

Aprendizado de Máquina

Prof. Raphael

Carvalho

Aprendizado de Máquina



Aprendizado de Máquina

- Aprendizado de Máquina (Machine Learning) é a ciência e a arte de **programar computadores** para que eles possam aprender com os dados.
- Definição um pouco mais formal diria que se trata do campo de estudo que dá aos computadores a capacidade de aprender sem ser explicitamente programado.
- Gosto mais dessa segunda definição porque ela traz uma comparação entre algoritmo de programação tradicional e algoritmo de aprendizado de máquina.

O que é a programação tradicional?



Aprendizado de Máquina

- Imaginem um software que recebe uma imagem



de entrada qualquer e
identifica se há um gato
nela ou não

Aprendizado de Máquina

- Uma forma de tentar resolver esse problema é por meio do aprendizado de máquina.
- Ocorre que funciona de uma maneira praticamente inversa à programação tradicional.
- Continuamos entrando com dados, mas – em vez de um programador criar manualmente as regras – são inseridos exemplos de resultados passados.

- A saída de um algoritmo de aprendizado de máquina são justamente as regras.

Aprendizado de Máquina



Aprendizado de Máquina

- No caso do exemplo identificador de gatos, continuaríamos inserindo imagens de entrada quaisquer, mas também seriam inseridos diversos exemplos de resultados (isto é, imagens que efetivamente contém gatos).
- A saída do algoritmo de aprendizado de máquina seria capaz, por si só, de realizar um mapeamento estatístico entre os dados de entrada e os exemplos de resultados esperados a fim descobrir se há ou não um gato em uma imagem.
- Em vez de um programador dizer quais são as regras, quem diz é o algoritmo

Etapas de Treinamento

- Algoritmo de aprendizado de máquina consegue extrair regras de identificação de gatos por meio de padrões estatísticos comuns entre os dados de entrada e os resultados esperados.
- Quanto mais exemplos de resultados você oferece, mais o algoritmo é treinado, mais regras são aprendidas e mais ajustado se torna o modelo.
- Essa etapa do processo de aprendizado de máquina é chamada de **Treinamento**.

Etapa de Treinamento

- Trata-se de uma etapa custosa porque idealmente nós

temos que inserir quantidades massivas de exemplos de resultados para que o modelo fique o mais ajustado possível.

Etapas de Inferência

- Ocorre quando utilizamos uma programação bem próxima à programação tradicional com regras aprendidas na etapa anterior e novos dados para gerar inferir resultados

Aprendizado de Máquina



Aprendizado de Máquina

MACHINE LEARNING

SUBCAMPO DA INTELIGÊNCIA ARTIFICIAL QUE FORNECE AOS COMPUTADORES A HABILIDADE DE APRENDER SEM SEREM EXPLICITAMENTE PROGRAMADOS

PROCESSO EM QUE O DESEMPENHO DE UMA TAREFA AUMENTA COM A EXPERIÊNCIA EXTRAÍDA DE NOVOS DADOS

A MÁQUINA PODE APRENDER A PARTIR DE SEUS ERROS E FAZER INFERÊNCIAS SOBRE DADOS

DIFERE-SE DA PROGRAMAÇÃO TRADICIONAL POR GERAR REGRAS A PARTIR DE RESULTADOS E, NÃO, O CONTRÁRIO

MODELOS SÃO REGRAS ESTATÍSTICAS NA FORMA DE MODELOS MATEMÁTICOS E PARÂMETROS

A ETAPA DE TREINAMENTO É MAIS CUSTOSA E A ETAPA DE INFERÊNCIA MENOS CUSTOSA

Aprendizado Supervisionado

- Trata-se de um conjunto de técnicas de aprendizado para treinar um modelo com dados rotulados manualmente
- Um especialista/supervisor externo diz qual é a saída esperada para cada dado histórico utilizado no treinamento.
- São as técnicas mais comumente empregadas no contexto de aprendizado de máquina e geralmente são utilizadas para identificar padrões específicos, prever resultados dado um conjunto de amostras de treinamento.

