

Introdução

Prof.s.: Eduardo Vargas Ferreira
Walmes Marques Zeviani

O que é Machine Learning?



Estatística

Machine Learning

**Ciência da
computação**

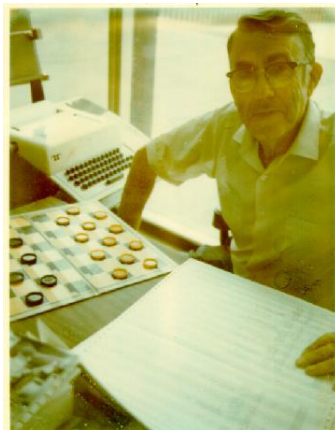


Métodos

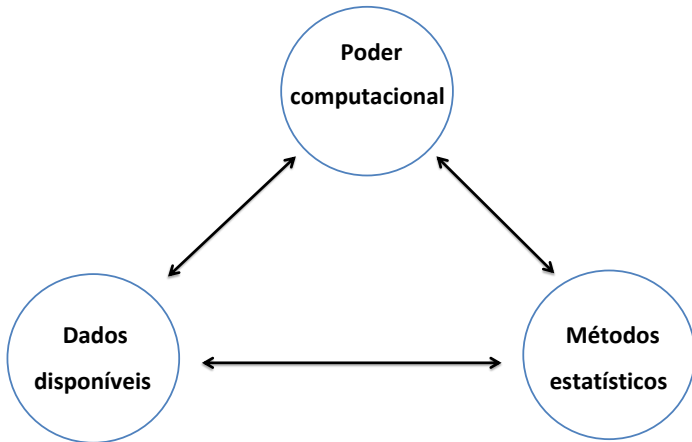
aplicados a

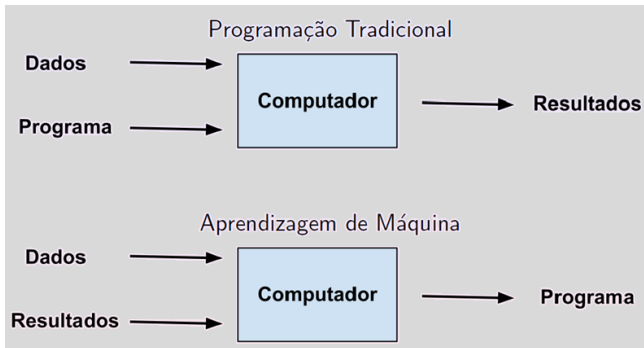
problemas

- Área de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados (Arthur Samuel, 1959).



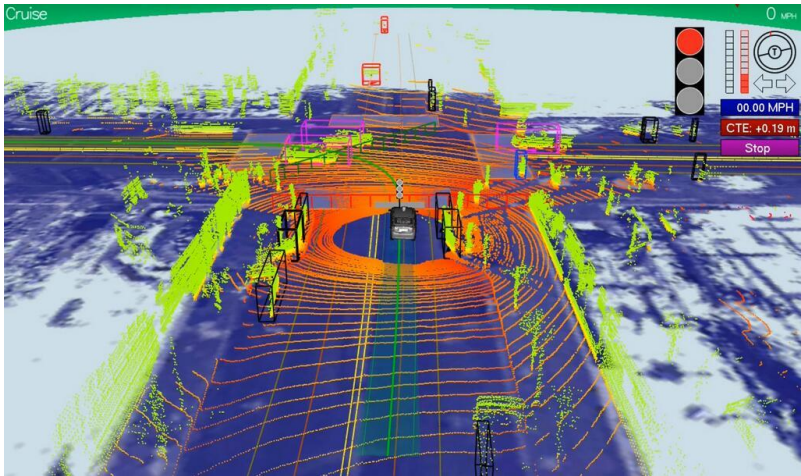
- Estamos interessados em capacitar os algoritmos para desenvolverem ações inteligentes a partir dos dados;





- Machine Learning é mais bem sucedido quando aumenta o conhecimento do especialista, ao invés de substituir;
- Qualquer organização que gere ou agregue dados, provavelmente emprega ao menos um algoritmo de ML para fazer sentido aos mesmos.
- Embora seja impossível listar todas as utilizações dessas técnicas, um levantamento de sucessos recentes inclui aplicações em:
 - ★ Identificação de mensagens de e-mail indesejadas com spam;
 - ★ Segmentação dos clientes para publicidade direcionada;
 - ★ Redução de transações fraudulentas de cartão de crédito;
 - ★ Desenvolvimento de algoritmos para *self-driving cars*;
 - ★ Descobertas de sequências gênicas ligadas à doença etc.

