

Apresentação da Disciplina

Prof. Leandro Balby



Aprendizagem de Máquina

Sobre modelos

All models are wrong but some are useful

– *George Box*

Roteiro

1. Motivação
2. Definição
3. Tipos de AM
4. Componentes da Aprendizagem
5. Organização da Disciplina

Sobre AM nos dias de hoje

Andrew Ng¹ no curso de deep learning no Coursera:²

- ▶ *AI is the new Eletricity.*
- ▶ *Eletricity had once transformed countless industries.*
- ▶ *AI will now bring about an equally big transformation.*

¹<http://www.andrewng.org/>

²<https://www.coursera.org/specializations/deep-learning>

Que tipo de transformação AM está trazendo?

- ▶ AI Fashion Designer: <https://www.technologyreview.com/s/608668/amazon-has-developed-an-ai-fashion-designer/>
- ▶ Climate-Change Research:
<https://www.technologyreview.com/the-download/608726/climate-change-research-is-getting-a-big-dose-of-ai/>
- ▶ AI and Sexism:
<https://www.technologyreview.com/the-download/608700/ai-learns-sexism-just-by-studying-photographs/>
- ▶ AI and games: <https://blog.openai.com/dota-2/>

Todos os links são de 2017 e AI está associada a AM em todos os casos.

Roteiro

1. Motivação
2. Definição
3. Tipos de AM
4. Componentes da Aprendizagem
5. Organização da Disciplina

O que é aprendizagem?

A persisting change in human performance or performance potential... [which] must come about as a result of the learner's experience and interaction with the world.

– From Psychology of Learning for Instruction by M. Driscoll

O que é aprendizagem de máquina?

Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.

– Arthur 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 .

– Tom Mitchell, 1997

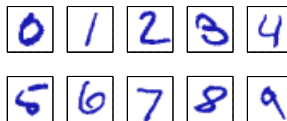
Machine Learning is the science (and art) of programming computers so they can learn from data.

– Aurélien Geron, 2016

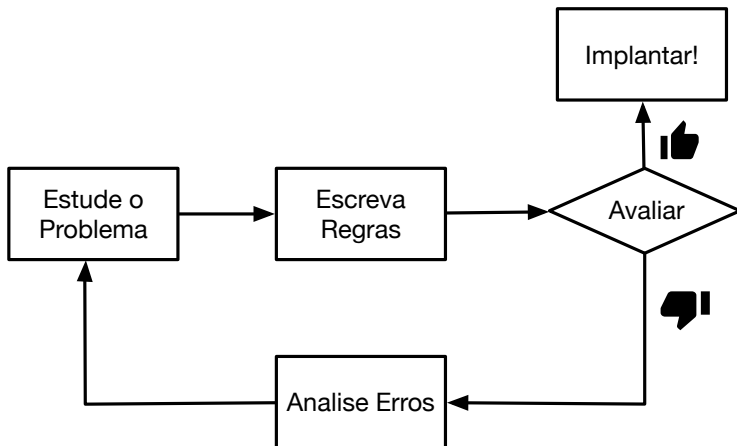
Programação Tradicional vs AM

Considere escrever dois programas usando técnicas tradicionais de programação:

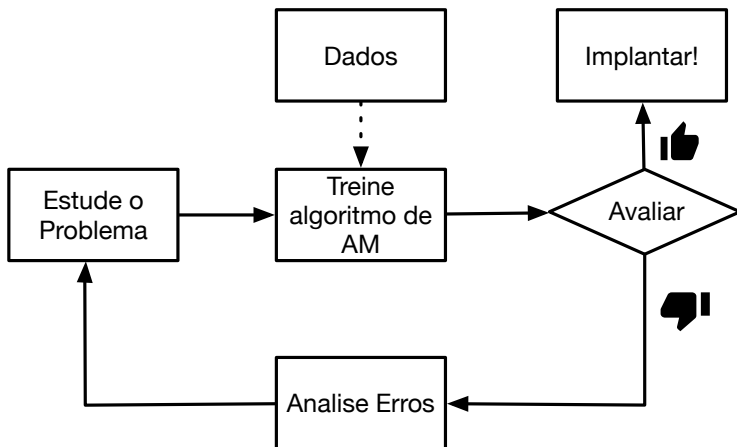
- ▶ detectar e-mails spam
- ▶ detecção de dígitos manuscritos.