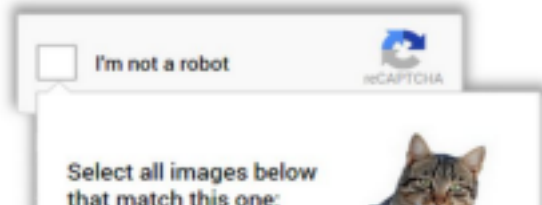
Aprendizado Supervisionado

- A saída desejada para cada exemplo de entrada já é conhecida no aprendizado supervisionado, isto é, os dados de saída são previamente rotulados.
- Essa abordagem – também chamada de tarefa de previsão – é bastante semelhante à aprendizagem humana sob a supervisão de um professor

Aprendizado Supervisionado



Aprendizado Supervisionado



- Sabe quando vamos acessar uma página web e aparece uma caixinha falando para nós selecionarmos imagens com algum objeto específico para indicar que não somos um robô?
- É reCAPTCHA!
- O objetivo principal desse teste não é verificar se somos robôs, mas o objetivo principal é treinar algoritmos de identificação de imagens

Aprendizado Supervisionado

- É uma abordagem de aprendizado de máquina em que um supervisor já conhece de antemão o resultado (rótulo/classe) e pode guiar o aprendizado mapeando as entradas em saídas por meio do ajuste de parâmetros em um modelo capaz de prever rótulos desconhecidos.
- Problemas de aprendizado supervisionado geralmente tratam de uma variável quantitativa ou qualitativa

Aprendizado Supervisionado

Regressão

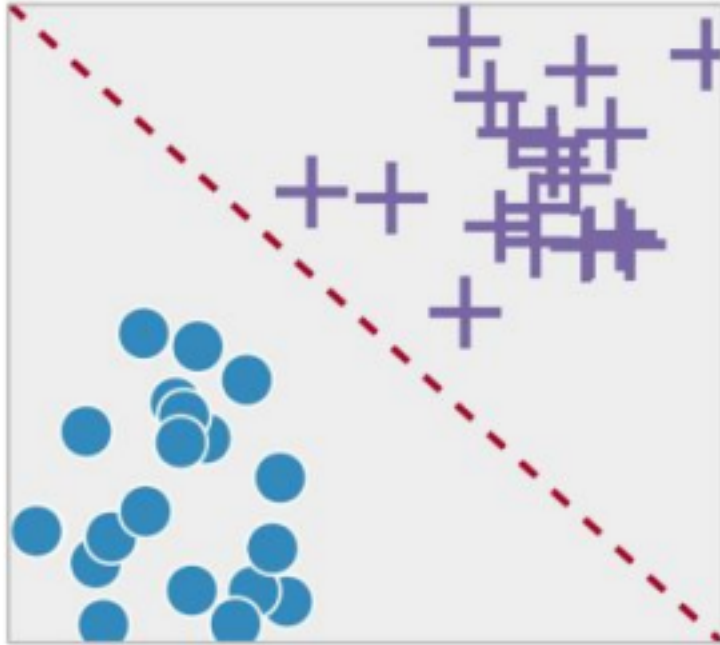
- Rótulos se referem a um conjunto infinito de valores numéricos contínuos

Classificação:

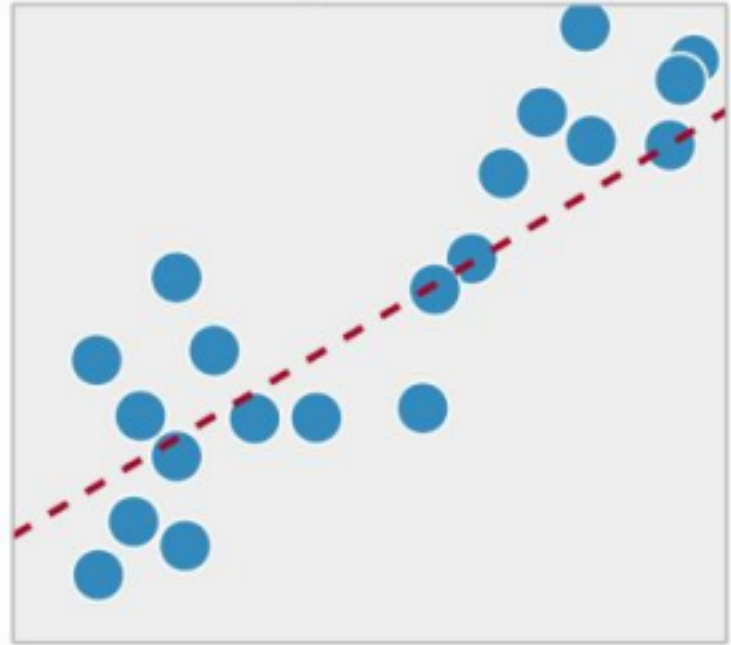
- Rótulos se referem a um conjunto finito e não ordenado de valores categóricos