Visão do self-driving car



Nos dias de hoje ...



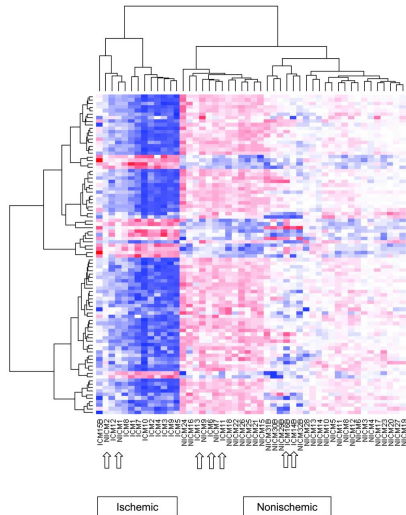
- Quando o DNA virou número;



- A vida virou informação.



Expressão gênica que diferencia pacientes



Serviços Financeiros

- Detecção de Fraude
- Visão 360° do cliente



Utilidades

- Smart Meter Analytics
- Gerenciamento dos Ativos



Transporte

- otimização logística
- congestionamento do tráfego



Mídia Digital

- Segmentação de anúncios em tempo real



Saúde

- Análise de registros médicos
- Monitoramento de doenças



Varejo

- Omni-channel Marketing
- Real-time promotions



Telecomunicação

- Perfil de Cliente
- Análise e Otimização de redes



Segurança

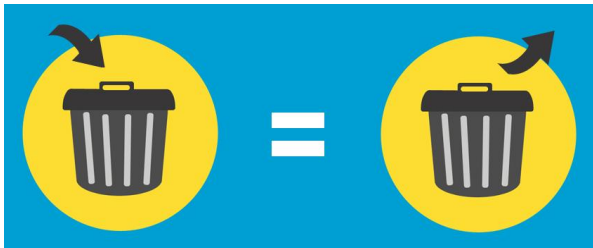
- Vigilância Multimodal
- Detecção de segurança cibernética



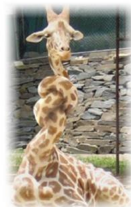
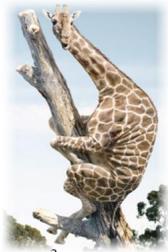
Os limites do Machine Learning



- Embora ML seja amplamente utilizado e tenha um grande potencial de aplicação, é importante entender seus limites;
- Ele tem pouca flexibilidade para extrapolar os parâmetros de aprendizagem e não conhece o senso comum!
- ML é tão bom quanto os dados são para ensinar. É um paradigma “Garbage in, garbage out!”



Exemplo: qual dessas imagens é uma girafa?



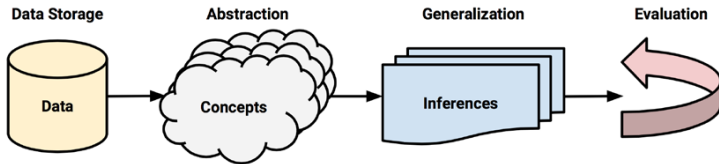
Como as máquinas aprendem?



- Machine Learning é a ciência de **descobrir estruturas** e **fazer previsões** em conjunto de dados;
- O aprendizado é efetuado a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema de aprendizado;



Como as máquinas aprendem?



Fonte: Livro "Machine Learning with R"

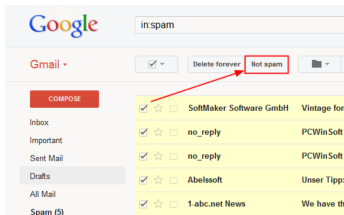
- **Armazenamento dos dados:** utiliza a observação para fornecer uma base para o raciocínio adicional;
- **Abstração:** envolve a tradução dos dados armazenados em representações e conceitos;
- **Generalização:** cria conhecimento e inferência que direcionam ações em novos contextos;
- **Avaliação:** fornece um mecanismo de *feedback* para medir a utilidade do conhecimento adquirido e informar potenciais melhorias.

