

ICC204 - Aprendizagem de Máquina e Mineração de Dados

Conceitos de Aprendizado de Máquina



Prof. Rafael Giusti
rgiusti@icomp.ufam.edu.br

Aprendizado ou Aprendizagem?

MICHAELIS

- **aprendizagem**
a·pren·di·za·gem
s.f.
- V. de **aprendizado**, acepções 1, 2, 3.

Aprendizado ou Mineração?

- Existe distinção entre Aprendizado de Máquina e Mineração de Dados?
- Normalmente encara-se Aprendizado de Máquina como parte das técnicas utilizadas para se fazer Mineração de Dados
- Existe uma grande interseção

O que precisamos saber?

- Pra que serve aprendizado de máquina?
- O que é aprendizado de máquina?
- Qual é a definição de aprendizado?
- Como se faz aprendizado de máquina?
- O que é aprendizado indutivo?
- O que é um modelo?
- Como se encontra um modelo?

O que precisamos saber?

- **Pra que serve** aprendizado de máquina?
- O que é aprendizado de máquina?
- Qual é a definição de aprendizado?
- Como se faz aprendizado de máquina?
- O que é aprendizado indutivo?
- O que é um modelo?
- Como se encontra um modelo?

Pra que serve?

- Muito se fala em Aprendizado de Máquina (AM)
 - É o que está por trás do sucesso do Google
 - E do GMail
 - Faz aqueles carros Tesla evitarem acidentes
 - Permite que a Siri (iPhone) e a Alexa (Amazon) conversem com os usuários
 - Ajuda o iFood a ganhar um monte de dinheiro
 - E muito mais

Pra que serve?

- Atualmente Aprendizado de Máquina está em uma fase de transição
 - Ainda é tema de muita pesquisa
 - Mas já é realidade de indústria

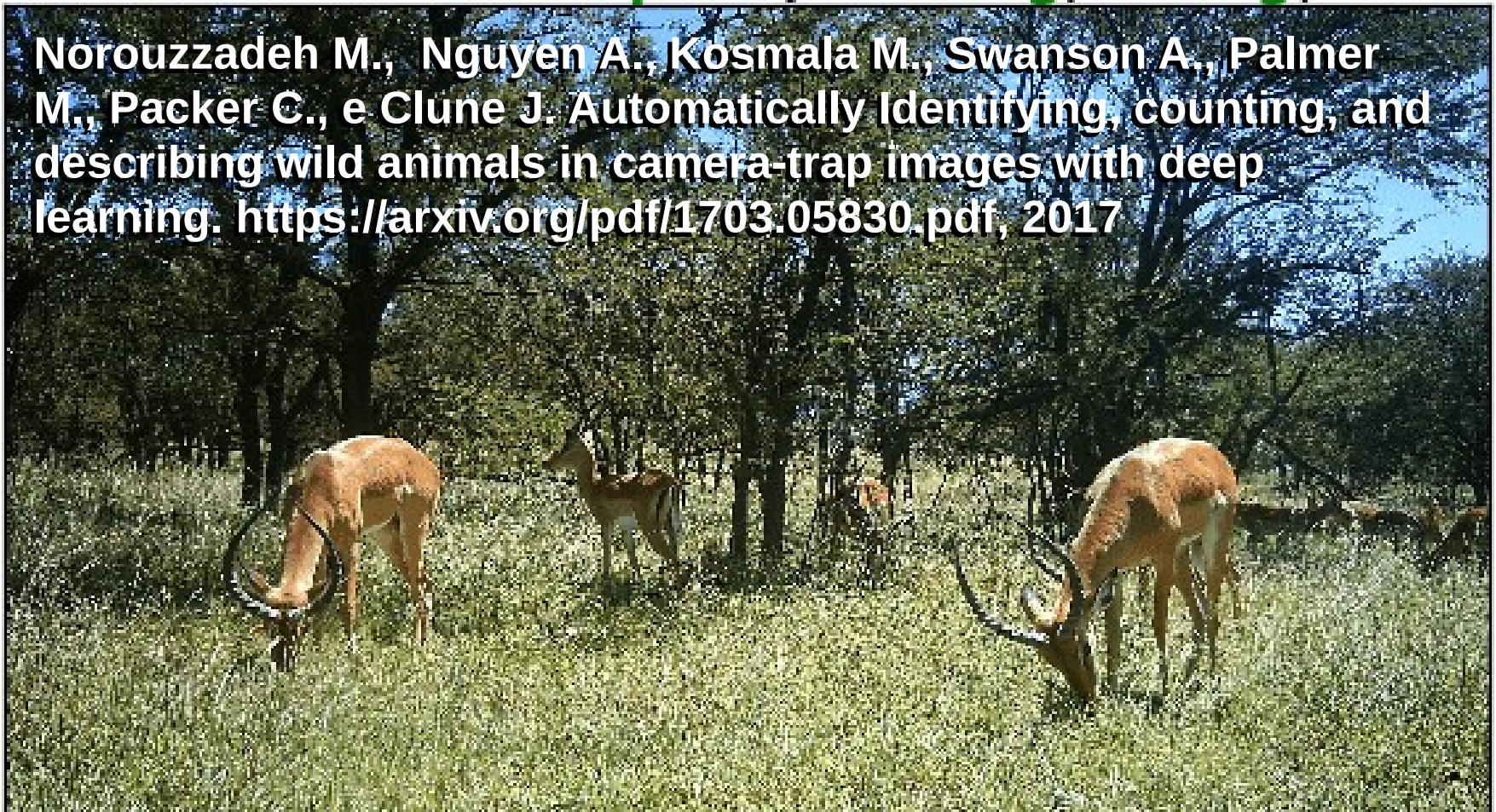
Exemplos de pesquisa em AM

- Identificação de animais em imagens

Human answer: 8 Impala (Standing, Eating)

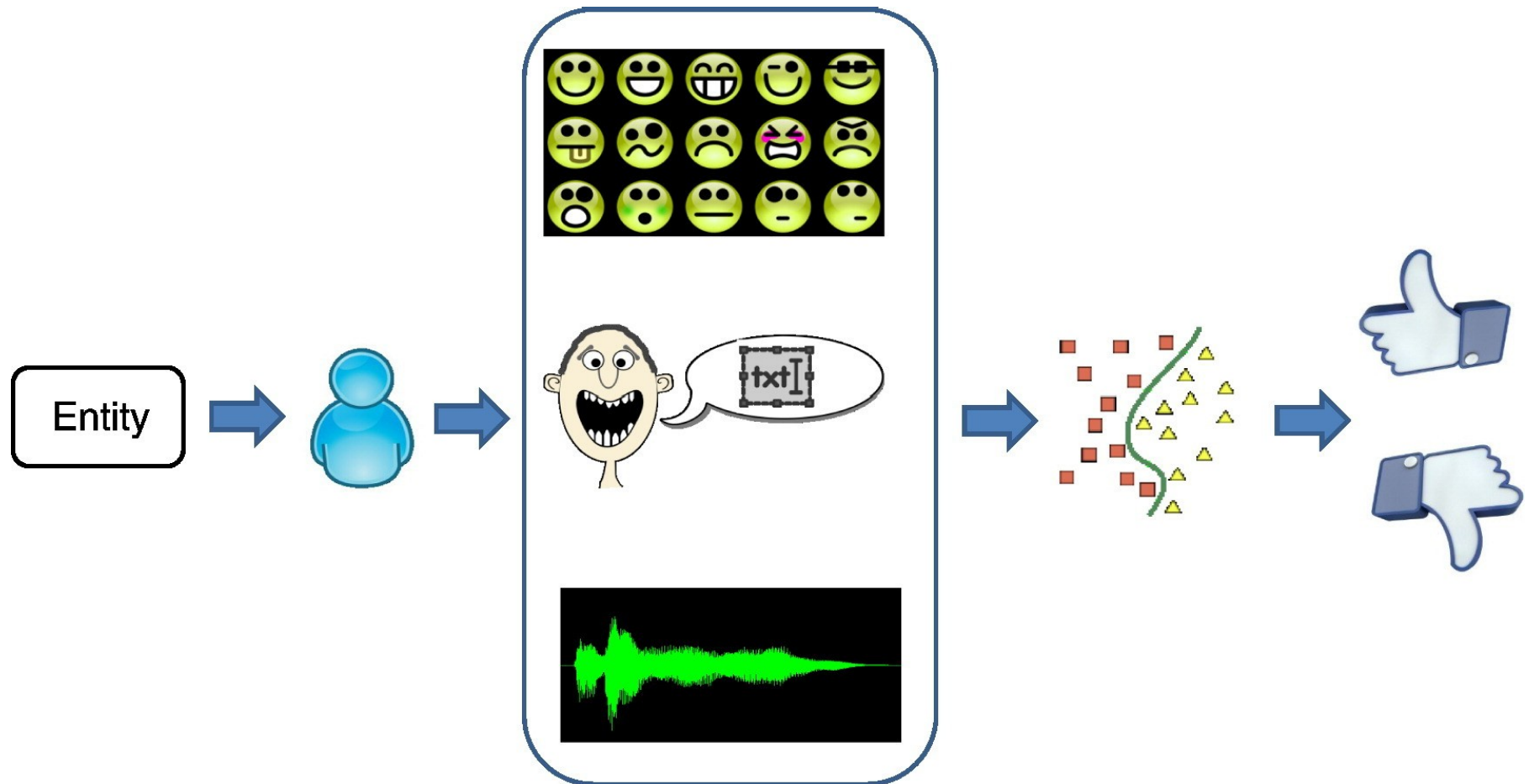
Model answer: 8 Impala (Standing, Eating)

Norouzzadeh M., Nguyen A., Kosmala M., Swanson A., Palmer M., Packer C., e Clune J. Automatically Identifying, counting, and describing wild animals in camera-trap images with deep learning. <https://arxiv.org/pdf/1703.05830.pdf>, 2017



