

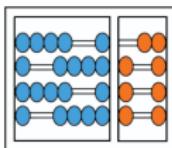
Introdução ao Aprendizado de Máquina para Bioinformática

Brazilian Symposium on Bioinformatics (BSB) 2023

Cristiano Gabriel de Souza Campos^{1,2},
Ronaldo Nogueira de Sousa^{3,4}

- ¹ Pós-graduação em C. Computação; ² Laboratório Recod.ai,
Instituto de Computação, Universidade Estadual de Campinas
³ Center of Toxins, Immune-response and Cell Signaling (CeTICS), I. Butantan
⁴ Pós-graduação em Bioinformática, Universidade de São Paulo

Curitiba, 14 de junho de 2023



**Instituto de
Computação**
UNIVERSIDADE ESTADUAL DE CAMPINAS



A UNICAMP



Antoninho Perri/Unicamp

A UNICAMP



Antoninho Perri/Unicamp

A IME-USP

- ▶ Programa Interunidades de Pós-graduação em Bioinformática
- ▶ Curso de Verão em Bioinformática USP



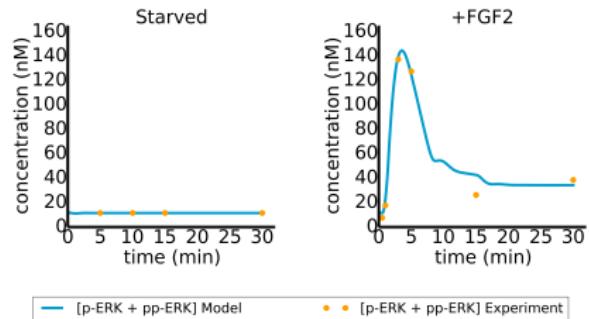
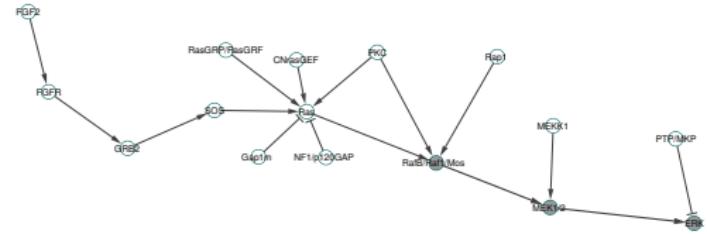
O Instituto Butantan



- ▶ Importante produtor de soros e vacinas
- ▶ É também um centro de pesquisas biomédicas

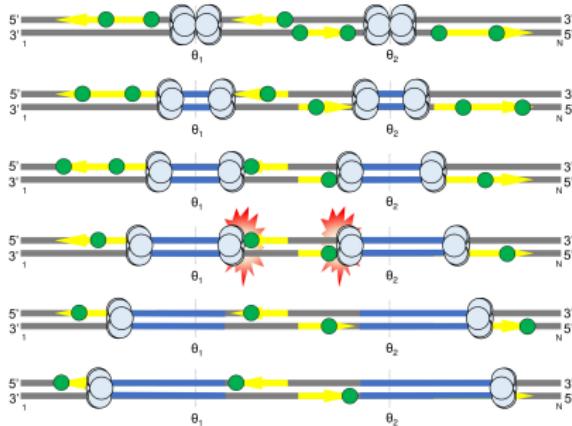


Nossas linhas de pesquisa em Bioinfo

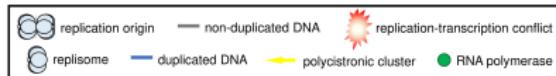


Identificação de vias de sinalização celular

Nossas linhas de pesquisa em Bioinfo



Modelagem dinâmica da replicação de DNA



Resumo de hoje:

Conceitos iniciais (definições, paradigmas, etc.)

Aprendizado Supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

Aprendizado Não-supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

E agora?

Sessão de perguntas e respostas (Q&A)

Resumo de hoje:

Conceitos iniciais (definições, paradigmas, etc.)