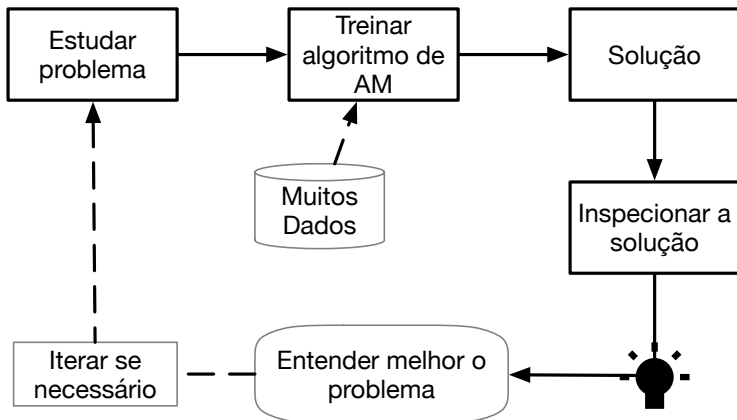
Abordagem Tradicional



Abordagem AM



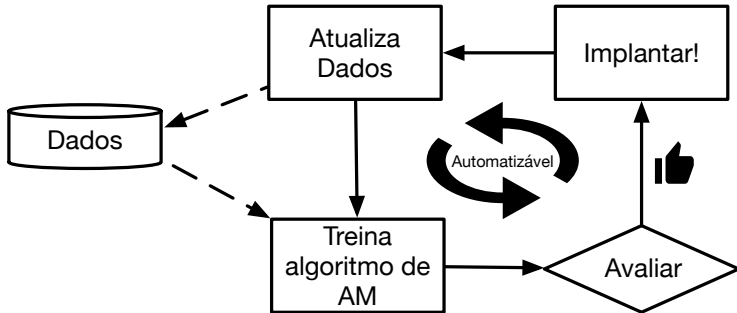
AM pode ajudar a entender melhor o problema



Por que queremos que computadores aprendam?

- ▶ Quando é muito difícil ou muito caro programá-los para resolver uma tarefa.
- ▶ Como: Faça o computador se auto programar mostrando exemplos de entradas e saídas.
 - ▶ Na prática escrevemos um programa parametrizável de forma que o algoritmo se encarrega de achar os parâmetros que melhor se aproximam da função ou comportamento desejável.

Adaptação Automática à Mudanças



Roteiro

1. Motivação
2. Definição
- 3. Tipos de AM**
4. Componentes da Aprendizagem
5. Organização da Disciplina

AM Supervisionada

- ▶ Dados rotulados (por humanos) com “resposta certa”.
- ▶ Métodos de previsão.
 - ▶ Regressão (variável alvo numérica).
 - ▶ Classificação (variável alvo categórica).
- ▶ **Dados exemplos de treino no formato (entrada, saída), produz saídas corretas para novas entradas.**

AM Supervisionada: Exemplo

Quem abandonará o curso na UFCG no próximo semestre?



AM Supervisionada: Exemplo





Como combater o crime antes que ele aconteça?

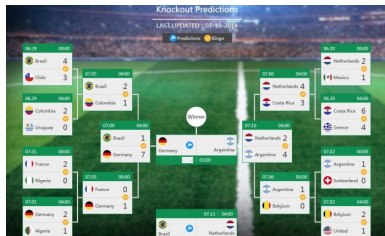


Real Time Crime Center of the “Blue C.R.U.S.H” in Memphis, US
(Crime Reduction Utilizing Statistical History).