Exemplos de pesquisa em AM

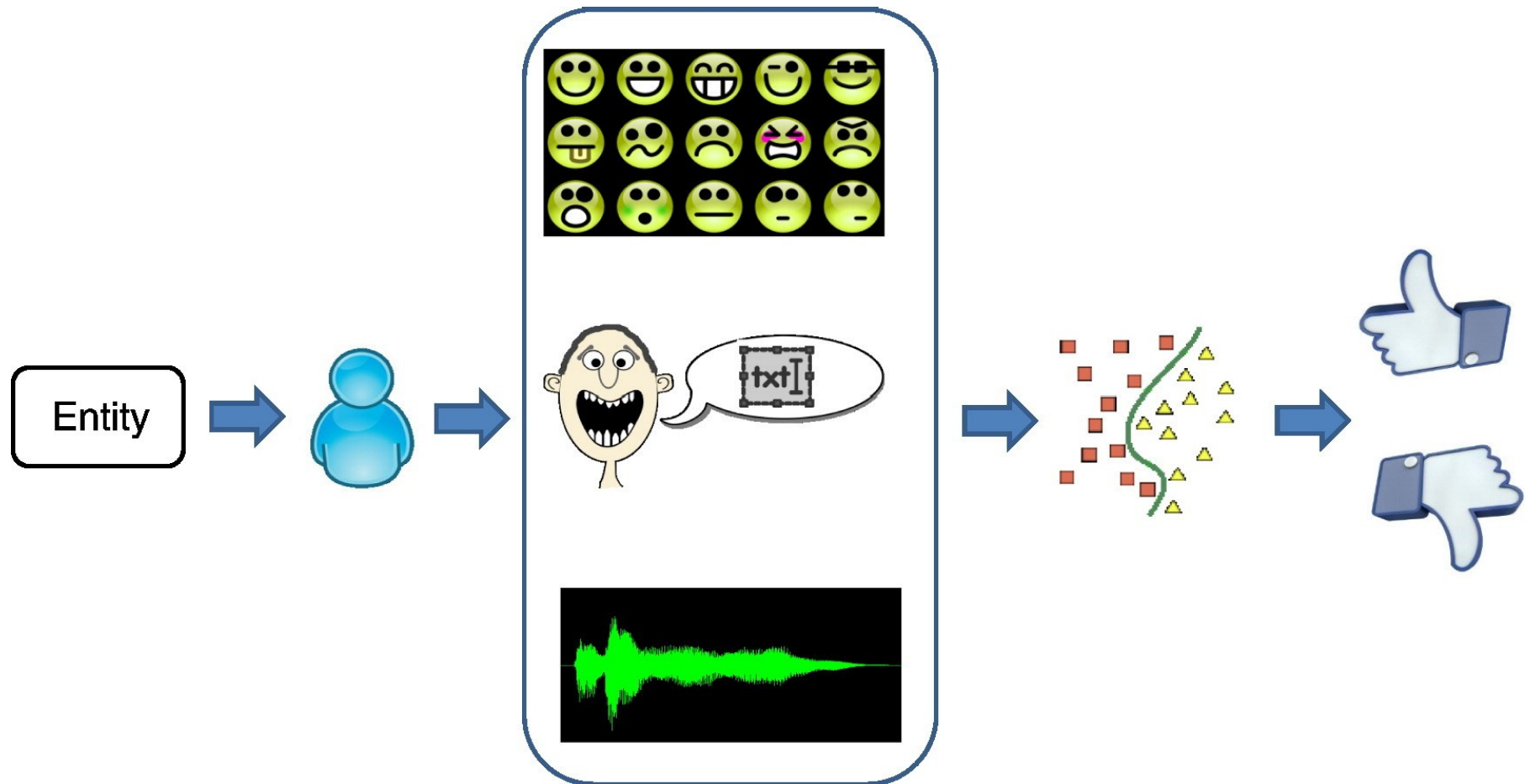
- Análise multimodal de sentimentos



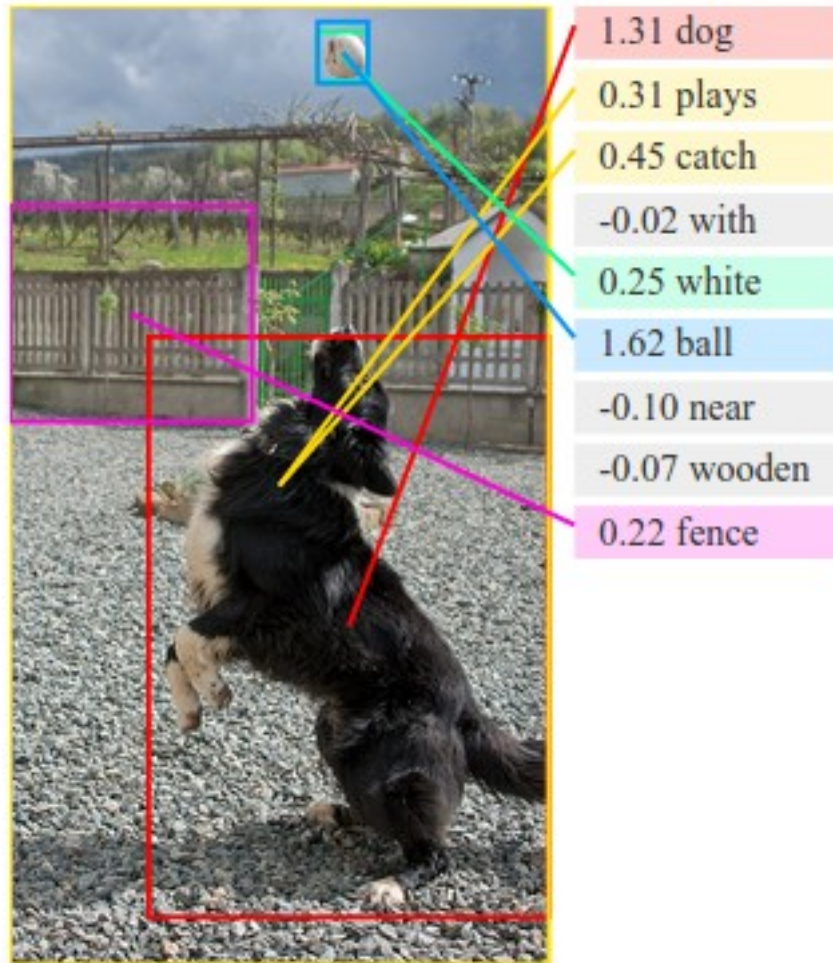
Exemplos de pesquisa e

Utiliza múltiplas mídias para a análise:
Tom de voz, texto, expressão facial etc.

- Análise **multimodal** de sentimentos



Exemplos de pesquisa em AM



Alinhamento

Fonte: <https://cs.stanford.edu/people/karpathy/deepimagesent/>



"black and white dog jumps over bar."



"a young boy is holding a baseball bat."

O que precisamos saber?

- Pra que serve aprendizado de máquina?
- O que é aprendizado de máquina?
- Qual é a definição de aprendizado?
- Como se faz aprendizado de máquina?
- O que é aprendizado indutivo?
- O que é um modelo?
- Como se encontra um modelo?

Definição de AM

- É o estudo científico de algoritmos e modelos estatísticos que sistemas computacionais utilizam para realizar uma tarefa sem serem explicitamente programados

Definição de AM

- É o estudo científico de algoritmos e modelos estatísticos que sistemas computacionais utilizam para realizar uma tarefa **sem serem explicitamente programados**
 - Como fazer um carro dirigir automaticamente?
 - *Spoiler*: não é programando um monte de **if** e casos especiais...

Definição de AM

- É o estudo científico de **algoritmos e modelos estatísticos** que sistemas computacionais utilizam para realizar uma tarefa sem serem explicitamente programados
 - Precisamos ser capazes de representar matematicamente uma experiência de aprendizado

Definição de AM

- É uma área multidisciplinar
- Aprendizado de máquina utiliza-se de
 - Estatística
 - Geometria
 - Álgebra linear
 - Cálculo
 - Conhecimentos de domínio de aplicação

Definição de AM

- É

-

Mas, **professor!!**

Preciso mesmo saber

tudo isso??

O que é importante para AM?

- Conhecimento nunca é demais
- Mas não se pode saber tudo
- O que é...

Útil

Importante

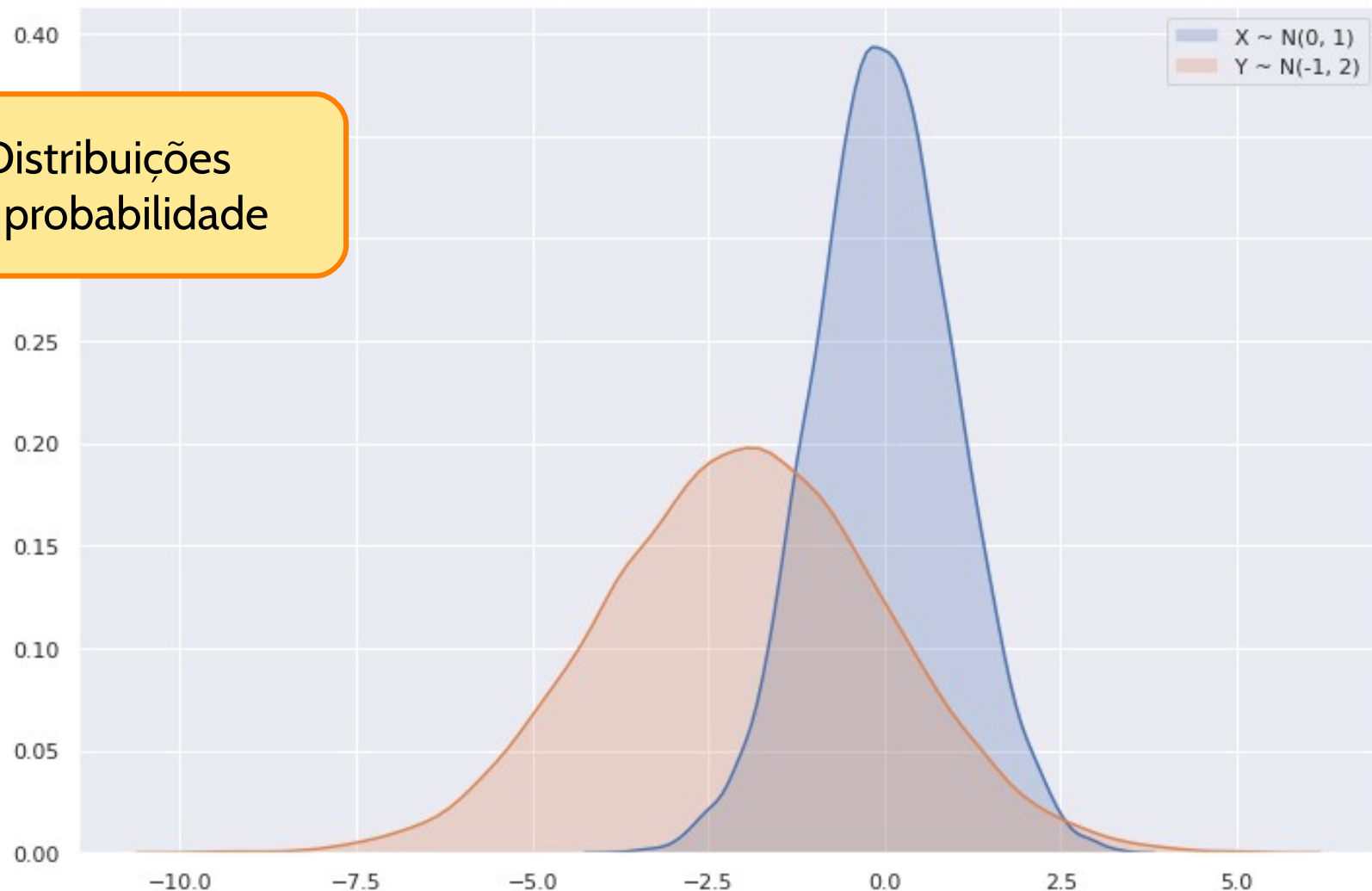
Crucial

...que eu tenha domínio para poder fazer
aprendizado de máquina?

O que é importante para AM?

- De estatística (e teoria das probabilidades)

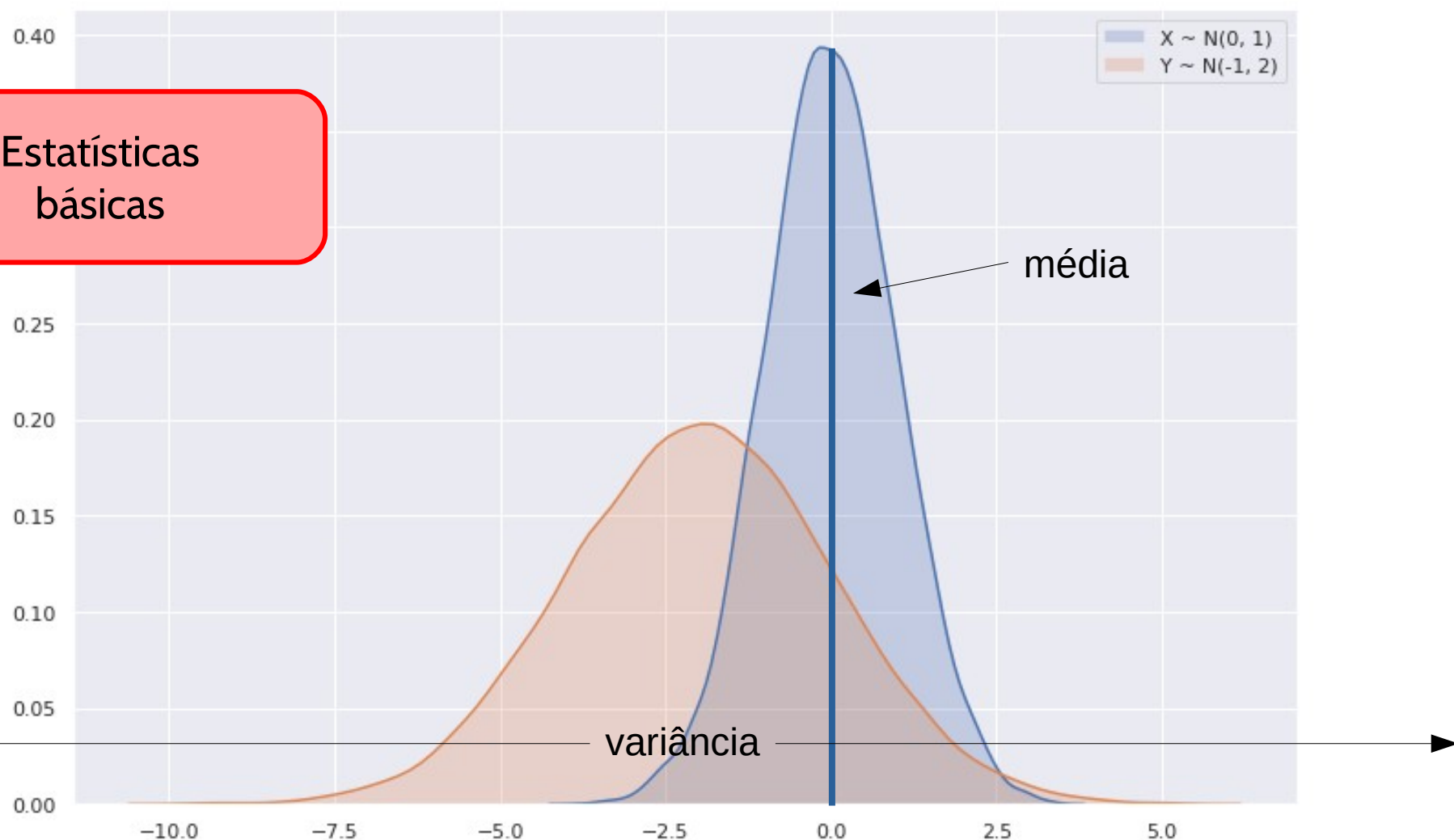
Distribuições
de probabilidade



O que é importante para AM?