Aprendizado Supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

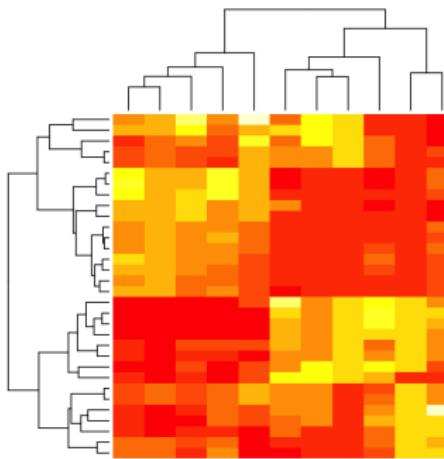
Aprendizado Não-supervisionado

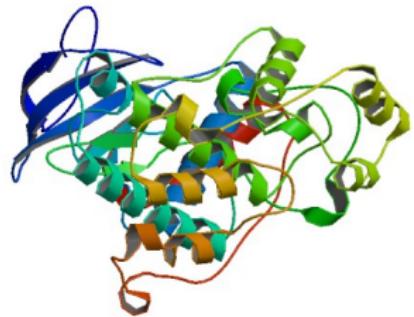
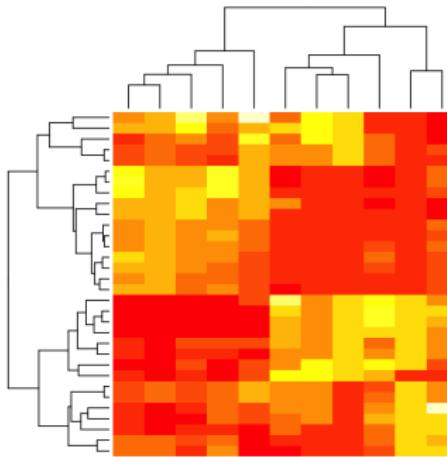
Alguma fundamentação teórica

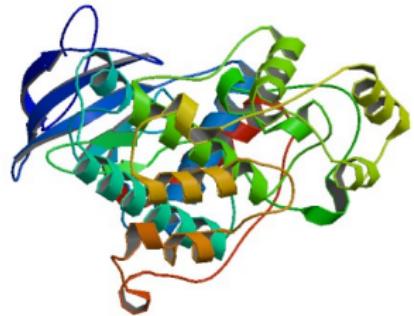
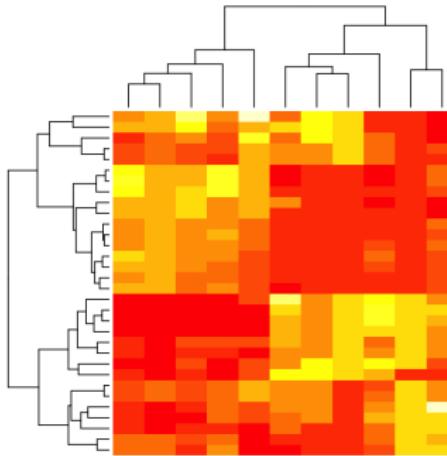
Parte prática (Colab)

E agora?

Sessão de perguntas e respostas (Q&A)







Relação de Aprendizado de Máquina com outras áreas

Da Wikipedia:

Relação de Aprendizado de Máquina com outras áreas

Da Wikipedia:

- ▶ “**Inteligência Artificial** é a aparente inteligência demonstrada por máquinas, ao contrário da inteligência natural de humanos e animais”

Relação de Aprendizado de Máquina com outras áreas

Da Wikipedia:

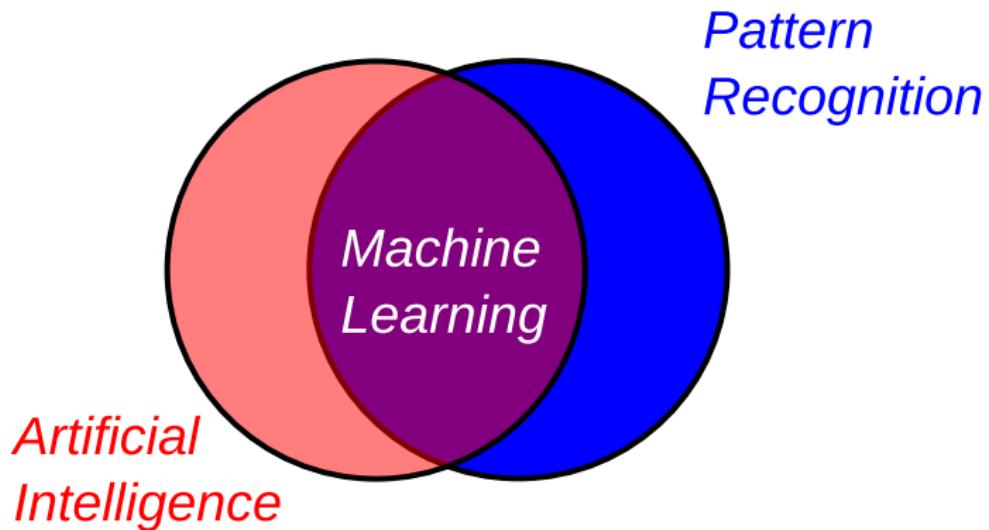
- ▶ “**Inteligência Artificial** é a aparente inteligência demonstrada por máquinas, ao contrário da inteligência natural de humanos e animais”

