

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

CONTEXTO INICIAL

Olá! Que bom que chegou até aqui.

No texto passado vimos um pouco sobre Machine Learning e que os dois principais tipos de aprendizados, que são aprendizado **supervisionado** e aprendizado **não-supervisionado**.

Depois vimos que no aprendizado supervisionado temos modelos de classificação para variáveis categóricas e regressão linear para targets numéricos contínuos.

Falando em modelos não-supervisionados vimos como funcionam alguns algoritmos de clusterização e de redução de dimensionalidade.

Agora que você já viu como funciona, está na hora de ver tudo (um pouquinho na verdade) o que eles podem fazer!

LEBRANDO QUE SE VOCÊ NÃO VIU O TEXTO PASSADO "APRENDIZADO SUPERVISIONADO E NÃO SUPERVISIONADO" TE ACONSELHO A VOLTAR E LER ELE PRIMEIRO. SE VOCÊ JÁ LEU, SIGA EM FRENTE!

AVANÇAR

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

RESOLVENDO PROBLEMAS COM MACHINE LEARNING

Machine Learning não é um fim. É um meio para atingir um fim. Ou seja, mais importante que ter um modelo robusto e bem treinado é saber para se e como ele funciona. Um dos clássicos exemplos disso é que a Netflix. Em outubro de 2006, o CEO Reed Hastings levantou a hipótese de que se o Cinematch, algoritmo proprietário de recomendações de filmes para usuários, melhorasse em 10%, a empresa teria uma queda no churn de seus usuários e um aumento de receitas que poderiam chegar a até U\$89 milhões.

Então, ao invés de contratar e treinar funcionários, ele optou pelo formato crowdsourcing. The Netflix Prize, uma competição aberta para o melhor algoritmo de filtragem colaborativa para prever classificações de usuários para filmes, com base em classificações anteriores. O prêmio distribuiu Progress Prizes de 2007 a 2009, quando a empresa finalmente anunciou o time vencedor que atingiu um score de 10,09%. Entretanto o algoritmo nunca foi utilizado e o motivo apurado pelos veículos de tecnologia no período foi:

O AUMENTO DE PRECISÃO QUE AS IDEIAS VENCEDORAS PROPICIARIAM NÃO PARECIA JUSTIFICAR O ESFORÇO DE ENGENHARIA QUE SERIA NECESSÁRIO FAZER PARA TRAZÊ-LAS PARA O AMBIENTE DE PRODUÇÃO.

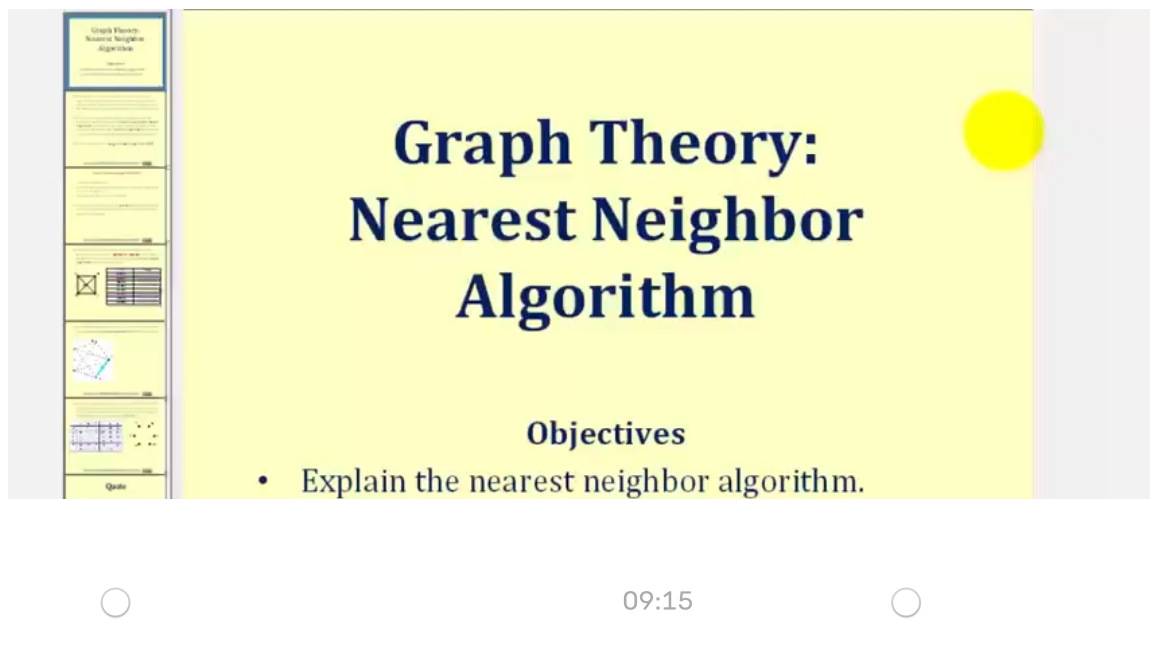
Em outras palavras, um modelo menos preciso, mas menos custoso pode ser mais útil do que um modelo mais preciso, mas que exige muita computação ou muito tempo de desenvolvimento, por exemplo.