- De estatística (e teoria das probabilidades)

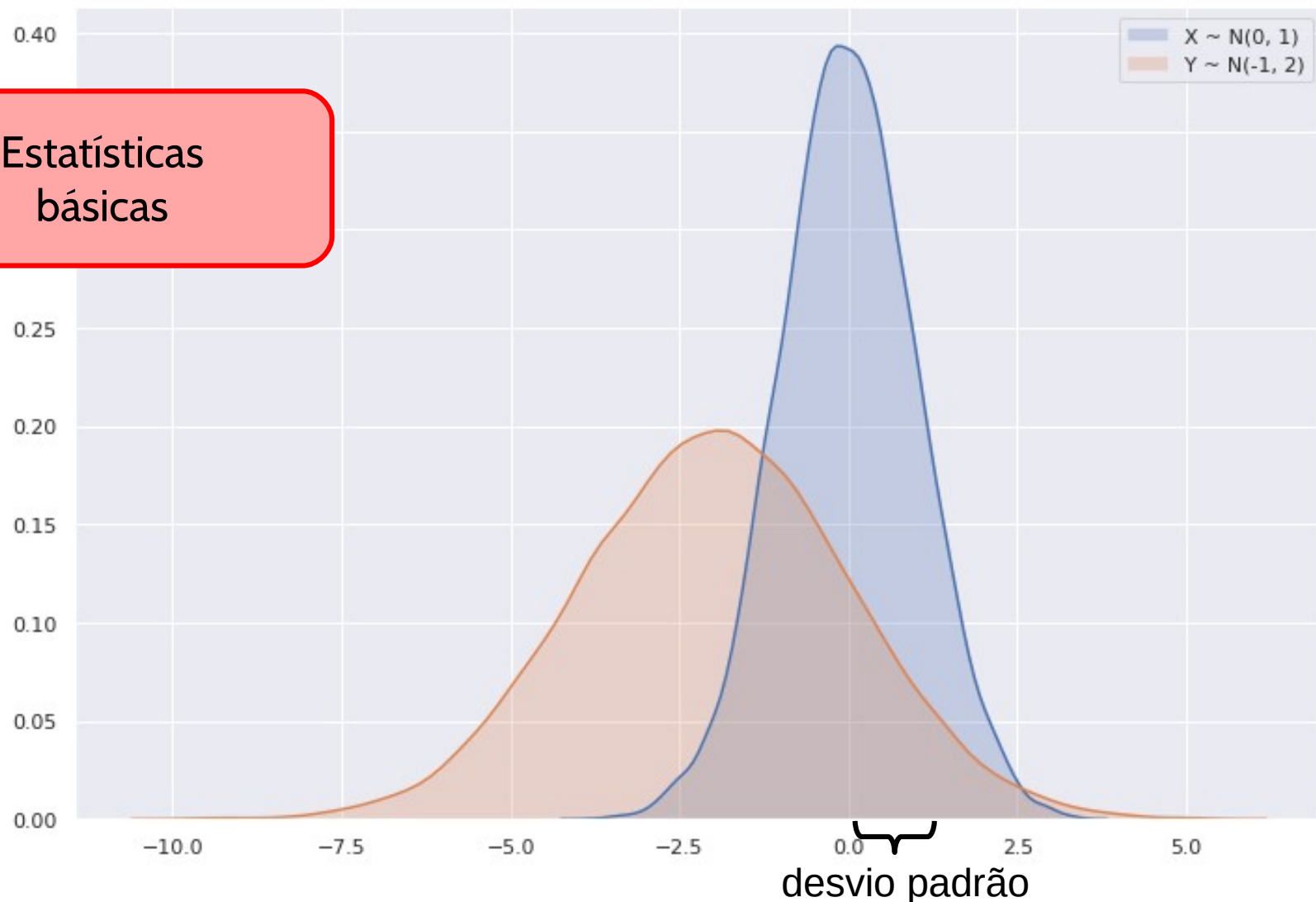
Estatísticas
básicas



O que é importante para AM?

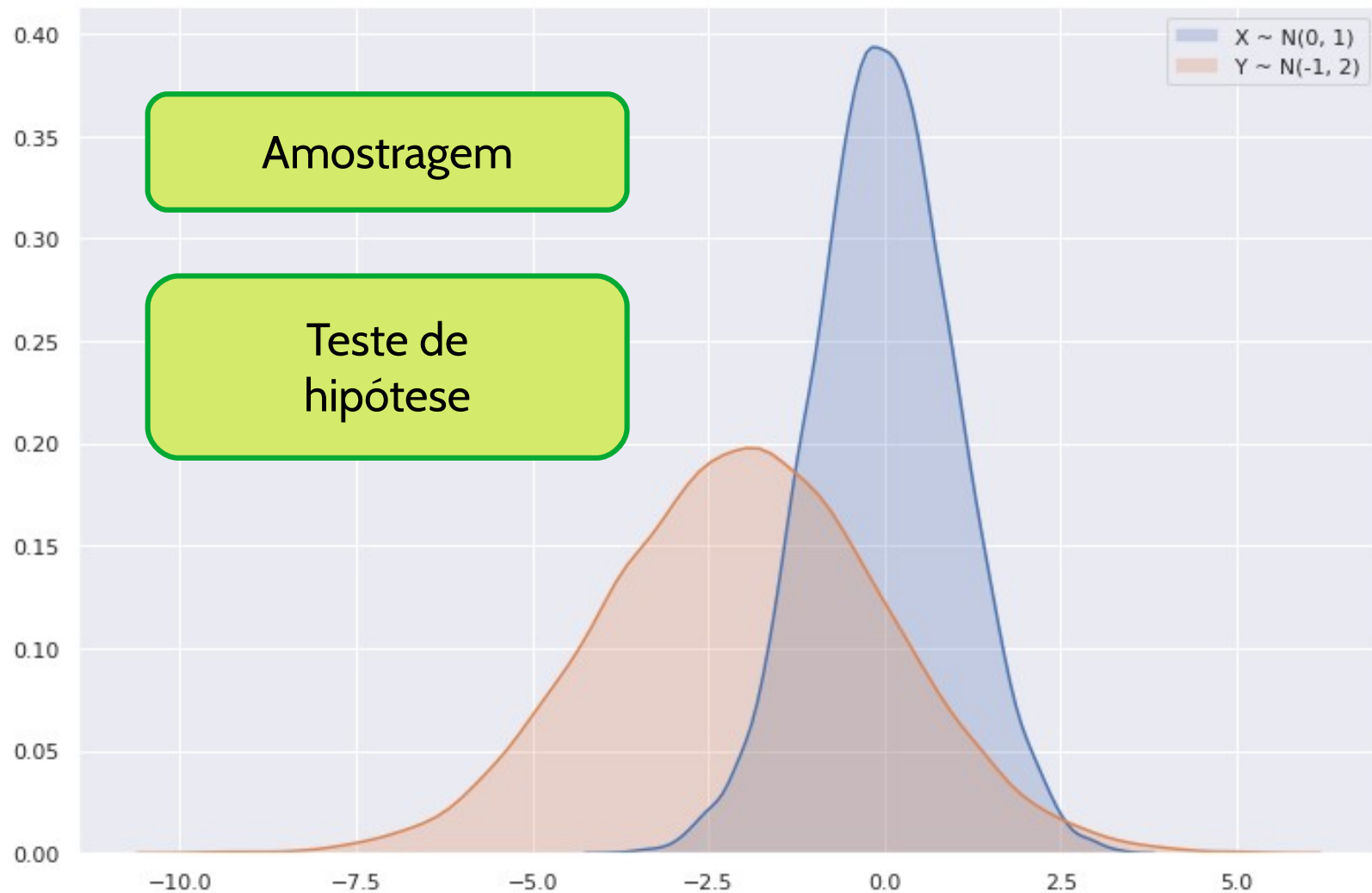
- De estatística (e teoria das probabilidades)

Estatísticas
básicas



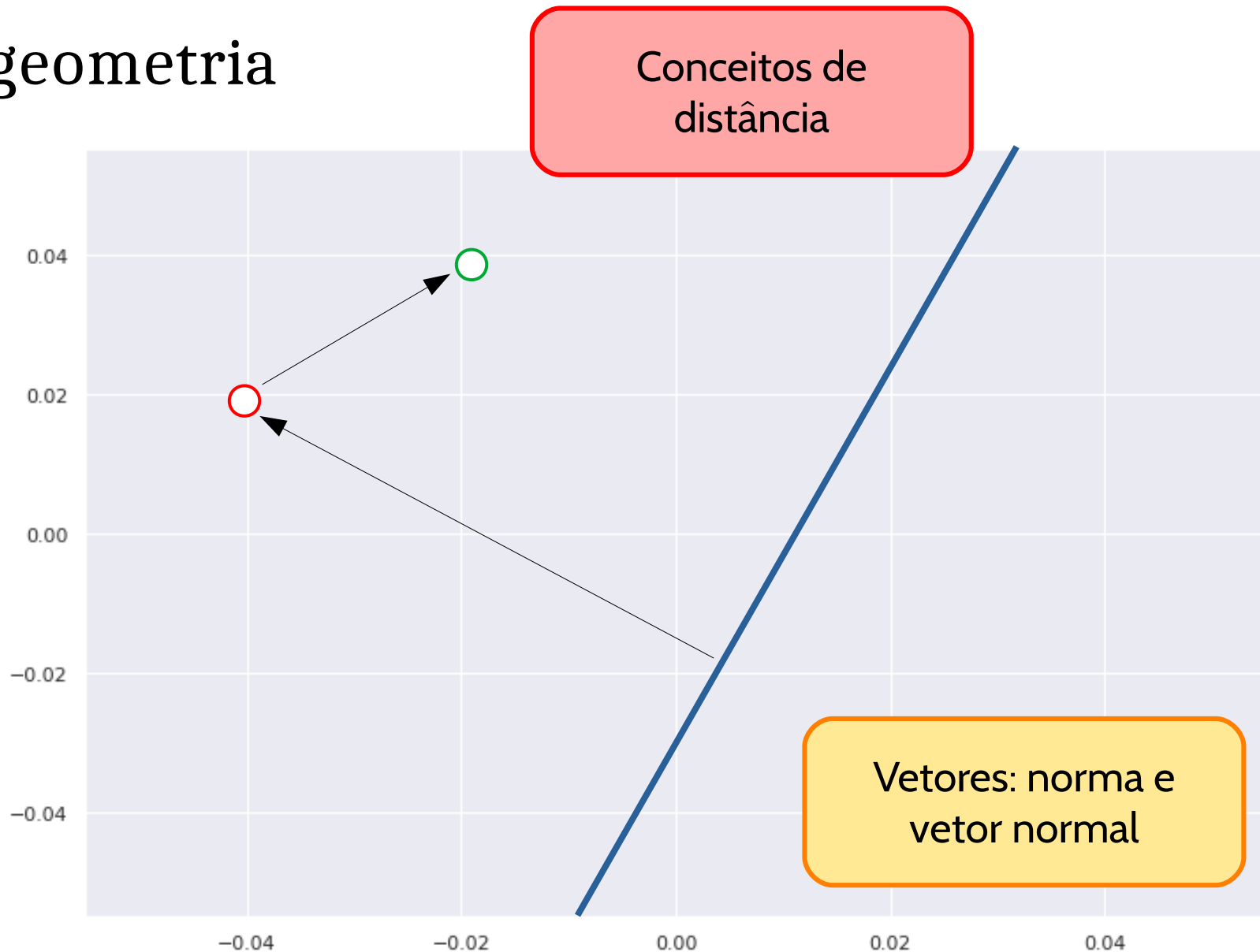
O que é importante para AM?