- ▶ “**Reconhecimento de Padrões** é o reconhecimento automático de padrões (regularidades) em dados”

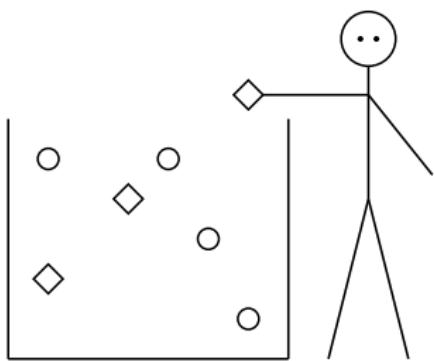
Relação de Aprendizado de Máquina com outras áreas

Da Wikipedia:

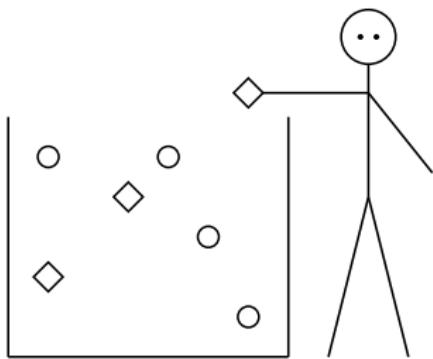
- ▶ “**Inteligência Artificial** é a aparente inteligência demonstrada por máquinas, ao contrário da inteligência natural de humanos e animais”
- ▶ “**Reconhecimento de Padrões** é o reconhecimento automático de padrões (regularidades) em dados”
- ▶ “**Aprendizado de Máquina** é a área de estudo de modelos que “aprendem” a partir de dados”