[AVANÇAR](#)

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

RESOLVENDO PROBLEMAS DO PASSADO

O *Nearest Neighbor Algorithm* (NNA) de 1967 foi um dos primeiros algoritmos desenvolvidos para ajudar vendedores ambulantes. A sua intenção era resolver o problema da rota de vendas dessas pessoas. Basicamente, ele tinha a função de ajudar a desenhar a rota mais eficiente, passando por todas as cidades necessárias, sem retornar àquelas já visitadas. No exemplo abaixo, ele usa o modo greedy, ou seja, busca pelo melhor parâmetro a cada decisão, sem se importar com as decisões e parâmetros subsequentes (vídeo, 9 min. Em inglês).



TUDO CERTO ATÉ AQUI?

AVANÇAR

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

RESOLVENDO PROBLEMAS DO PRESENTE

PREVENDO O PREÇO DE UM IMÓVEL DE ACORDO COM SUAS CARACTERÍSTICAS

Imagine que você vai trocar de casa ou apartamento. Como prever o preço do imóvel que você deseja adquirir? É possível fazer isso através de **algoritmos supervisionados**, fazendo uma **regressão linear** para identificar o preço de um imóvel dadas as suas características. Mesmo que este imóvel não exista de fato. Isso pode ser usado por imobiliárias para descobrir qual o tipo de imóvel mais rentável em uma região, para avaliar o preço de um imóvel, para identificar ofertas muito abaixo do mercado ou negociar preços que estão acima da média.

PREVENDO ACIDENTES DE TRÂNSITO

Uma seguradora decide investigar um pouco mais sobre os acidentes de trânsito de seus segurados. A forma como se deseja modular o problema interfere diretamente na forma de capturar e tratar os dados e qual modelo deverá ser usado. A seguradora poderia investigar o número de acidentes de acordo com o período do ano e as condições climáticas por exemplo. Esse seria um problema para um modelo de **aprendizado supervisionado de regressão**. Caso a seguradora decidisse investigar se determinados horários, condições climáticas e rotas teriam acidentes ou não, estaríamos falando de um **modelo de classificação com aprendizado supervisionado**. Agora, se ao invés disso, ela decidisse investigar os seus segurados quanto ao modo de dirigir, os horários que ficam no trânsito, as rotas que costumam fazer e clusterizar esses perfis, estaríamos falando de um modelo de **aprendizado não supervisionado**.

DETECÇÃO DE FRAUDE NO CARTÃO DE CRÉDITO

Quando acontece uma compra indevida no cartão de crédito, existem várias possíveis abordagens de resolução de problemas. Com **algoritmos não supervisionados**, é possível analisar a última compra feita e a compará-la com todas as anteriores. Valores, localização, moeda, horário, local de compra, forma de pagamento e muitos outros fatores seus e de outras pessoas usuárias com o mesmo perfil podem ser avaliados para detectar

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

ou não, pode ser muito mais simples e eficiente computacionalmente e em termos de infraestrutura.

PREVENÇÃO DE MANUNTEÇÃO DE MÁQUINAS OU AVALIAÇÃO DA QUALIDADE DE UM PRODUTO

Embora pareçam tarefas muito diferentes, ambas possuem o mesmo desenho de problema. São problemas que buscam anomalias, seja no funcionamento das máquinas ou na consistência e qualidade de um produto em uma linha de montagem, por exemplo. Nesses casos, podemos usar tanto modelos **não supervisionados** quanto **modelos supervisionados**, desde que, neste último tenhamos etiquetas para treinar os modelos previamente.

TIME SÉRIES TEMPORAIS OU USO DO TEMPO COMO UMA VARIÁVEL PRINCIPAL PARA GERAR MODELOS E ANÁLISES

Uma análise e previsão de demanda e estoque podem ser cruciais para o sucesso de uma empresa de varejo ou e-commerce. Além disso, previsões nas oscilações na bolsas de valores, previsões de vendas de veículos, cervejas e até mesmo aumento ou queda do número de seguidores no Instagram ou da venda de passagens aéreas. Todos esses são problemas de séries temporais. Embora eles possam ser resolvidos com alguns algoritmos de [árvore de decisão](#), sua abordagem e modelagem é um pouco mais complexa. Isso porque séries temporais são sensíveis a vários fatores externos como sazonalidade, estacionalidade, auto correlação, entre outros.