- De estatística (e teoria das probabilidades)



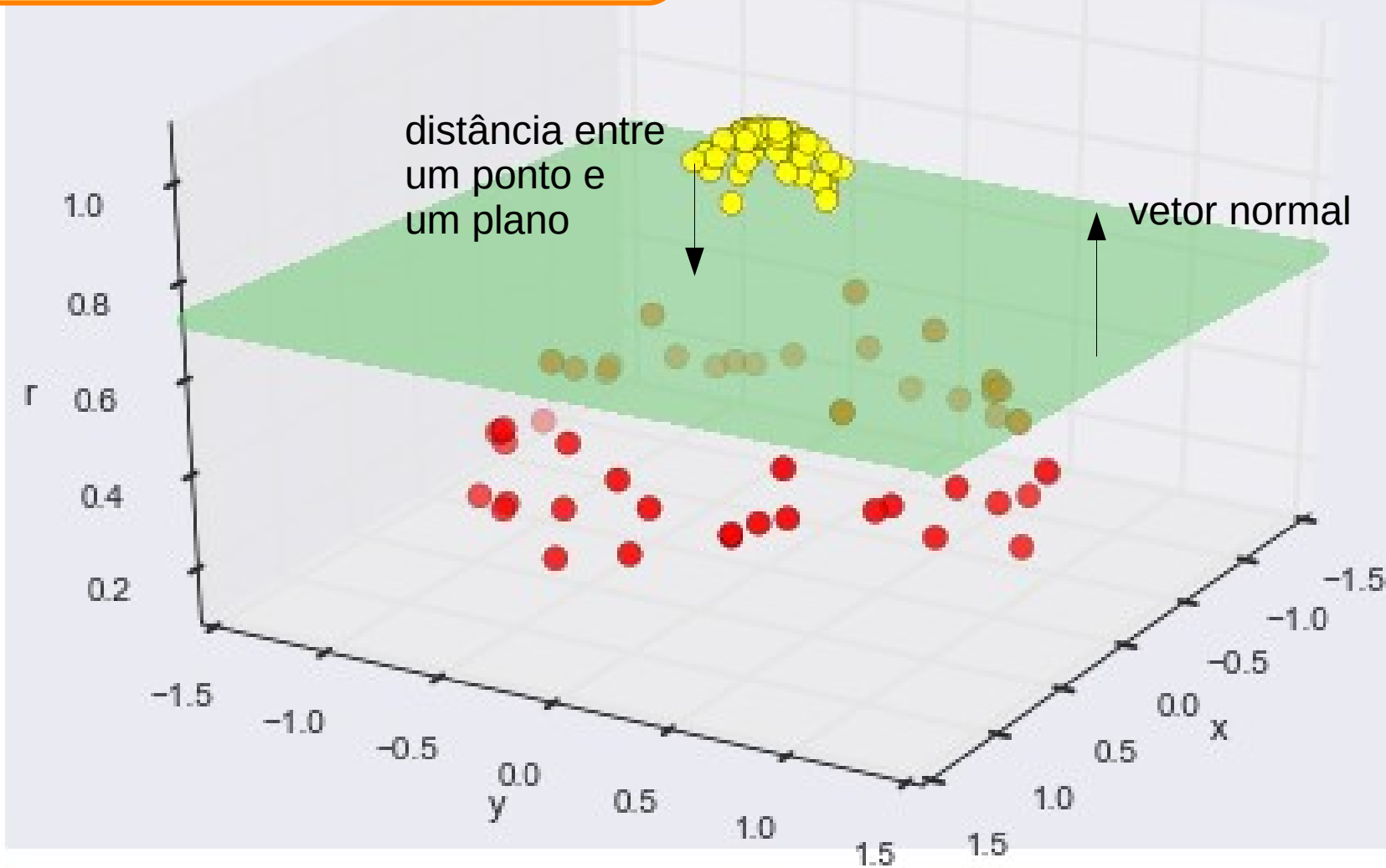
O que é importante para AM?

- De geometria



O que é importante para AM?

Abstração dos conceitos para espaços de alta dimensionalidade



O que é importante para AM?

- Álgebra linear (vamos rever)
 - Operações com matrizes e vetores
 - Determinante

$$\begin{pmatrix} 3 & 8 & 10 & 5 & 4 \\ 8 & 9 & 1 & 5 & 3 \\ 5 & 2 & 7 & 3 & 4 \\ 8 & 8 & 4 & 10 & 9 \\ 9 & 1 & 8 & 10 & 3 \end{pmatrix} \quad \begin{pmatrix} 4 & 5 \\ 1 & 9 \\ 10 & 2 \\ 3 & 9 \\ 6 & 4 \end{pmatrix}$$

O que é importante para AM?

- Cálculo (vamos rever)
 - Limite e derivada são úteis

$$\lim_{n \rightarrow \infty} \sum_{i=0}^n \frac{X_i}{n} = \overline{X}$$

Teorema do limite central

$$\frac{d\phi(z)}{dz} = \phi(z) (1 - \phi(z))$$

Derivada parcial

O que é importante para AM?

- É uma área multidisciplinar
- Aprendizado de
 - Estatística
 - Geometria
 - Álgebra linear
 - Cálculo
 - Conhecimentos de...

Se você coçou a cabeça em alguma das caixinhas vermelhas dos slides anteriores, faça uma visita à biblioteca!

As caixinhas verdes são tópicos sobre os quais vamos falar bastante **ou** dos quais faremos uma boa revisão quando chegar a hora

O que é importante para AM?

- Conhecimentos de domínio de aplicação?
 - Aprendizado de Máquina raramente é uma atividade-fim
 - Fazemos AM para realizar alguma outra tarefa
 - Esses são os domínios de aplicação
 - Classificação de plantas? Vá estudar botânica
 - Veículos autônomos? Estude física e trânsito
 - Ou contrate um especialista

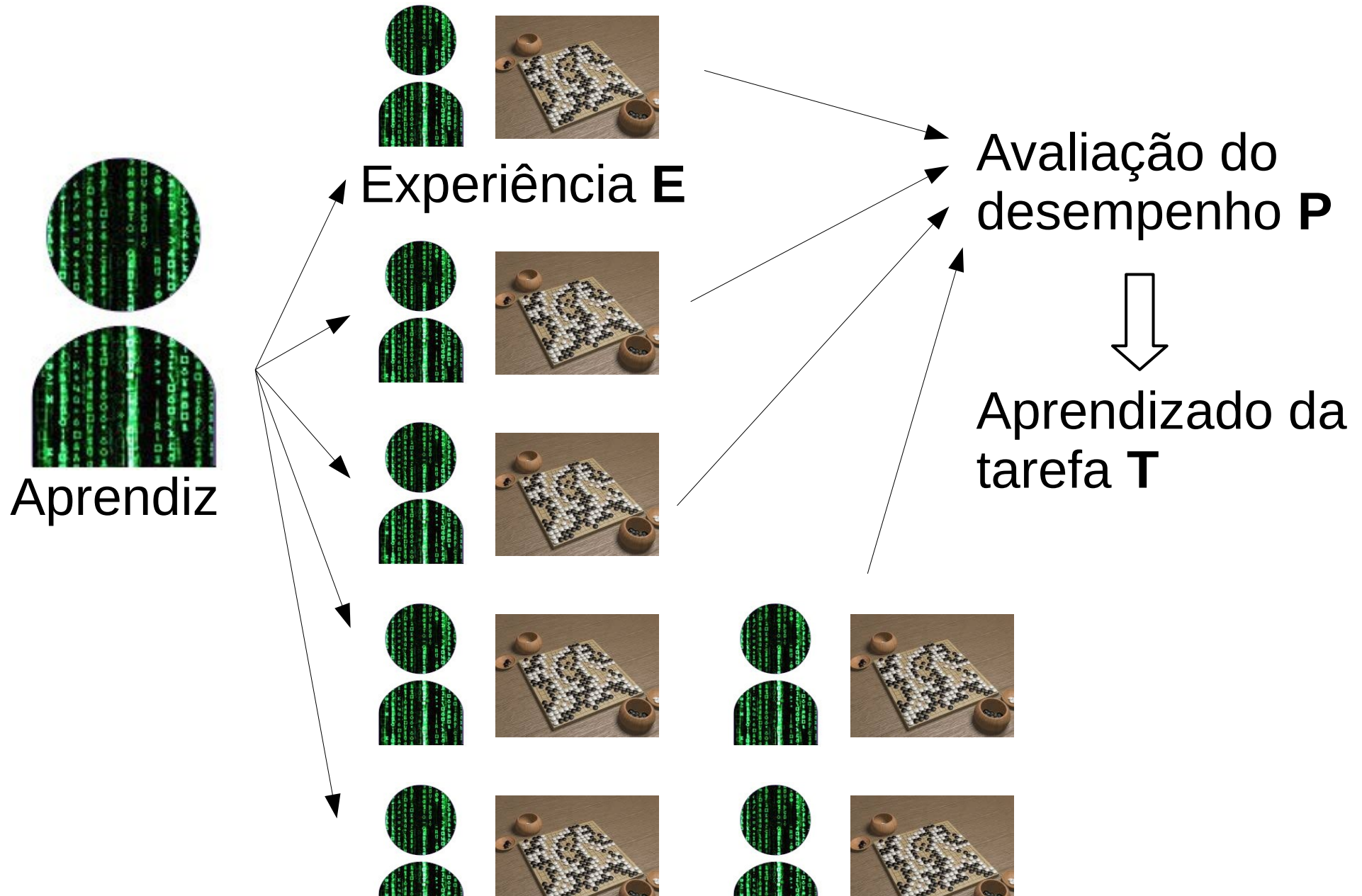
O que precisamos saber?

- Pra que serve aprendizado de máquina?
- O que é aprendizado de máquina?
- Qual é a **definição de aprendizado?**
- Como se faz aprendizado de máquina?
- O que é aprendizado indutivo?
- O que é um modelo?
- Como se encontra um modelo?

Definição de aprendizado

- De acordo com Tom Mitchell
 - Diz-se que um programa aprende com a experiência **E** a realizar uma classe de tarefas **T**, com respeito a uma métrica de desempenho **P**, se...
 - ...o seu desempenho nas tarefas **T**, com respeito a **P**, aumenta com a experiência **E**

Definição de aprendizado



Exemplos de aprendizado

- Aprender a dirigir um veículo
 - Tarefa **T**: dirigir em uma via de duas faixas utilizando sensores óticos
 - Medida **P**: distância média dirigida antes de cometer um erro (a definição de "erro" é dada por um especialista humano)
 - Experiência **E**: vídeos de humanos dirigindo com registro do uso dos controles do carro (volante, pedais, marcha etc)

Exemplos de aprendizado



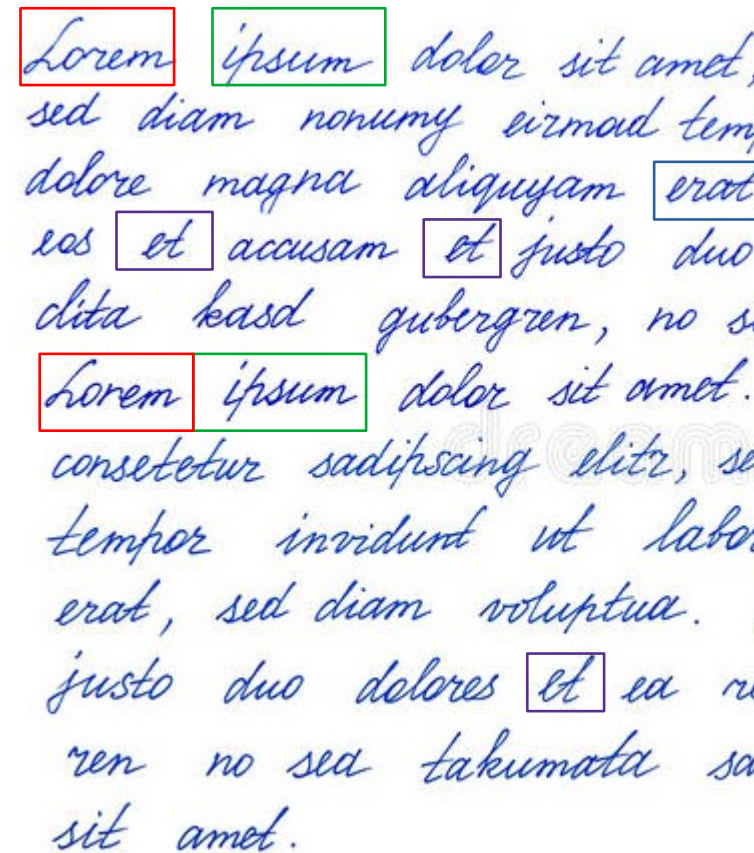
- Jogar Go

- Tarefa **T**: jogar Go
- Medida **P**: número de jogos vencidos
- Experiência **E**: praticar jogando contra si mesmo