Exemplo

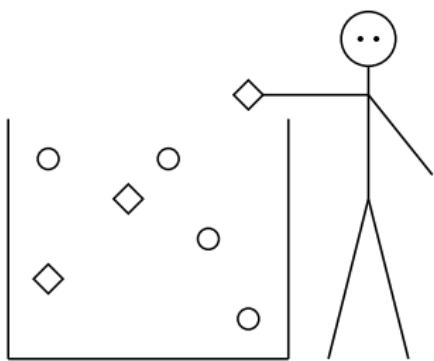


Exemplo

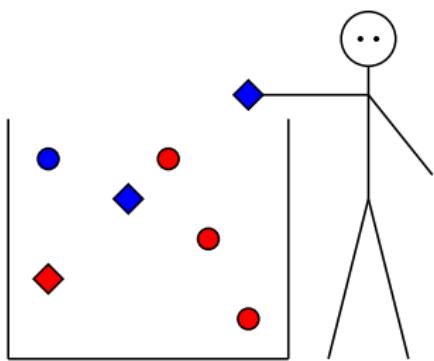


<i>circle</i>	4
<i>diamond</i>	3

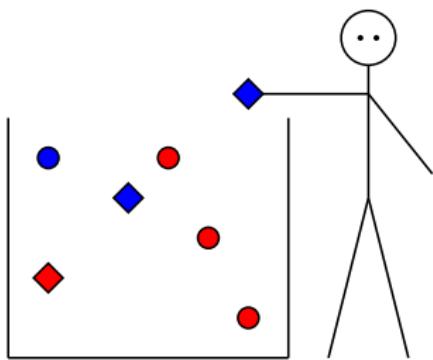
Exemplo



Exemplo

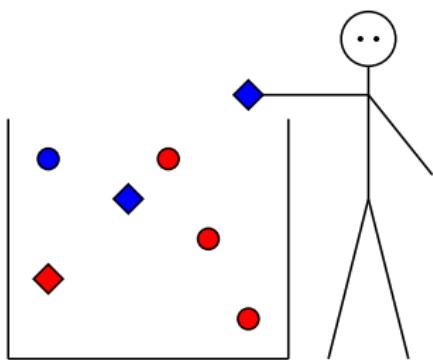


Exemplo



	<i>blue</i>	<i>red</i>
<i>circle</i>	1	3
<i>diamond</i>	2	1

Exemplo

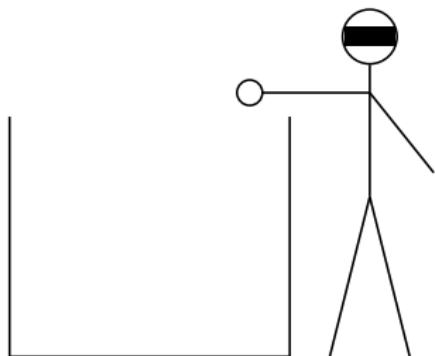


$$h(\textcircled{O}) = \text{red}$$

$$h(\textdiamondsuit) = \text{blue}$$

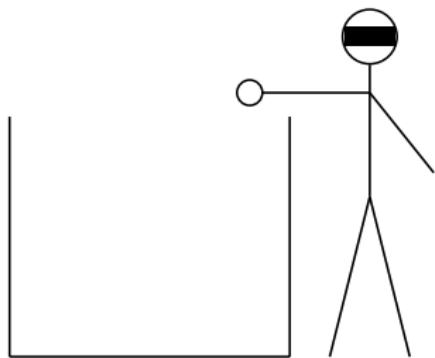
	blue	red
circle	1	3
diamond	2	1

Usando o classificador treinado



$$h(\textcircled{O}) = ?$$

Usando o classificador treinado



$$h(\text{ } \circ \text{ }) = \text{red}$$

Paradigmas de Aprendizado de Máquina

Paradigmas de Aprendizado de Máquina

1. Aprendizado Supervisionado

Paradigmas de Aprendizado de Máquina

1. Aprendizado Supervisionado : *(atributo, resposta)*

Paradigmas de Aprendizado de Máquina

1. Aprendizado Supervisionado : *(atributo, resposta)*
2. Aprendizado Não-supervisionado

Paradigmas de Aprendizado de Máquina

1. Aprendizado Supervisionado : *(atributo, resposta)*
2. Aprendizado Não-supervisionado : *atributo*

Paradigmas de Aprendizado de Máquina

1. Aprendizado Supervisionado : *(atributo, resposta)*
2. Aprendizado Não-supervisionado : *atributo*
3. Outros (Por Reforço, Semi-supervisionado, etc.)

Princípios básicos em Aprendizado

Princípios básicos em Aprendizado

- ▶ Existe um padrão associado a um conceito

Princípios básicos em Aprendizado

- ▶ Existe um padrão associado a um conceito
- ▶ Não conseguimos descrever matematicamente esse padrão

Princípios básicos em Aprendizado

- ▶ Existe um padrão associado a um conceito
- ▶ Não conseguimos descrever matematicamente esse padrão
- ▶ Temos observações desse padrão (dados)

Resumo de hoje:

Conceitos iniciais (definições, paradigmas, etc.)

Aprendizado Supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

Aprendizado Não-supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

E agora?

Sessão de perguntas e respostas (Q&A)

Um exemplo de Aprendizado Supervisionado em Bioinfo

- ▶ Suponha que queremos classificar um conjunto de sequências proteicas de acordo com o fato delas serem ou não exportadas para fora da célula

Um exemplo de Aprendizado Supervisionado em Bioinfo

- ▶ Suponha que queremos classificar um conjunto de sequências proteicas de acordo com o fato delas serem ou não exportadas para fora da célula
- ▶ Nossos dados são sequências de aminoácidos:

Um exemplo de Aprendizado Supervisionado em Bioinfo

- ▶ Suponha que queremos classificar um conjunto de sequências proteicas de acordo com o fato delas serem ou não exportadas para fora da célula
- ▶ Nossos dados são sequências de aminoácidos:
 1. Ala-Cys-Gln-Pro-Thr-Ala-Ala-Tyr-Val-Met...

Um exemplo de Aprendizado Supervisionado em Bioinfo

- ▶ Suponha que queremos classificar um conjunto de sequências proteicas de acordo com o fato delas serem ou não exportadas para fora da célula
- ▶ Nossos dados são sequências de aminoácidos:
 1. Ala-Cys-Gln-Pro-Thr-Ala-Ala-Tyr-Val-Met...
 2. Cys-Tyr-Gln-Lys-Thr-Ala-Leu-Ley-Met-His...

Um exemplo de Aprendizado Supervisionado em Bioinfo

- ▶ Suponha que queremos classificar um conjunto de sequências proteicas de acordo com o fato delas serem ou não exportadas para fora da célula
- ▶ Nossos dados são sequências de aminoácidos:
 1. Ala-Cys-Gln-Pro-Thr-Ala-Ala-Tyr-Val-Met...
 2. Cys-Tyr-Gln-Lys-Thr-Ala-Leu-Ley-Met-His...
 3. Ala-Cys-Gln-Asp-Thr-Gly-Met-His-Val-Met...

