



Conceitos Básicos de Machine Learning com Python



Bibliotecas que serão usadas no workshop



O Tsunami *Machine Learning*

- As redes neurais foram teorizadas a muitas décadas
- Com o passar do tempo, foram consideradas impossíveis
- Em 1990, os estudos sobre a área foram abandonados
- Em 2006, Geoffrey Hinton publica artigo sobre Redes Neurais
- Conseguiu precisões jamais antes vistas
- Nenhum algoritmo clássico de ML conseguia se equiparar
- O nome da técnica inventada foi *Deep Learning*
- Recebeu o prêmio Turing da Computação em 2019



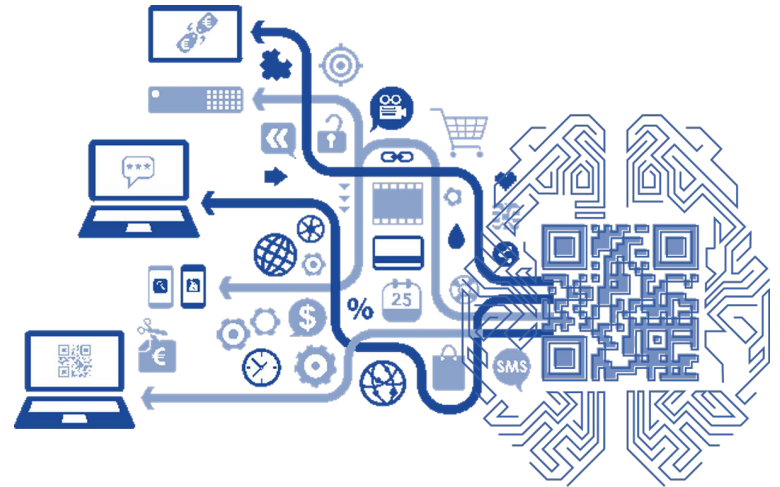
O Tsunami *Machine Learning*

- Logo em seguida, a comunidade científica começou a explorar a técnica extensivamente
- Foram atingidos feitos jamais imagináveis
- Uma década depois, ML já havia conquistado a indústria e a área científica



Machine Learning

- A problemática da popularização e do uso indiscriminado do termo
- O que é? Onde começa? Até onde vai?





O que é Machine Learning?



Machine Learning - Definição

O termo pode ser definido como a arte de programar algoritmos que aprendem a partir de dados

“Field of study that gives computers the ability to learn without being explicitly programmed” - Artur Samuel, 1959

“A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E ” - Tomas Mitchell, 1997



Por que usar Machine Learning?

- Exemplo do detector de SPAM



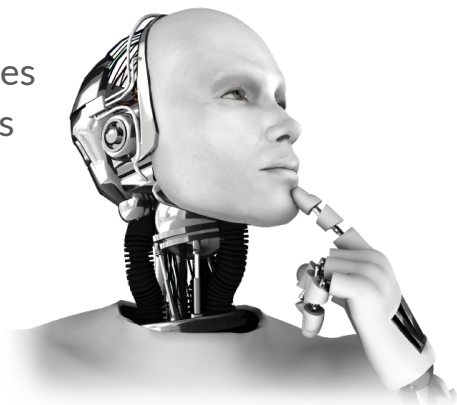


Áreas em que técnicas de ML brilham

- Problemas como o caso do detector de SPAM, onde soluções que não usam ML requerem muitas listas de regras, ainda mais quando essas regras tendem a ser atualizadas ao longo do tempo
- Problemas muito complexos onde o uso de algoritmos tradicionais não atingem bons resultados
- Sistemas que precisam se adaptar constantemente a novas situações e dados
- Conseguir informações sobre grandes quantidades de datas