<https://techcrunch.com/2017/05/24/alphago-beats-planets-best-human-go-player-ke-jie/>

Exemplos de aprendizado

- Aprender a reconhecer texto manuscrito (OCR)
 - Tarefa **T**: reconhecer e classificar palavras através de imagens
 - Medida **P**: fração de palavras corretamente identificadas
 - Experiência **E**: um conjunto de imagens contendo palavras corretamente classificadas



Handwritten text in cursive script, likely Latin, with several words highlighted by colored boxes (red and green) to illustrate OCR results. The text is arranged in two paragraphs. The first paragraph contains: "Lorem ipsum dolor sit amet, sed diam nonumy eizmod tempor, dolore magna aliquyam erat, eos et accusam et justo duo clita kasd gubergren, no s". The second paragraph contains: "Lorem ipsum dolor sit amet. consetetur sadipscing elitr, se tempor invidunt ut labor erat, sed diam voluptua. justo duo dolores et ea re ren no sea takumata sa sit amet." The boxes highlight specific words: "Lorem" (red), "ipsum" (green), "dolor" (red), "sit" (green), "amet" (red), "erat" (green), "et" (red), "accusam" (green), "et" (red), "justo" (green), "duo" (red), "clita" (green), "kasd" (red), "gubergren" (green), "no" (red), "s" (green), "consetetur" (red), "sadipscing" (green), "elitr" (red), "se" (green), "tempor" (red), "invidunt" (green), "ut" (red), "labor" (green), "erat" (red), "sed" (green), "diam" (red), "voluptua" (green), "justo" (red), "duo" (green), "dolores" (red), "et" (green), "ea" (red), "re" (green), "ren" (red), "no" (green), "sea" (red), "takumata" (green), "sa" (red), "sit" (green), "amet" (red).

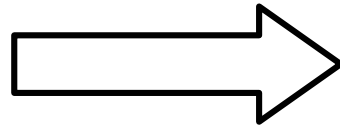
O que precisamos saber?

- Pra que serve aprendizado de máquina?
- O que é aprendizado de máquina?
- Qual é a definição de aprendizado?
- **Como** se faz aprendizado de máquina?
- O que é aprendizado indutivo?
- O que é um modelo?
- Como se encontra um modelo?

Modalidades de AM

- AM é tipicamente empregado sob uma ótica de **redução de problemas**
- Para **reduzir** um problema **A** em um problema **B**
 - Especifique **A** em termos de **B**

Entrega de produtos

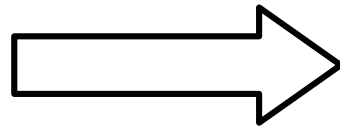


Caixeiro viajante

Modalidades de AM

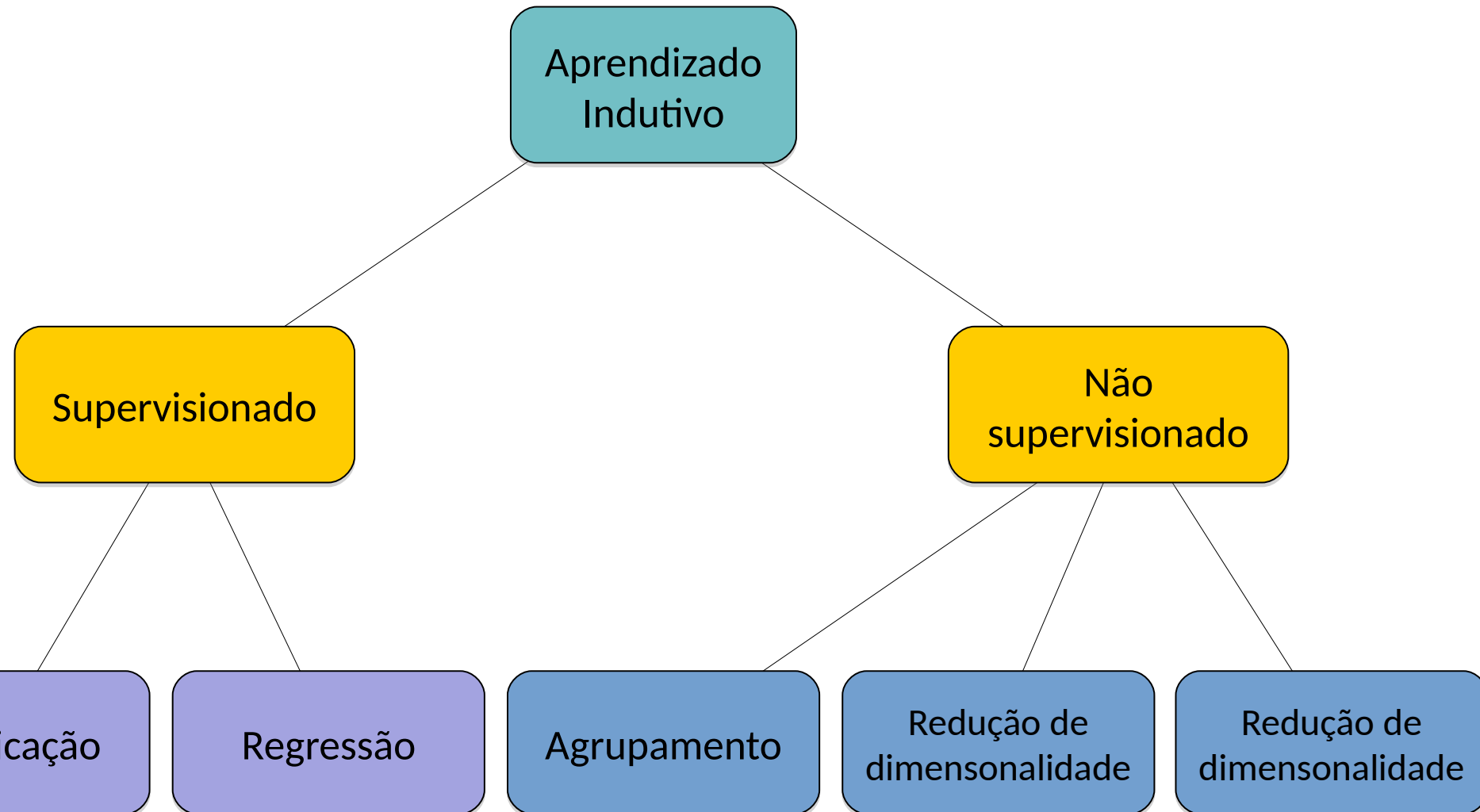
- AM é tipicamente empregado sob uma ótica de **redução de problemas**
- Para **reduzir** um problema **A** em um problema **B**
 - Especifique **A** em termos de **B**

Identificar animais
que aparecem em
uma foto



Classificação

Modalidades de AM



Modalidades de AM

Aprendizado

Qual é a experiência **E**?
Como avalio o desempenho **P**?

Supervisionado

Não supervisionado

Classificação

Regressão

Agrupamento

Redução de
dimensionalidade

Redução de
dimensionalidade

Modalidades de AM

Aprendizado

Vou dizer explicitamente para o aprendiz o que ele deve aprender?

Supervisionado

Classificação

Regressão

Agrupamento

Redução de
dimensionalidade

Redução de
dimensionalidade

Modalidades de AM

Aprendizado

Vou deixar o aprendiz inferir por conta própria o que ele deve obter da experiência?

Não supervisionado

Classificação

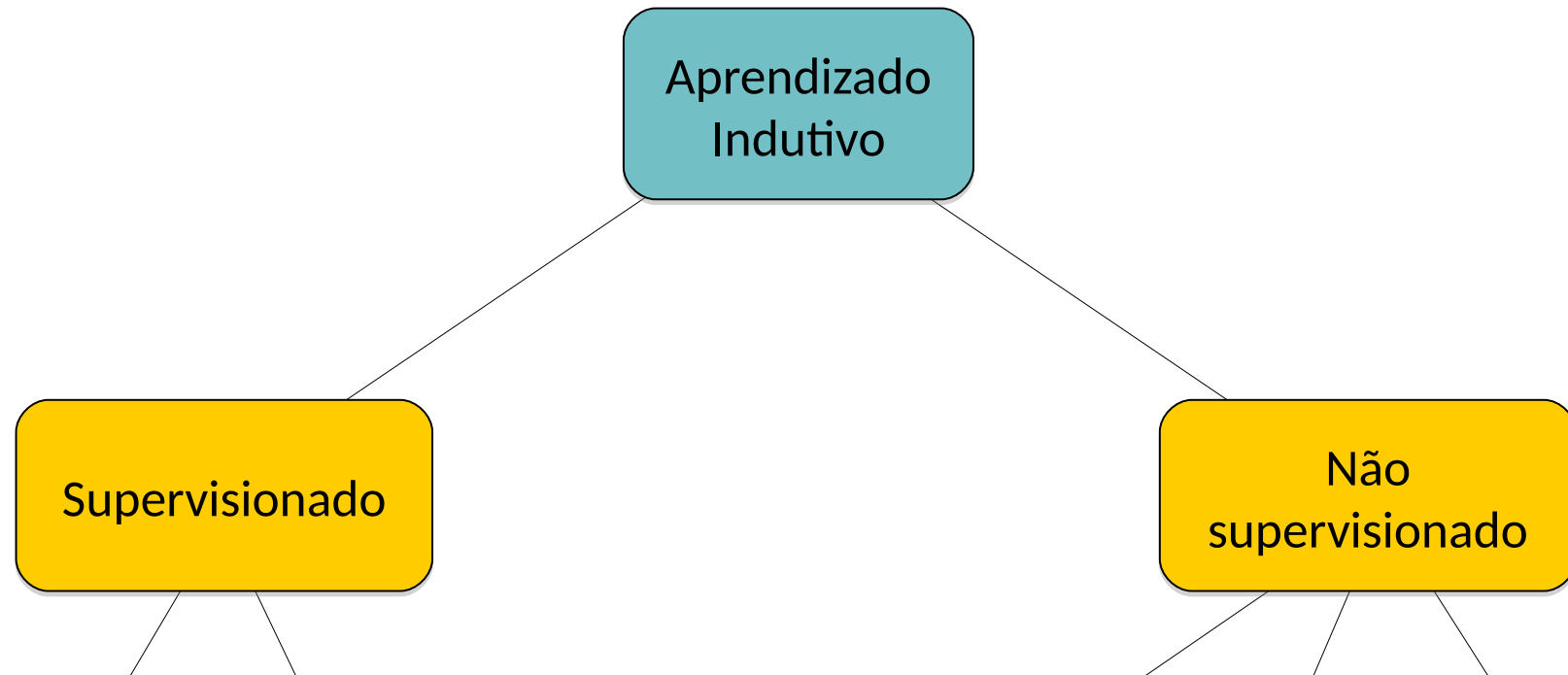
Regressão

Agrupamento

Redução de
dimensionalidade

Redução de
dimensionalidade

Modalidades de AM



Qual redução é mais adequada para a tarefa T?