MOTORES DE RECOMENDAÇÕES

Esta é uma das soluções de dados mais presentes no nosso cotidiano. Estamos falando do Cinematch do Netflix, do motor de recomendações do Spotify, dos algoritmos de sugestão de conteúdo do Youtube. Isso tudo apenas em entretenimento. Mas não pára por aqui. Temos o motor de recomendações da Amazon, que sugere produtos correlacionados ao que você acabou de clicar, muitas vezes com um combo ou oferta especial, o autocomplete das buscas do Google e até o smart reply do Gmail, que sugere duas ou três opções de resposta automáticas para emails simples. Motores de recomendação são altamente personalizáveis. Muitas vezes é isso que os torna eficazes. Vão desde modelos mais discretos como o Spotify, que analisa o quanto você gosta de uma música a partir do número de plays, visitas ao perfil da banda ou do álbum, interação com outros fãs e análise sobre o arquivo de áudio bruto para fazer recomendações, quanto

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

AVANÇAR

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

RESOLVENDO PROBLEMAS DO FUTURO

Imagine uma pizzeria. Agora imagine a complexidade de fatores que envolvem o processo de produção de uma pizza: desde escolher os ingredientes, elaborar as receitas, garantir que elas estejam no padrão tanto de elaboração quanto de cozimento, entender a demanda de cada receita em cada região da cidade, pensar na melhor localização para colocar as lojas. Imagine que cada um desses processos pode ser executado por máquinas. Desde quantas rodela de tomate devem ir numa pizza mozzarella, até onde devem estar os foodtrucks que cozinham e entregam a pizza em tempo recorde, passando por fornos inteligentes que possuem uma forma específica para assar cada receita.

Agora pare de imaginar e conheçam a Zume Pizza (vídeo, 36min - **você pode assistir a partir de 28:29**).

<https://www.youtube.com/watch?v=f2aocKWrPG8>

AVANÇAR

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

CONCLUSÃO

As soluções de Machine Learning resolvem os mais variados tipos de problema. Mas diferentes problemas pedem diferentes abordagens. Em tudo. Na forma de pensar, de capturar e modelar os dados. Por isso é importante pensar e modelar o problema antes de escolher a abordagem e os modelos de machine learning ou ainda quais dados vão alimentar estes modelos e o que esperar deles.

PRÓXIMOS PASSOS

Apesar de tudo isto, ainda temos mais **três tópicos** importantes:

- 01** Um produto ou serviço, baseado em dados, não se resume ao modelo de Machine Learning que utiliza. Existe todo o entregável, a experiência da pessoa usuária. Pode ser um app, uma plataforma online, pode ser uma api mas seu cliente final dificilmente aceitará um arquivo de Jupyter Notebook.
- 02** Alguns produtos e serviços possuem mais de uma aplicação ou solução de Machine Learning rodando de forma encadeada ou independente. Note que não existem soluções “one fits all”, ou seja, uma única solução para todos os problemas. Diferentes problemas precisam de diferentes modelos e aplicações.
- 03** Alguns temas possuem um grau de profundidade e personalização muito maior e por isso são menos generalizáveis. Esses são os casos das séries temporais e dos motores de recomendações. Por esse motivo, eles não estão no escopo principal do curso.

Aprofundaremos esses tópicos em outros momentos. :)

MAS NUNCA É DEMAIS LEMBRAR...

ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

com você:

Então nos vemos no próximo texto. Vem com a gente!

PARA SABER MAIS...

Sobre o **Netflix Prize**: [Netflix Prize](#)

[Netflix Never Used Its \\$1 Million Algorithm Due To Engineering Costs](#)

[BellKor's Pragmatic Chaos Wins \\$1 Million Netflix Prize by Mere Minutes](#)

[Netflix: Designing the Netflix Prize](#)

Sobre **NNA**:

[Nearest neighbour algorithm](#)

Sobre os usos de **Machine Learning**:

[Using Machine Learning to Predict Car Accident Risk](#)

[8 problems that can be easily solved by Machine Learning](#)

[Treating cancer, stopping violence... How AI protects us](#)

Sobre **Séries Temporais**:

[The Complete Guide to Time Series Analysis and Forecasting](#)

Sobre **Algoritmo de recomendações do Spotify**:

[Behind Spotify Recommendation Engine](#)



ELABORAÇÃO DE SOLUÇÕES DE MACHINE LEARNING

Deixe seu feedback para continuarmos melhorando sua experiência.

 1 MIN

AVALIAR

VOLTAR PARA O CURSO