Algumas áreas que usam ML extensivamente

- Classificação de Produtos em lojas virtuais
- Detecção de tumores cerebrais
- Classificação de notícias
- Detecção de SPAM
- Flagging de conteúdo ofensivo na internet
- Resumo automático de documentos
- Detecção de Clonagem de Cartão de Crédito via compras
- Pesquisa por voz
- Pesquisa por imagem
- Recomendações de produtos, vídeos e notícias
- Games
- Robótica
- Engenharia de Aviões
- Veículos autônomos
- Muitas outras



Tipos de Sistemas de Machine Learning

- Supervised X Unsupervised Learning
- Batch X Online Learning
- Instance Based X Model Based Learning





Supervised X Unsupervised Learning



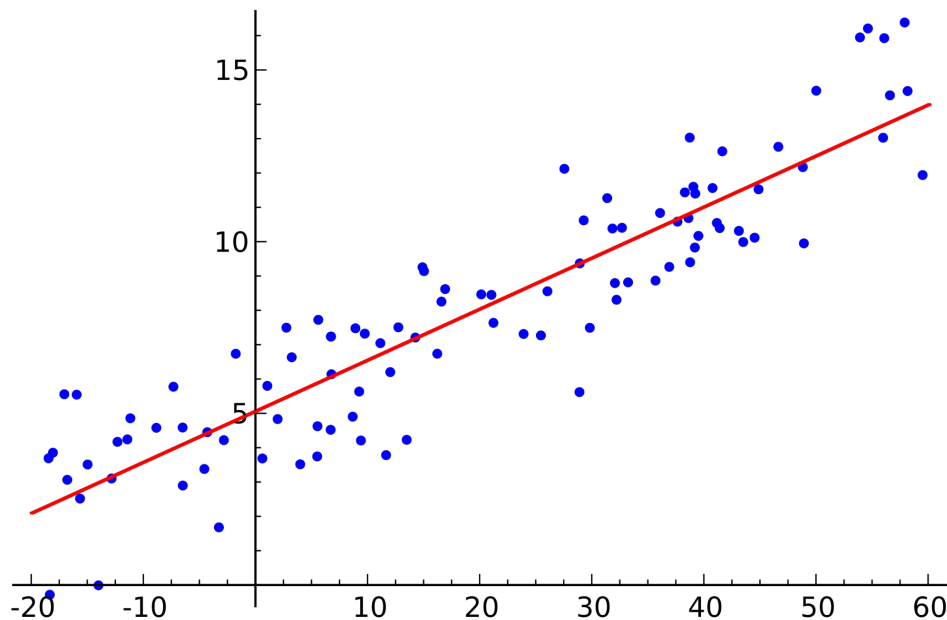
Supervised Learning - Aprendizado Supervisionado

- Alimentamos o algoritmo com dados e soluções (labels)
- Ex: Se queremos prever preços de casas, alimentamos o algoritmo com informações de casas e o preço de cada uma, para que ele possa prever qual o preço de casas tendo apenas as outras informações
- Muito utilizado em tarefas de classificação e regressão

Supervised Learning

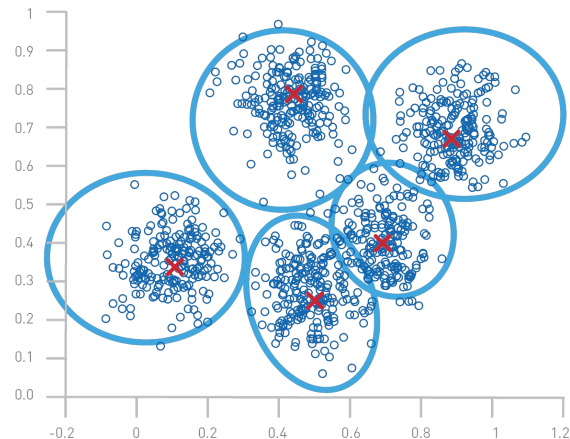
Algoritmos importantes que usam:

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural Networks



Unsupervised Learning - Aprendizado não supervisionado

Training data não possui rótulos (labels). “O sistema tenta aprender sem um professor”