- 1 **Coleta dos dados:** envolve a coleta de material que o algoritmo utilizará para gerar conhecimento;
- 2 **Exploração e preparação dos dados:** é exigido um trabalho adicional na preparação desses, recodificando-os de acordo com os *inputs* esperados;
- 3 **Formação do modelo:** depois dos dados preparados, o pesquisador já é capaz de dizer o que é possível aprender deles, e como;
- 4 **Avaliação dos modelos:** avaliamos a qualidade do aprendizado, não pode ser pouco (underfitting) nem decorar os dados (overfitting);
- 5 **Melhoria do modelo:** se necessário, podemos melhorar o desempenho do modelo através de estratégias avançadas (p. ex., combinando modelos);

- Os algoritmos de aprendizado são divididos em duas categorias segundo sua finalidade:
 - 1 **Aprendizado supervisionado:** é fornecido um conjunto de exemplos para os quais o rótulo da classe associada é conhecido;
 - Dadas as medições $(x_1, y_1), \dots, (x_n, y_n)$, ensinamos o modelo a **prever** um novo y baseado em x ;
 - Ou seja, o objetivo é construir um classificador que possa determinar corretamente a classe de exemplos não rotulados;
 - Para rótulos discretos, esse problema é conhecido como **classificação** e para valores contínuos como **regressão**.

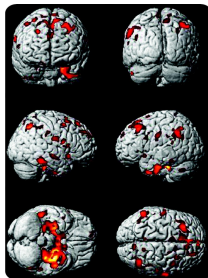
Detecção de Spams

- x_i : e-mail;
- y_i : spam/não spam;

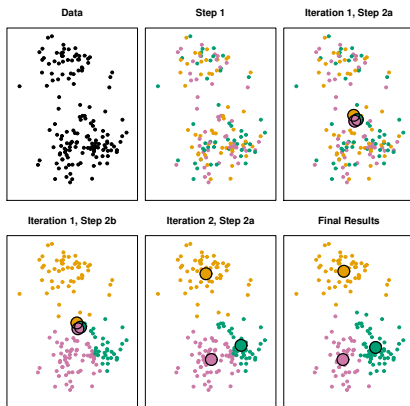


Predição de Alzheimer

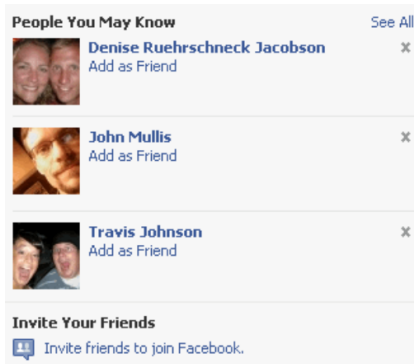
- x_i : imagem do cérebro;
- y_i : Paciente com/sem Alzheimer;






- 2 **Aprendizado não-supervisionado:** a partir dos exemplos (x' s) fornecidos tenta-se determinar algum agrupamento desses.
- Dado x_1, \dots, x_n , descobrir alguma estrutura baseada na similaridade.



- Recomendação de amizades
 - ★ x_i : existe um link entre dois usuários do Facebook;



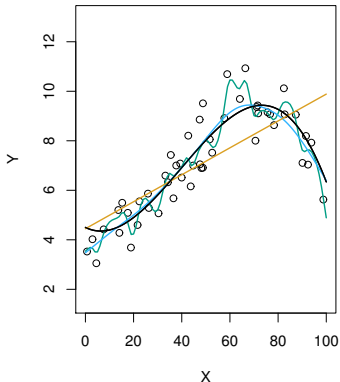
The screenshot shows the 'People You May Know' section of a Facebook profile. It features three suggested friends, each with a profile picture, name, and an 'Add as Friend' button. To the right of each name is a small 'x' icon. At the top right of the section is a 'See All' link. Below the suggestions is a section titled 'Invite Your Friends' with a button that says 'Invite friends to join Facebook.'

Profile Picture	Name	Action
	Denise Rueherschneck Jacobson Add as Friend	x
	John Mullis Add as Friend	x
	Travis Johnson Add as Friend	x

Entendendo a “Avaliação dos modelos”



- Suponha que estamos interessados em estudar a relação entre X e Y ;
- Podemos definir varias funções, $h(x)$, mas qual fornece a melhor predição? **Resposta:** a que apresentar menor custo (ou risco).



- A Função custo descreve o quão bem a superfície de resposta $h(\mathbf{x})$ ajustou-se aos dados.

