ex: "pessoas que gostaram do filme A também gostaram do filme B").
  + **Detecção de Anomalias:** Identificar pontos de dados que não se encaixam em nenhum grupo, podendo ser fraudes ou defeitos de fabricação.

**3. Aprendizado por Reforço (Reinforcement Learning)**

Paradigma mais complexo dos outros dois. Ele não se baseia em um dataset estático.

* **Conceito:** Um agente (modelo) aprende a tomar ações em um ambiente para maximizar uma recompensa cumulativa. Ele aprende por tentativa e erro.
* **A Analogia:** É exatamente como adestrar um animal de estimação. Quando o cão executa a ação correta (ex: sentar), ele recebe um petisco (recompensa positiva). Quando faz algo errado, não recebe nada ou pode receber uma repreensão (recompensa negativa). Com o tempo, ele aprende a política de ações que maximiza os petiscos.
* **O "Idioma":**
  + **Agente:** A entidade que aprende e toma decisões.
  + **Ambiente:** O mundo (real ou simulado) com o qual o agente interage.
  + **Ação:** Uma decisão que o agente pode tomar.
  + **Estado:** A situação atual do ambiente.
  + **Recompensa:** O feedback (positivo ou negativo) que o ambiente dá ao agente após uma ação.
* **Exemplos:**
  + **Inteligência Artificial para Jogos:** Ensinar um agente a jogar xadrez, Go ou videogames em nível sobre-humano.
  + **Robótica:** Treinar um robô para andar, pegar objetos ou navegar em um ambiente desconhecido.
  + **Otimização de Sistemas:** Gerenciamento de tráfego, otimização de portfólios de investimento ou controle de sistemas de refrigeração em data centers.

**Conceitos Essenciais Adicionais**

* **Modelo:** É o artefato que resulta do processo de treinamento. É uma representação matemática dos padrões encontrados nos dados. É o "cérebro" treinado que você usa para fazer novas previsões.
* **Overfitting (Sobreajuste) vs. Underfitting (Subajuste):**
  + **Overfitting:** O modelo aprendeu os dados de treino *bem demais*, incluindo o ruído e os detalhes irrelevantes. Ele "decorou" as respostas em vez de aprender a regra geral. Ele se sai perfeitamente nos dados de treino, mas falha miseravelmente com novos dados.
  + **Underfitting:** O modelo é simples demais e não conseguiu capturar os padrões subjacentes nos dados de treino. Ele se sai mal tanto nos dados de treino quanto em novos dados.
  + **O Objetivo:** Encontrar um modelo que generaliza bem, ou seja, que tem um bom desempenho em dados que nunca viu antes.