AM Supervisionada: Exemplo

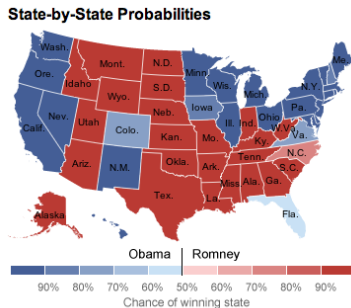
Quem ganhará a copa do mundo?

	Prediction Accuracy - Quarterfinals	Prediction Accuracy - Round of 16	Prediction Accuracy - Group Stage
	100%	100%	58.33%
	100%	100%	56.25%
	100%	100%	37.5%
	75%	100%	/



AM Supervisionada: Exemplo

Quem ganhará as próximas eleições para presidente nos EUA?



Nate Silver previu corretamente o resultado das eleições nos EUA nos 50 estados da federação em 2008. E tentou fazer o mesmo em 2016³

³<http://projects.fivethirtyeight.com/2016-election-forecast/>

AM Supervisionada: Exemplo

Como saber o que as pessoas estão pensando?



AM Supervisionada: Exemplo

Qual a tendência do BM&FBOVESPA no dia seguinte?








AM Supervisionada: Exemplo

Qual a polaridade de um certo tweet?



AM Supervisionada: Exemplo

Várias competições de data science voltadas à predição (e.g. www.kaggle.com)

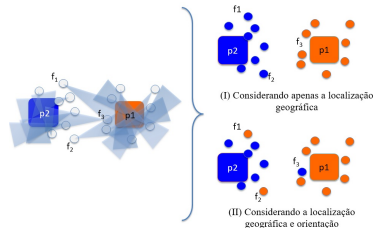
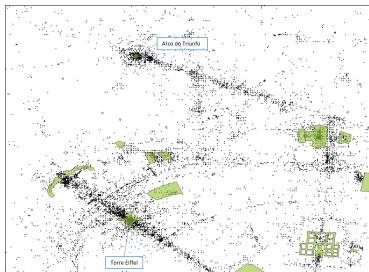
Active	All	Entered	All Categories	Q
	Santander Product Recommendation Can you pair products with people? Featured · a month to go · 154 kernels	\$60,000 390 teams		
	Bosch Production Line Performance Reduce manufacturing failures Featured · 5 days to go · 431 kernels	\$30,000 1,346 teams		
	Outbrain Click Prediction Can you predict which recommended content each user will click? Featured · 2 months to go · 191 kernels	\$25,000 361 teams		
	Allstate Claims Severity How severe is an insurance claim? Recruitment · a month to go · 364 kernels	Jobs 1,734 teams		
	Melbourne University AES/MathWorks/NIH Seizure Prediction Predict seizures in long-term human intracranial EEG recordings Research · 25 days to go · 139 kernels	\$20,000 58 teams		

AM Não Supervisionada

- ▶ Dados não possuem rótulos ou resposta certa.
- ▶ Sistema tenta descobrir estruturas latentes nos dados.
- ▶ Métodos de descrição.
 - ▶ Agrupamento.
 - ▶ Regras de associação.
 - ▶ Visualização e redução de dimensionalidade.
- ▶ **Sistema tenta aprender sem supervisor.**

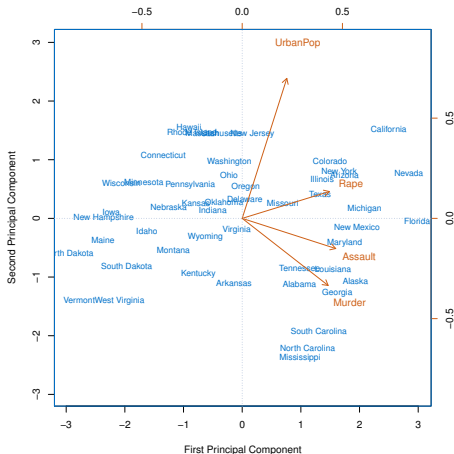
AM Não Supervisionada: Exemplo

Como identificar áreas de interesse através de fotos georeferenciadas?