★ **Soma de quadrado dos desvios (SQD)**

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n [y_i - h(\mathbf{x}_i)]^2$$

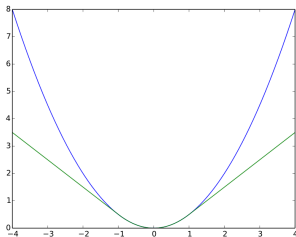
★ **Soma dos desvios absolutos (SDA)**

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n |y_i - h(\mathbf{x}_i)|$$

★ **Huber-M cost**

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n \begin{cases} \frac{1}{2} [y_i - h(\mathbf{x}_i)]^2, & \text{para } |y - h(\mathbf{x}_i)| \leq \delta, \\ \delta |y_i - h(\mathbf{x}_i)| - \frac{1}{2} \delta^2, & \text{caso contrário.} \end{cases}$$

- Abaixo, a curva azul representa a **SQD** e verde a **Função Huber-M**.



- Ela combina as qualidades da **SQD** e **SQA**.
- O parâmetro δ é obtido automaticamente para um específico percentil dos erros absolutos.

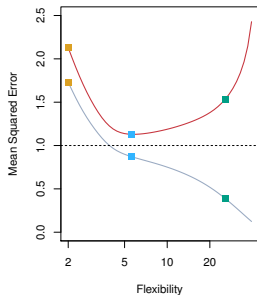
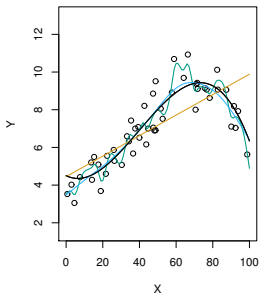
- Fase de treino:
 - ★ Entra com os **dados** e os **verdadeiros níveis**;
 - ★ Obtém um “classificador”.
- Fase de teste (ou aplicação do modelo):
 - ★ Entra com os **dados**;
 - ★ Obtém seu nível (de acordo com o “classificador”).

- ATCCGTATAGTCGATCAGTCAGCTACTATGCGTAT **CANCER**
- TGCATGCATGCAGATCGATCGATCGCCAACGTAC **NO CANCER**
- ATTATATTCTGCGATCGAAGCTATGCGATCGTCGA **CANCER**
- TATGCGCGCGAGTTTTATGAGGCGATCGATGCTA **CANCER**
- ATCGCATCGACGTACGATGCTGATTATTATAGCCG **NO CANCER**
- GATCATGCTGCGAGAGGAGATTTTATGCGATAGA **CANCER**

...

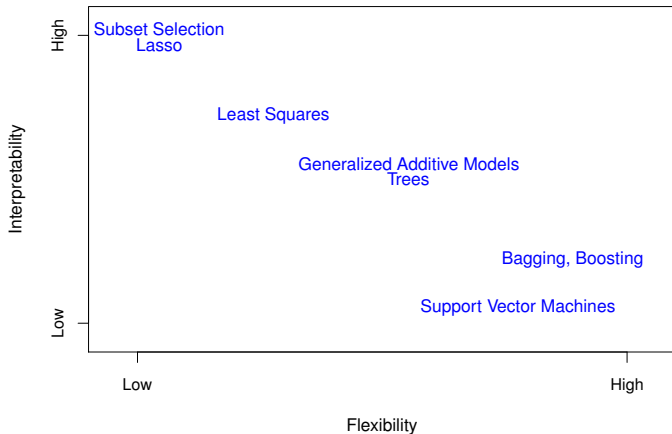
ATCGTCTGATGCAGCGAGCTATGCGTACGTAGCA ?????

- Valores pequenos da função custo corresponde à melhores ajustes;
- No gráfico da direita, a linha cinza representa os dados de treinamento e a vermelha de teste.



- Em inferência em geral **assume-se que o modelo é correto**.
- Isso ocorre pois o principal objetivo está na interpretação dos parâmetros.
- P. ex., *quais parâmetros são significantes? Qual o efeito do aumento da dose do remédio no medicamento?* etc.
- Já em predição, o nosso objetivo maior é simplesmente criar $h(x)$ que tenha **bom poder preditivo**.
- P. ex., não estamos assumindo que a verdadeira regressão é linear!!
- Podemos continuar interpretando os resultados, mas esse - em geral - não é o foco das análises.

Predição versus Inferência



- Duas culturas no uso de modelos estatísticos:
 - ★ **Data Modeling Culture:** Domina a comunidade estatística. Testar suposições é fundamental. Foco em inferência;
 - ★ **Algorithmic Modeling Culture:** Domina a comunidade de machine learning. O modelo é utilizado apenas para criar bons algoritmos preditivos.

L. Breiman. **Statistical modeling: The two cultures.** *Statistical Science*, 16(3):199-231, 2001.