Unsupervised Learning

Algoritmos Importantes que usam

- Clustering
 - K-Means
 - DBSCAN
 - Hierarchical Cluster Analysis (HCA)
- Anomaly Detection
 - One-class SVM
 - Isolation Forest
- Visualization and dimensionality reduction
 - PCA
 - Kernel PCA
 - Locally Linear Embedding
 - t-Distributed Stochastic Neighbor Embedding
- Visualization and dimensionality reduction
 - PCA (Principal Component Analysis)
 - Kernel PCA
 - Locally Linear Embedding
 - t-Distributed Stochastic Neighbor Embedding
- Association rule learning
 - Apriori
 - Eclat



Semisupervised Learning

Combinações diversas de algoritmos supervisionados e não supervisionados

Reinforcement Learning

Extremamente diferente das outras formas de aprendizado

Um agente (sistema de ML) é posto num ambiente, com a capacidade de realizar determinadas ações, as ações que ele escolhe realizar recebem recompensas (positivas / negativas) ou não. Ele no fim aprende o comportamento que gera mais recompensas em cada situação (a melhor estratégia), que recebe o nome de *policy* (política). O sistema é muito usado em robótica, jogos. Curiosidade: “Alpha Go” utiliza muito reinforcement learning!





Batch X Online Learning



Batch Learning

- O sistema não consegue se atualizar (aprender) constantemente. Toda vez que deseja-se adaptação à novos dados, deve se treinar o sistema encima do conjunto inteiro dos dados (antigos e novos)
- Bom quando os dados não se atualizam constantemente
- Custoso com dados que mudam em curtos períodos de tempo



Online Learning

- Treina-se o sistema de forma incremental. Os dados são passados para o algoritmo sequencialmente. Instâncias podem ser passadas individualmente ou em conjunto (mini-batches). O sistema consegue se adaptar a novos dados facilmente e constantemente.
- Excelente para dados que mudam ou incrementam constantemente
- Existe uma preocupação para com o quanto os algoritmos devem atualizar (learning rate - podem se modificar rapidamente, esquecendo dados antigos ou podem modificar de forma mais lenta)



Instance Based X Model Based Learning



Instance Based Learning

- Sistemas generalizam a novos casos por medidas de similaridade aplicadas entre esses e os casos conhecidos



Model Based Learning

- É construído um novo modelo dos novos exemplos de dados para fazer previsões.



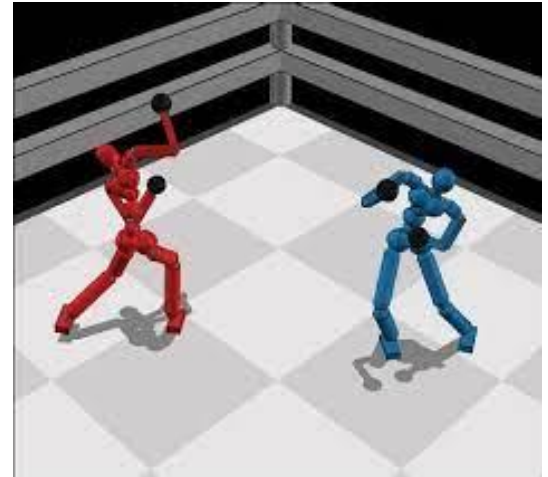
Por hoje é isso!

Até amanhã!



Artigos interessantes recomendados

- Control Strategies for Physically Simulated Characters Performing Two-player Competitive Sports (WON, Jungdam)
 - Vídeo resumo:
https://www.youtube.com/watch?v=SsJ_AusntiU&ab_channel=TwoMinutePapers



Artigos interessantes recomendados

- Emergent Tool Use from Multi-Agent Interaction (BAKER, Bowen)
 - Vídeo resumo:
https://www.youtube.com/watch?v=Lu56xVIZ40M&t=13s&ab_channel=TwoMinutePapers



Livros Recomendados

Python Machine Learning: A Practical Beginner's Guide (Brandon Railey)

Python For Data Analysis (Samuel Burns)

Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow (Aurélien Géron)