Classificação

Regressão

Agrupamento

Redução de
dimensionalidade

Redução de
dimensionalidade

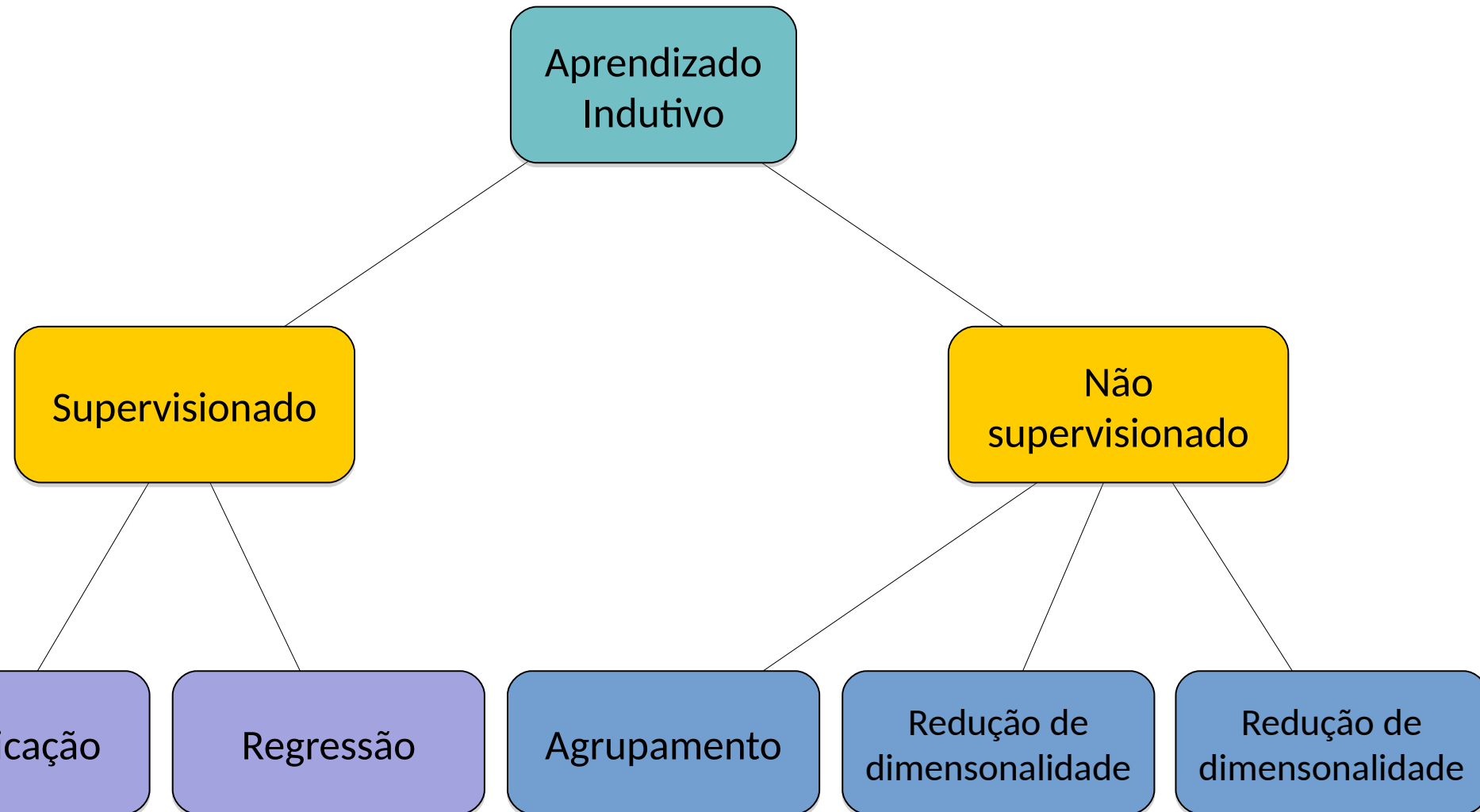
Como se faz AM?

- O que define um bom especialista de AM?
 - Saber identificar uma tarefa de aprendizado
 - Saber reformular essa tarefa em termos de problemas de AM com soluções conhecidas
 - Saber avaliar o desempenho da solução obtida
 - Conhecer e saber empregar algoritmos de AM
 - Para processar/prepara os dados, obter a solução, testar a solução e convencer o cliente de que a solução é boa

O que precisamos saber?

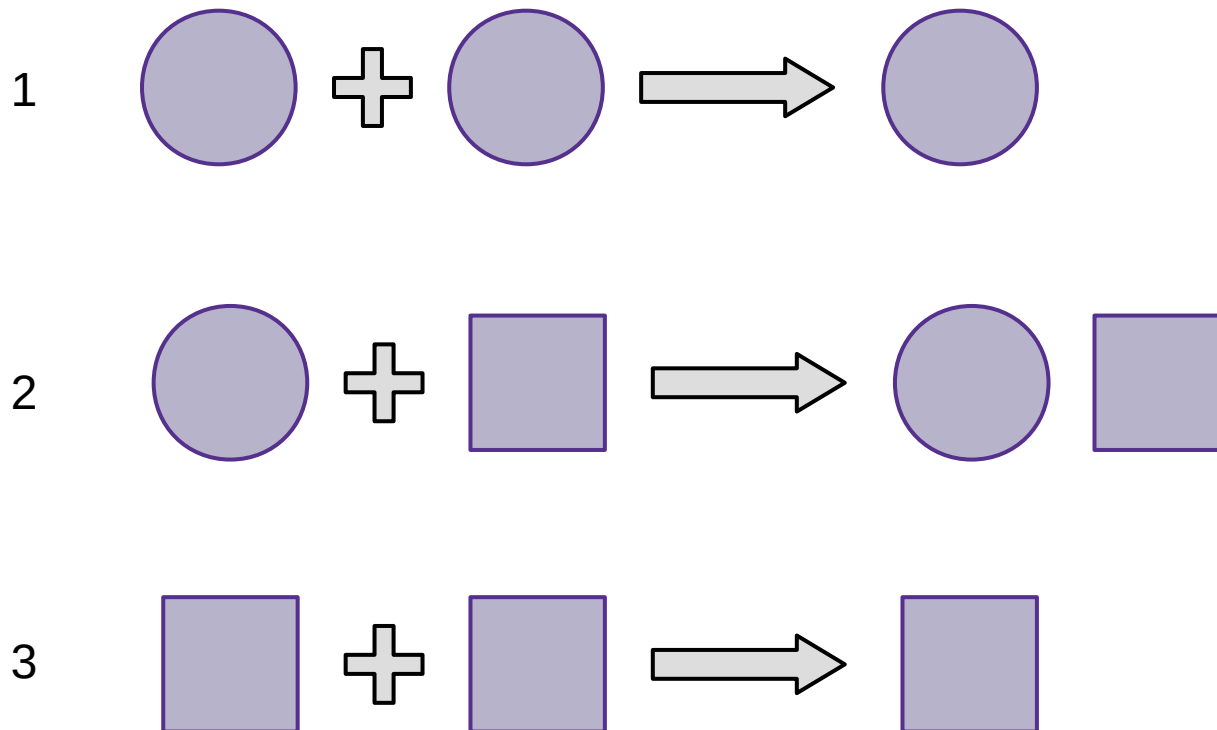
- Pra que serve aprendizado de máquina?
- O que é aprendizado de máquina?
- Qual é a definição de aprendizado?
- Como se faz aprendizado de máquina?
- O que é **aprendizado indutivo**?
- O que é um modelo?
- Como se encontra um modelo?

Modalidades de AM



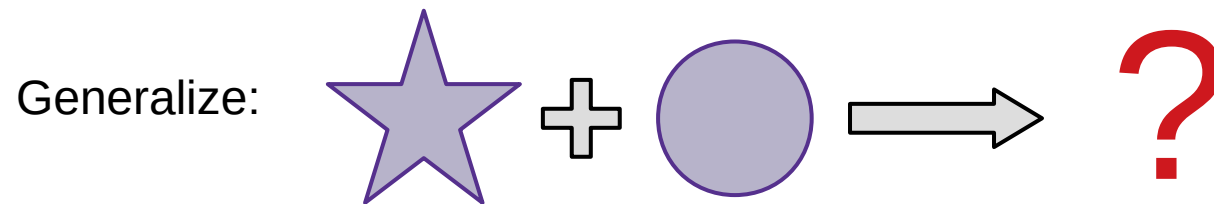
Aprendizado indutivo

- Tem base no raciocínio indutivo
 - Generalização de um conceito com base na observação



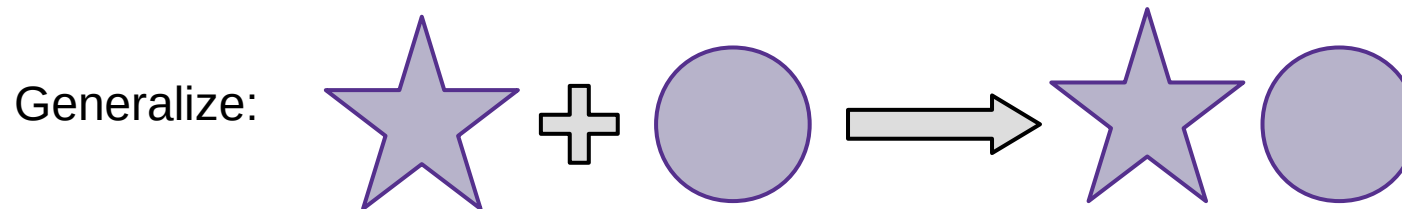
Aprendizado indutivo

- Tem base no raciocínio indutivo
 - Generalização de um conceito com base na observação



Aprendizado indutivo

- Tem base no raciocínio indutivo
 - Generalização de um conceito com base na observação



Aprendizado indutivo

- Como e quando é possível aprender através de indução?
 - Assume-se que a tarefa **T** envolve um **conceito**
 - Assume-se que esse conceito pode ser representado através de **dados**
 - Um registro de motoristas dirigindo corretamente
 - Um registro de jogos de Go
 - Um registro de textos com caligrafia anotada

Aprendizado indutivo

- Como e quando é possível aprender através de indução?
 - Assume-se que existe uma **função-conceito** que representa matematicamente um mapeamento dos **dados** para o **conceito**
 - **Dados**: uma descrição do tabuleiro de Go e de todas as jogadas feitas até o momento
 - **Função-conceito**: um mapeamento desses dados para a melhor jogada a ser feita

Aprendizado indutivo

- Como e quando é possível aprender através de indução?
 - Assume-se que existe uma **função-conceito** que representa matematicamente um mapeamento dos **dados** para o **conceito**
 - **Dados**: uma descrição da posição de todos os objetos na estrada e da velocidade e direção do veículo
 - **Função-conceito**: um mapeamento desses dados para um conjunto de ações que devem ser tomadas para dirigir corretamente o veículo

Aprendizado indutivo

- Como e quando é possível aprender através de indução?
 - Assume-se que existe uma **função-conceito** que representa matematicamente um mapeamento dos **dados** para o **conceito**
 - **Dados**: uma foto
 - **Função-conceito**: um mapeamento da foto para o conjunto de animais que estão registrados nela

Aprendizado indutivo

- Normalmente não é possível aprender a função-conceito
 - Não temos uma **quantidade suficiente** de dados para induzir essa função
 - Ou não sabemos exatamente **quais dados** representam essa função
 - Pode **não existir** um mapeamento exato dos dados para o conceito
 - O conceito pode **mudar com o tempo**

Aprendizado indutivo

- Então o objetivo prático do aprendizado indutivo é obter um **modelo** da função-conceito

O que precisamos saber?

- Pra que serve aprendizado de máquina?
- O que é aprendizado de máquina?
- Qual é a definição de aprendizado?
- Como se faz aprendizado de máquina?
- O que é aprendizado indutivo?
- O que é um **modelo**?
- Como se encontra um modelo?

Modelo

- O objetivo prático do aprendizado indutivo é aprender uma **aproximação** da função-conceito
 - Essa aproximação é chamada **modelo**
 - (Ou *hipótese*, dependendo do contexto)
 - O modelo é obtido através de uma **amostra** dos dados
 - Essa amostra é composta por **exemplos de treinamento**

Modelo

- Hipótese fundamental do aprendizado indutivo (Tom Mitchell)
 - *Qualquer hipótese que aproxime razoavelmente a função-conceito para um conjunto suficientemente grande de exemplos de treinamento também irá aproximar razoavelmente bem a função-conceito para exemplos nunca observados.*

Modelo

- Hipótese fundamental do aprendizado indutivo (Tom Mitchell)
 - *Qualquer hipótese que aproxime razoavelmente a função-conceito para um conjunto suficientemente grande de exemplos de treinamento também irá aproximar razoavelmente bem a função-conceito para exemplos nunca observados.*

A amostra deve ser **representativa** da tarefa T que desejamos aprender

Modelo

- Hipótese fundamental do aprendizado indutivo (Tom Mitchell)
 - *Qualquer hipótese que **aproxime razoavelmente** a função-conceito para um conjunto suficientemente grande de exemplos de treinamento também irá aproximar razoavelmente bem a função-conceito para exemplos nunca observados.*

É necessário estabelecer uma experiência **E** e uma métrica **P** para avaliar a qualidade do modelo

Modelo

- Hipótese fundamental do aprendizado indutivo (Tom Mitchell)
 - *Qualquer hipótese* que aproxime razoavelmente a função-conceito para um conjunto suficientemente grande de exemplos de treinamento também irá aproximar razoavelmente bem a função-conceito para exemplos nunca observados.