AM Não Supervisionada: Exemplo

Principal Component Analysis (PCA)

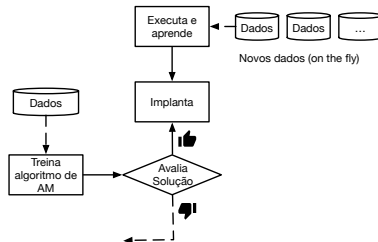


Aprendizagem por Reforço (similar a como animais aprendem)

- ▶ Um agente recebe uma entrada do ambiente e realiza ações que afetam o mesmo.
- ▶ Ocasionalmente, o agente recebe uma recompensa ou punição.
- ▶ O objetivo é aprender a produzir sequências de ações que maximizam a recompensa esperada.
- ▶ Super Mario com Aprendizagem por reforço:
https://www.youtube.com/watch?v=L4KBBAwF_bE

Aprendizagem Batch e Online

- ▶ Batch: treinamento offline usando todos os dados.
 - ▶ Para considerar novos dados o modelo precisa ser retreinado do zero.
- ▶ Online: treinamento incremental.
 - ▶ Também usado para lidar com dados massivos.



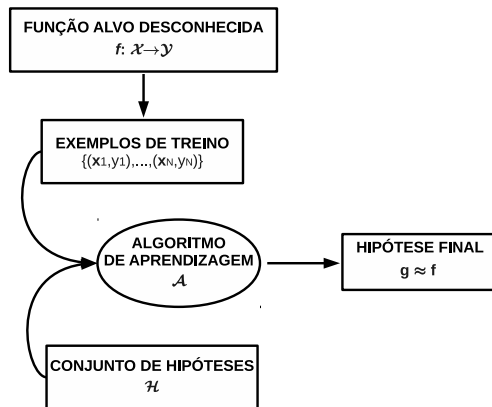
Roteiro

1. Motivação
2. Definição
3. Tipos de AM
- 4. Componentes da Aprendizagem**
5. Organização da Disciplina

Componentes da Aprendizagem

- ▶ Entrada: x (Anos de Educação)
- ▶ Saída: y (salário)
- ▶ Função alvo: $f : \mathcal{X} \rightarrow \mathcal{Y}$ (função ideal de predição de salário)
- ▶ Dados de Treino: $\mathcal{D}^{\text{train}} := \{(x_1, y_1), \dots, (x_N, y_N)\}$ (registros históricos)
- ▶ Hipótese: $g : \mathcal{X} \rightarrow \mathcal{Y}$

Componentes da Aprendizagem [Yaser, 2012]



Desafios

- ▶ Poucos dados.
- ▶ Qualidade ruim dos dados.
- ▶ Dados não representativos.
- ▶ Features irrelevantes.
- ▶ Overfitting e Underfitting.

Roteiro

1. Motivação
2. Definição
3. Tipos de AM
4. Componentes da Aprendizagem
5. Organização da Disciplina

Metodologia

- ▶ Aprenda fazendo (em python).
- ▶ Há muito material disponível para autoaprendizagem, muitos deles serão postados no canvas.
- ▶ Laboratórios de implementação e aplicação.

Ementa

- ▶ Regressão Linear, Multivariada e Polinomial (Previsão Numérica)
- ▶ Regressão Logística, SVM, Árvores e Florestas de Decisão, Boosting (Previsão Categórica)
- ▶ Aprendizagem Estatística com Naive Bayes
- ▶ Sistemas de Recomendação com Fatoração de Matrizes
- ▶ Deep Learning com Redes Neurais
- ▶ **Conteúdo detalhado no Canvas da disciplina**

Avaliação

- ▶ Laboratórios.
- ▶ Atividades de Leitura.
- ▶ Projeto Final.

Referências



Yaser S. Abu-Mostafa, Malik Magdon-Ismail. Learning from Data. AMLBook, 2012.