Aprendizado Supervisionado

Classification



Regression



Resumindo

APRENDIZADO SUPERVISIONADO

TRATA-SE DA ABORDAGEM QUE BUSCA ENCONTRAR UM CONJUNTO DE REGRAS OU FUNÇÕES (TAMBÉM CHAMADAS DE MODELO) A PARTIR DOS DADOS DE TREINAMENTO QUE POSSAM SER UTILIZADAS PARA PREVER UM RÓTULO OU VALOR QUE CARACTERIZE UM NOVO EXEMPLO, COM BASE NOS VALORES DE SEUS ATRIBUTOS DE ENTRADA. O TERMO SUPERVISIONADO VEM DA PRESENÇA DE UM SUPERVISOR EXTERNO, QUE CONHECE A SAÍDA (RÓTULO) DESEJADA PARA CADA EXEMPLO. COM ISSO, ELE PODE AVALIAR A CAPACIDADE DO ALGORITMO DE PREDIZER O VALOR DE SAÍDA PARA NOVOS EXEMPLOS.

Aprendizado Não-Supervisionado

- Trata-se de um conjunto de técnicas para treinar um modelo em que não se sabe a saída esperada para cada dado usado no treinamento.

- Diferentemente do aprendizado supervisionado, você não utiliza rótulos/categorias para as amostras de treinamento.
- Os algoritmos são formulados de tal forma que podem encontrar estruturas e padrões adequados nos dados por conta própria

Aprendizado Não-Supervisionado

- O algoritmo identifica as semelhanças nos dados apresentados e reage com base na presença ou ausência dessas tais semelhanças.
- Essa abordagem – também chamada de tarefa de

descrição – permite que o algoritmo aprenda a categorizar dados autonomamente.

- A ideia aqui não é prever nada e, sim, organizar os dados de alguma forma ou descrever sua estrutura

Aprendizado Não-Supervisionado

- O processo de aprendizado não supervisionado é mais complexo porque não existe um supervisor para treinar o algoritmo e nem existem categorias pré-definidas.
- Logo, existe realmente uma chance de o algoritmo gerar categorias completamente diferentes do que você esperava.

- Existem dois grandes sub-grupos:
 - Agrupamentos (Clustering)
 - Regras de Associação (Association Rules)

Aprendizado Não-Supervisionado

Regras de Associação (Association Rules)

- Tipo de aprendizado não-supervisionado que permite estabelecer regras capazes de verificar como determinados elementos em um conjunto estão intimamente associados, isto é, se a presença de um elemento implica a presença de outro dentro em uma mesma transação.
- Os principais modelos são:
 - Apriori

- PCA (Principal Component Analysis)

Aprendizado Não-Supervisionado

Agrupamentos (Clustering)

- É um tipo de aprendizado não-supervisionado em que se busca encontrar padrões em um conjunto de dados e agrupá-los em subconjuntos que – ao comparar dados de um mesmo grupo – sejam o máximo possível homogêneos ou semelhantes e – ao comparar dados de grupos diferentes – sejam o máximo possível heterogêneos ou diferentes. ●

Principais modelos são:

- k-Means
- Agrupamento Hierárquico

Aprendizado Não-Supervisionado

APRENDIZADO NÃO-SUPERVISIONADO

TRATA-SE DA ABORDAGEM EM QUE O ALGORITMO BUSCA ENCONTRAR UM PADRÃO SUBJACENTE NOS DADOS SEM A UTILIZAÇÃO DE UM SUPERVISOR EXTERNO PARA ATRIBUIR RÓTULOS OU CATEGORIAS PRÉ-DEFINIDAS PARA AS AMOSTRAS DE TREINAMENTO. OS ALGORITMOS SÃO FORMULADOS DE FORMA QUE POSSAM ENCONTRAR PADRÕES AUTONOMAMENTE COM O INTUITO DE EXPLORAR DADOS DESCONHECIDOS E ENCONTRAR ESTRUTURAS INTERESSANTES OU OCULTAS NOS DADOS QUE NÃO ERAM VISÍVEIS ANTERIORMENTE PARA OS CIENTISTAS DE DADOS.