A experiência de aprendizado **E** deve nos permitir selecionar um modelo

Modelo

- Hipótese fundamental do aprendizado indutivo (Tom Mitchell)
 - *Qualquer hipótese que aproxime razoavelmente a função-conceito para um conjunto suficientemente grande de exemplos de treinamento também irá aproximar razoavelmente bem a função-conceito para exemplos nunca observados.*

O aprendizado bem sucedido resulta em um modelo capaz de **generalizar**

O que precisamos saber?

- Pra que serve aprendizado de máquina?
- O que é aprendizado de máquina?
- Qual é a definição de aprendizado?
- Como se faz aprendizado de máquina?
- O que é aprendizado indutivo?
- O que é um modelo?
- **Como se encontra um modelo?**

Análise dos dados

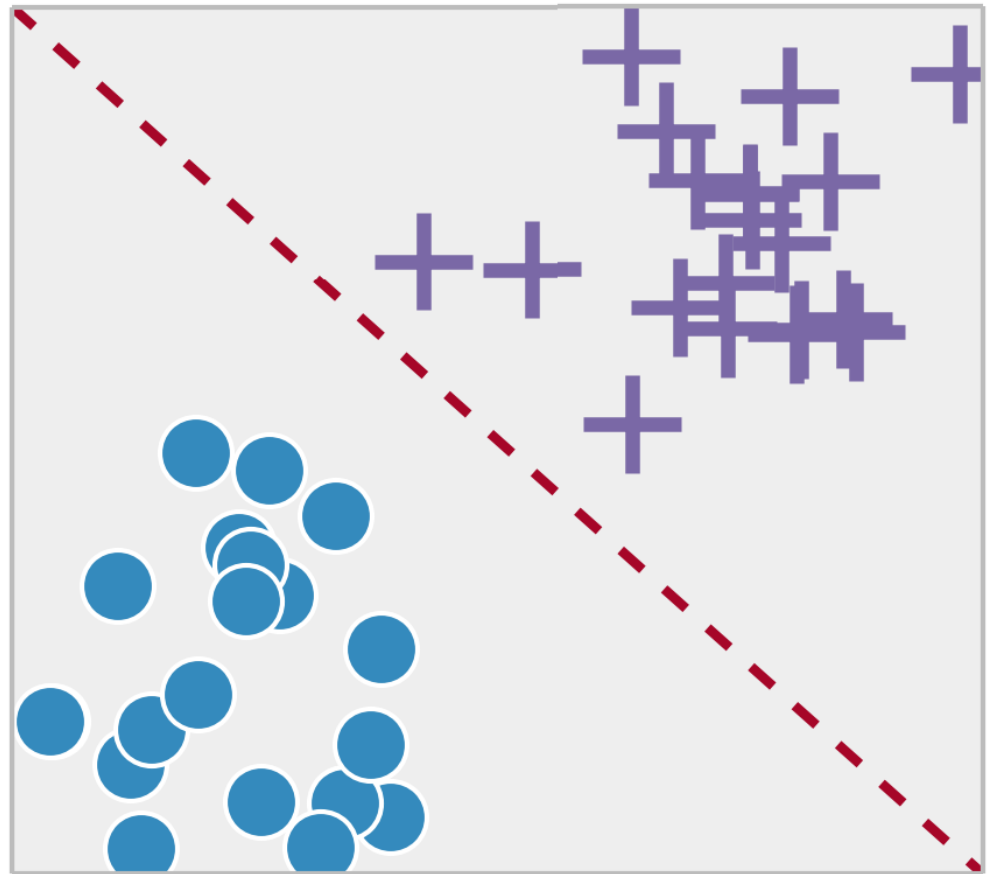
- Precisamos definir nossa tarefa **T** e como a experiência **E** pode ser representada em dados
- Precisamos explorar esses dados e entender nosso **domínio de aplicação**
 - Os dados possuem características que tornam certas **classes de modelo** mais ou menos adequadas
 - A natureza da tarefa **T** torna certas **classes de modelo** mais ou menos adequadas

Classes de modelo

- Como saber quando um modelo é mais ou menos adequado que outro?
 - Uma pergunta em geral sem resposta clara
 - Mas existem heurísticas
 - Alguns modelos são adequados para classificação
 - Outros para regressão
 - Outros para agrupamento

Classes de modelo

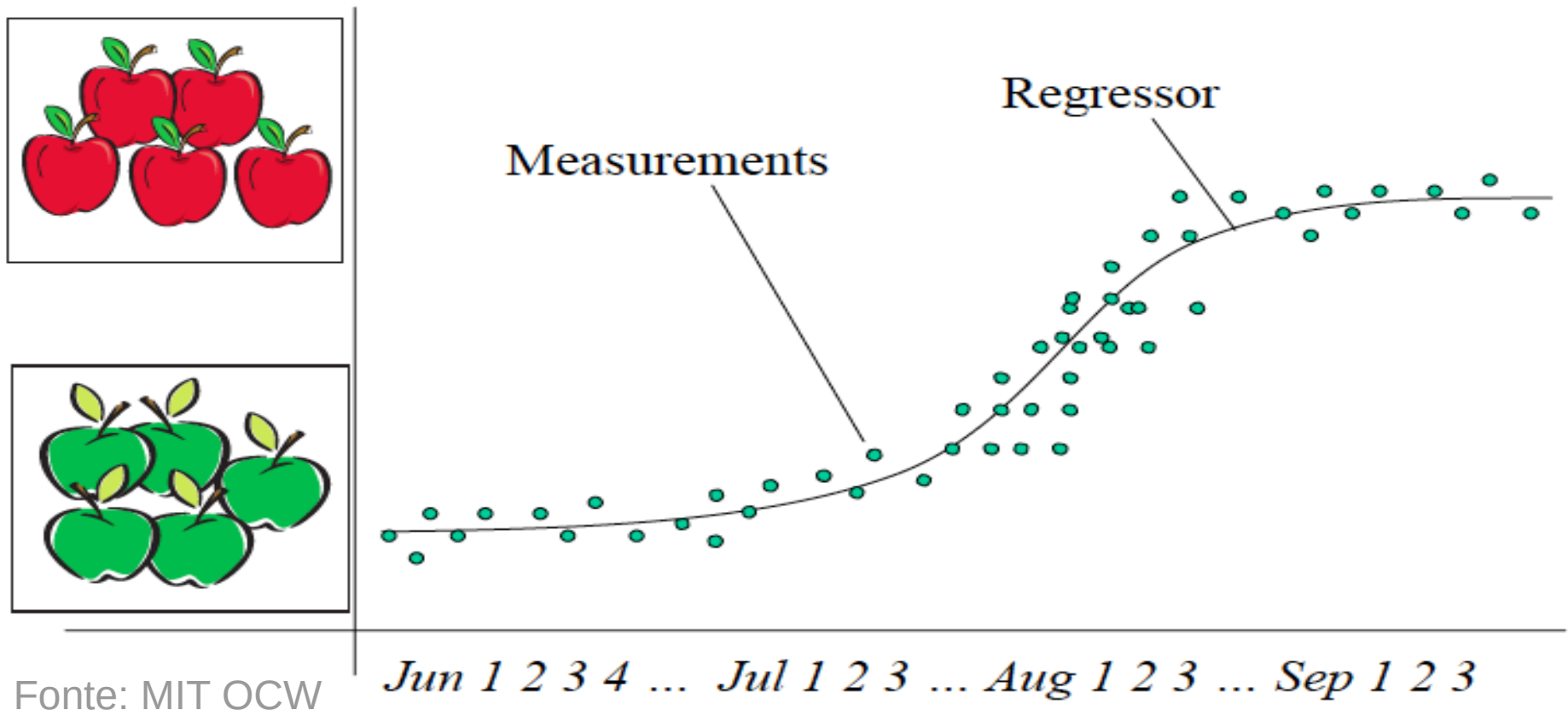
- Classificação
 - O modelo identifica os dados como pertencentes a categorias conhecidas



Fonte: <http://ipython-books.github.io/featured-04/>

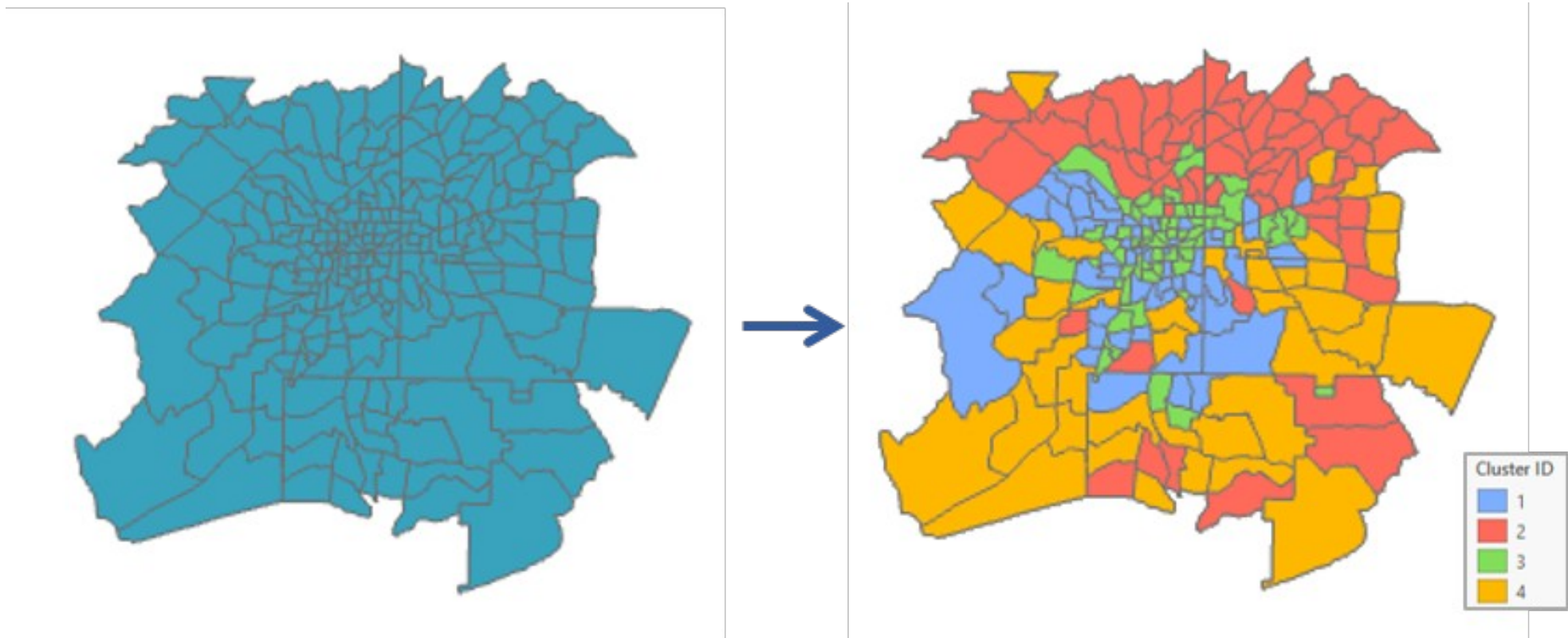
Classes de modelo

- Regressão
 - O modelo extrapola um valor para os dados



Classes de modelo

- Agrupamento
 - O modelo agrupa os dados em categorias

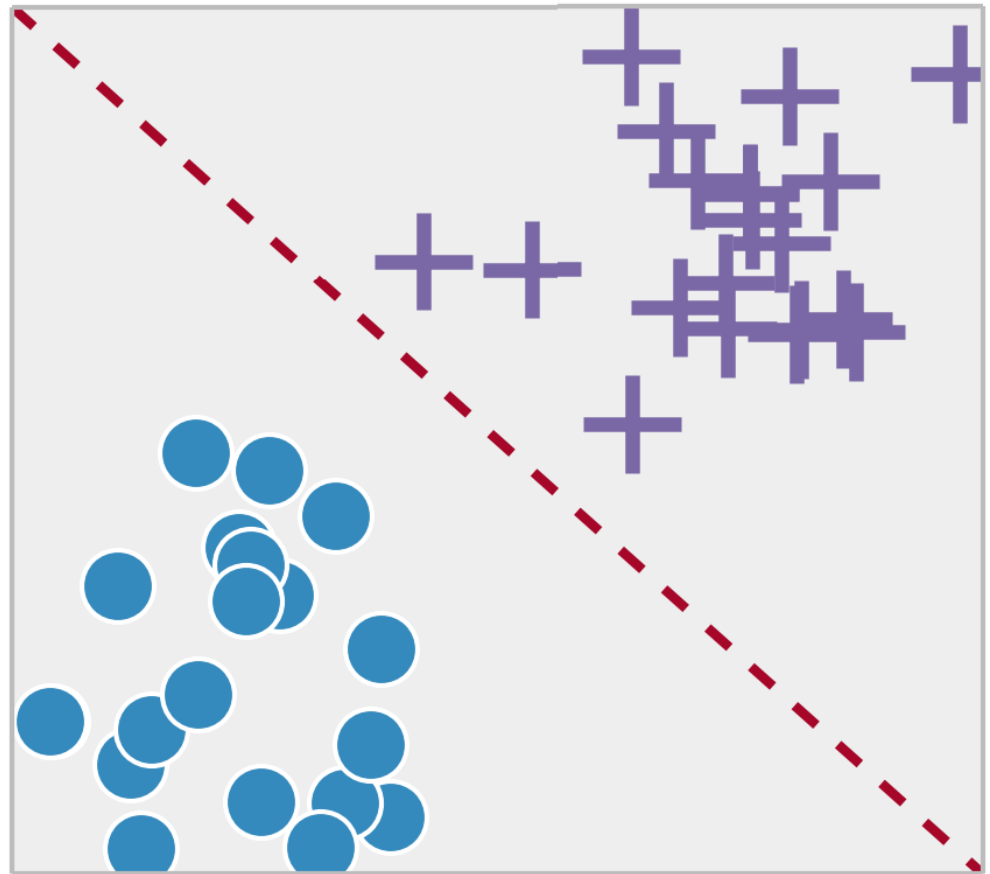


Linguagem de descrição

- Como saber quando um modelo é mais ou menos adequado que outro?
 - Heurística: linguagem de descrição
 - Cada classe de modelo aproxima a função-conceito de uma maneira diferente
 - Cada modelo tem uma **linguagem** para descrever o modelo
 - Algumas linguagens são em geral mais ou menos adequadas para certos tipos de dados

Linguagem de descrição

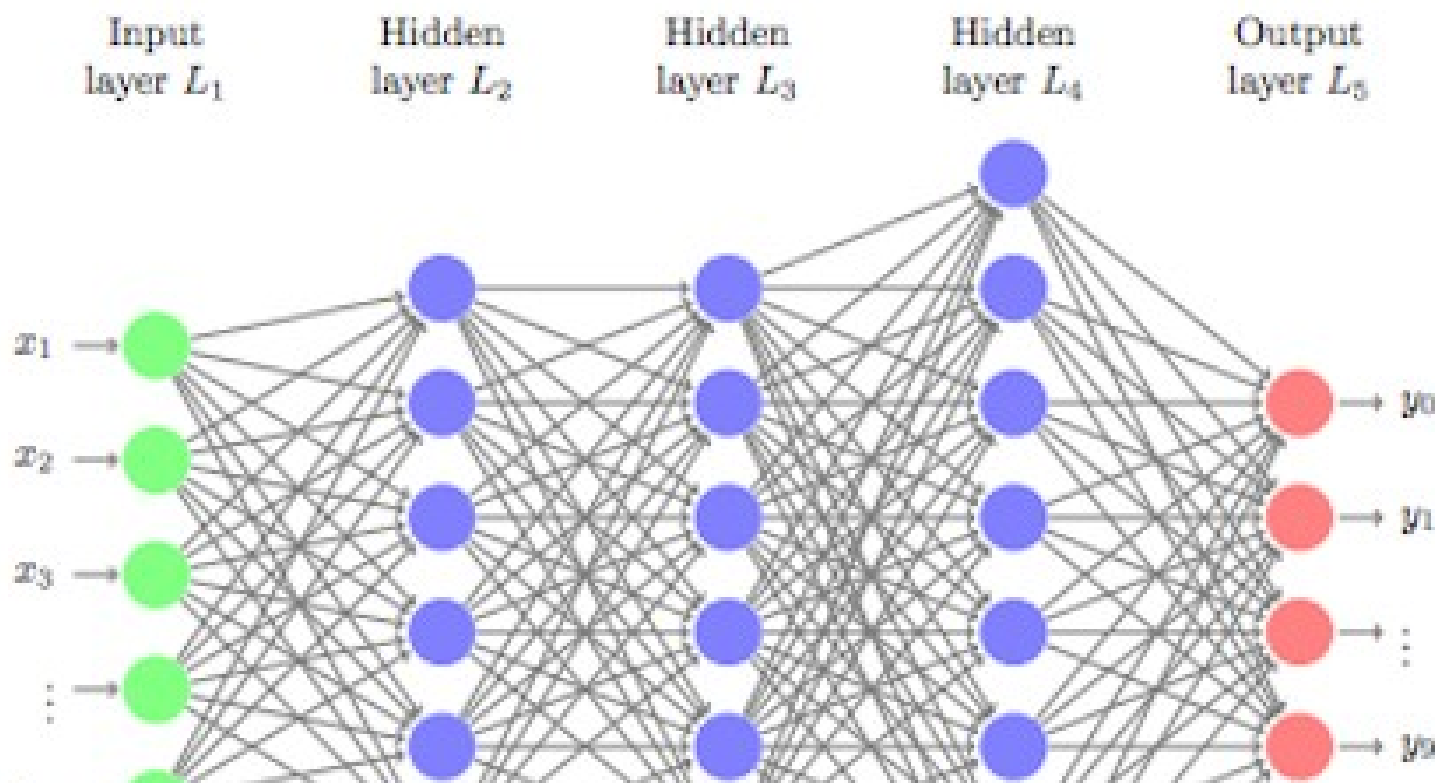
- Modelos de separação linear
 - Assumem que os dados podem ser geometricamente separados com um hiperplano



Fonte: <http://ipython-books.github.io/featured-04/>

Linguagem de descrição

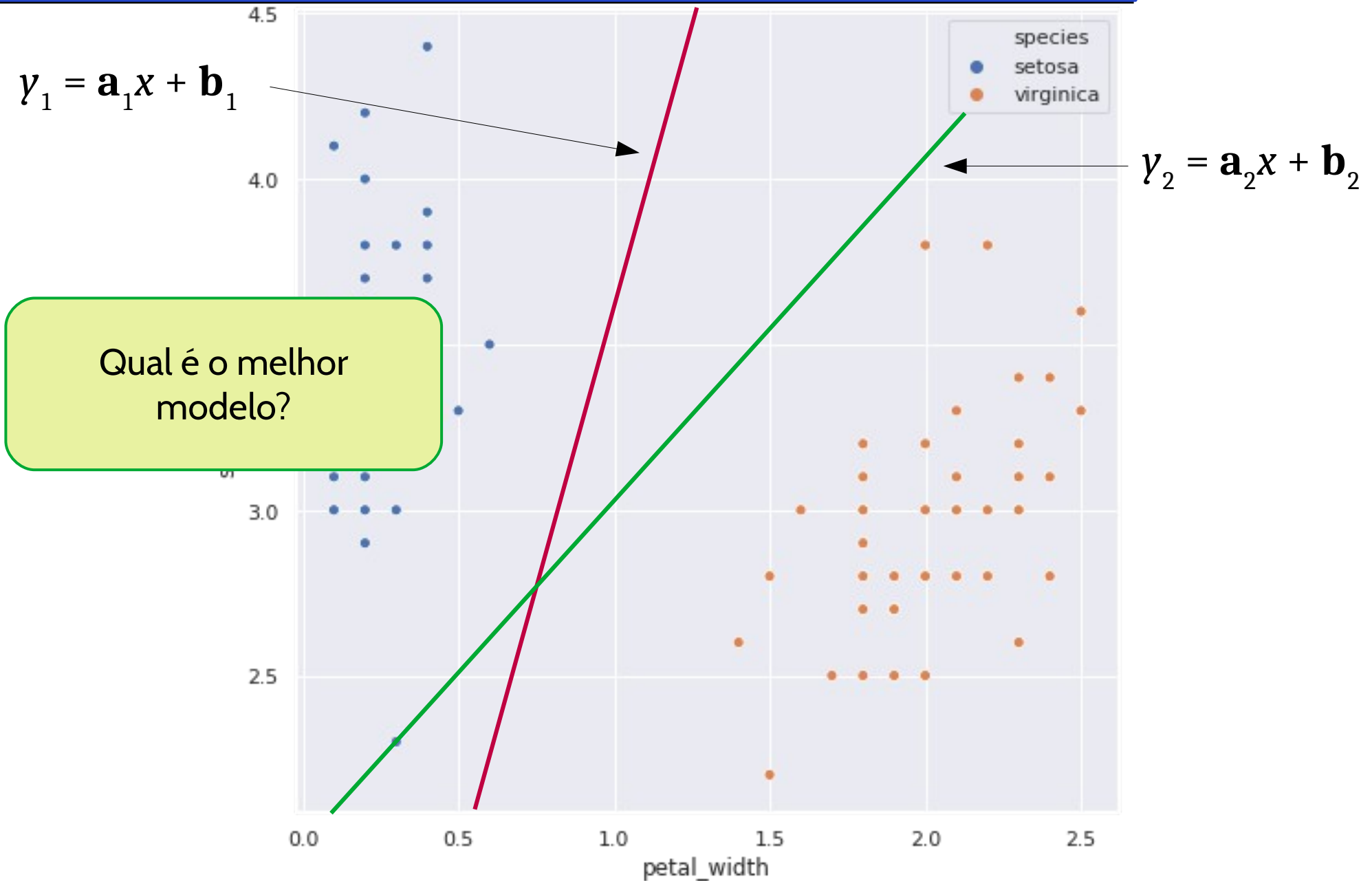
- Redes neurais profundas
 - Aprendem uma relação matemática entre os dados
 - Exigem **grande volume de dados**



Parâmetros e hiperparâmetros

- Para empregar um modelo em um domínio de aplicação, é necessário gerar uma **instância** do modelo
- Isso é feito encontrando um conjunto de **parâmetros** mais adequados para o domínio de aplicação

Parâmetros e hiperparâmetros



Parâmetros e hiperparâmetros

- De acordo com a hipótese fundamental do aprendizado indutivo
 - O melhor modelo é o mais adequado para os exemplos de treinamento
 - O trabalho de encontrar esse modelo é realizado por um algoritmo denominado **indutor**

Parâmetros e hiperparâmetros

- O trabalho do inutor é **buscar** os parâmetros ideais do modelo para um domínio de aplicação
- Utilizando...
 - Os exemplos de treinamento como uma referência do domínio
 - Um conjunto de **preferências** e de **hiperparâmetros** para determinar como essa **busca** deve ser realizada

Aprendizado como busca

- O aprendizado é encarado como um algoritmo de busca
 - **def** induzir_modelo(dados, classe, hiperparam, threshold):
... modelo = classe.novo(dados, hiperparam)
... **while** (modelo.desempenho < threshold):
... modelo.atualizar(dados, hiperparam)
... **return** modelo

Isto é apenas um exemplo conceitual. Não necessariamente um algoritmo de um indutor.

Aprendizado como busca

- O aprendizado é encarado como um algoritmo de busca
 - E nesse processo de busca o indutor insere suas preferências
 - Essas preferências são denominadas **viés**
 - **Não existe aprendizado sem viés**
 - Teorema "*No Free Lunch*"
 - Falaremos mais sobre isso nas próximas aulas

Nas próximas aulas...

- A(s) próxima(s) aula(s) serão sobre
 - Análise de dados
 - Utilizando Python, Jupyter *notebooks* e as bibliotecas NumPy, Pandas e PyPlot
 - Os computadores do Laboratório de Graduação 1 estão instalados e corretamente configurados
 - Recomendo instalar no seu *desktop* ou *laptop*
 - Sugestão: Anaconda

No restante do curso...

- O restante do curso será inteiramente dedicado ao estudo de
 - Classes de modelos
 - Algoritmos de indução
 - Algoritmos de validação
 - Técnicas de pré-processamento