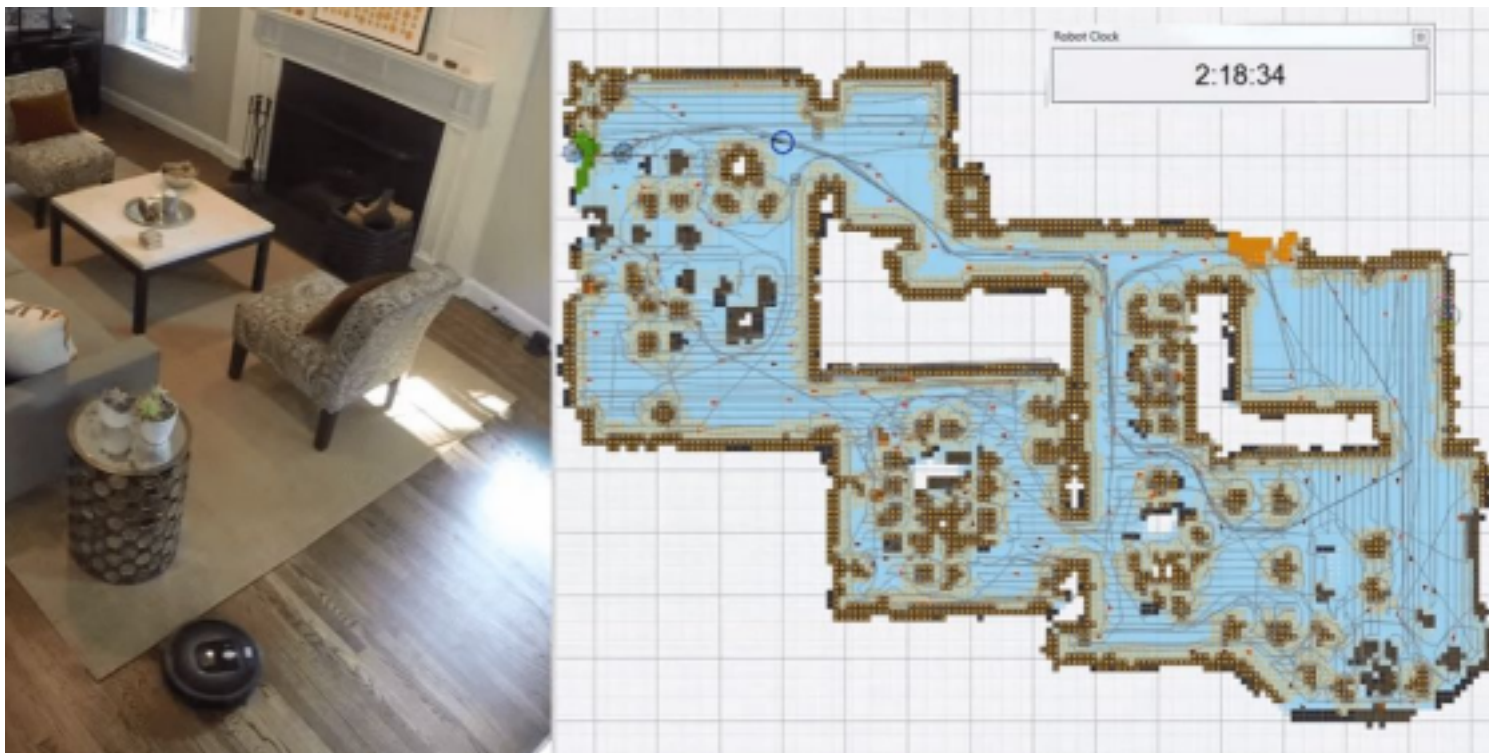
Aprendizado por Reforço

- Trata-se de um conjunto de técnicas que utilizam

tentativa e erro para descobrir decisões ótimas de como interagir com ambiente ou com outros agentes.

- Ele tem como meta reforçar ou recompensar uma ação considerada positiva e punir uma ação considerada negativa
- Exemplo: são os famosos robôs aspiradores

Aprendizado por Reforço



Aprendizado por Reforço

- Algoritmos de aprendizado por reforço baseiam-se em reforço positivo/negativo para otimização de resultado. ●
- No caso dos robôs aspiradores, eles punem a passagem por trechos pouco promissores e recompensam a passagem por trechos mais promissores