Um exemplo de Aprendizado Supervisionado em Bioinfo

- ▶ Suponha que queremos classificar um conjunto de sequências proteicas de acordo com o fato delas serem ou não exportadas para fora da célula
- ▶ Nossos dados são sequências de aminoácidos:
 1. Ala-Cys-Gln-Pro-Thr-Ala-Ala-Tyr-Val-Met...
 2. Cys-Tyr-Gln-Lys-Thr-Ala-Leu-Ley-Met-His...
 3. Ala-Cys-Gln-Asp-Thr-Gly-Met-His-Val-Met...

: : : : :

Extração de características

- ▶ Primeira abordagem: usar todas as sequências

Extração de características

- ▶ Primeira abordagem: usar todas as sequências
- ▶ Todavia, suponha que proteínas exportadas tenham presente um padrão em uma determinada localização

Extração de características

- ▶ Primeira abordagem: usar todas as sequências
- ▶ Todavia, suponha que proteínas exportadas tenham presente um padrão em uma determinada localização
- ▶ Além disso, o padrão tem uma carga média

Extração de características

- ▶ Primeira abordagem: usar todas as sequências
- ▶ Todavia, suponha que proteínas exportadas tenham presente um padrão em uma determinada localização
- ▶ Além disso, o padrão tem uma carga média

Ala-Cys-Gln-Asp-Thr-Gly-Met-His-Val-Met...

Extração de características

- ▶ Primeira abordagem: usar todas as sequências
- ▶ Todavia, suponha que proteínas exportadas tenham presente um padrão em uma determinada localização
- ▶ Além disso, o padrão tem uma carga média

Ala-Cys-Gln-Asp-Thr-**Gly-Met-His-Val-Met...**

Extração de características

- ▶ Primeira abordagem: usar todas as sequências
- ▶ Todavia, suponha que proteínas exportadas tenham presente um padrão em uma determinada localização
- ▶ Além disso, o padrão tem uma carga média

Ala-Cys-Gln-Asp-Thr-**Gly-Met-His-Val-Met...**

- ▶ Localização: 6

Extração de características

- ▶ Primeira abordagem: usar todas as sequências
- ▶ Todavia, suponha que proteínas exportadas tenham presente um padrão em uma determinada localização
- ▶ Além disso, o padrão tem uma carga média

Ala-Cys-Gln-Asp-Thr-**Gly-Met-His-Val-Met...**

- ▶ Localização: 6
- ▶ Carga: 0.33

Definindo o conjunto de treinamento

Definindo o conjunto de treinamento

- ▶ Temos então dois atributos:

Definindo o conjunto de treinamento

- ▶ Temos então dois atributos:
 - ▶ Localização do padrão

Definindo o conjunto de treinamento

- ▶ Temos então dois atributos:

- ▶ Localização do padrão
- ▶ Carga do padrão

Definindo o conjunto de treinamento

- ▶ Temos então dois atributos:
 - ▶ Localização do padrão
 - ▶ Carga do padrão
- ▶ Para cada proteína, temos um par $\underbrace{(\textit{localization}, \textit{charge})}_{\textit{atributos}}$

Definindo o conjunto de treinamento

- ▶ Temos então dois atributos:
 - ▶ Localização do padrão
 - ▶ Carga do padrão
- ▶ Para cada proteína, temos um par $\underbrace{(\textit{localization}, \textit{charge})}_{\textit{atributos}}$
- ▶ Para cada proteína que conhecemos (através de experimentos) nós definimos:

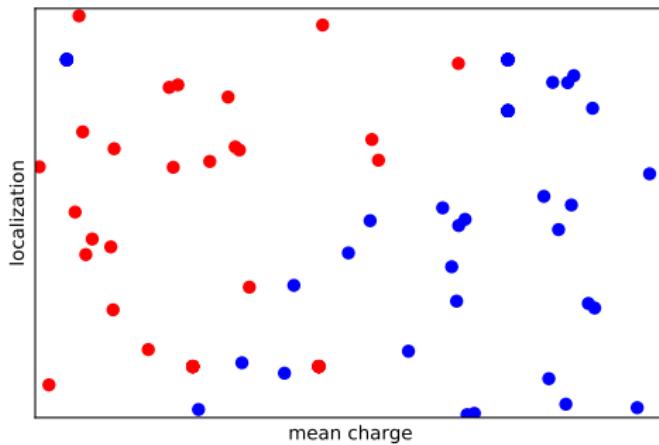
Definindo o conjunto de treinamento

- ▶ Temos então dois atributos:
 - ▶ Localização do padrão
 - ▶ Carga do padrão
- ▶ Para cada proteína, temos um par $\underbrace{(\textit{localization}, \textit{charge})}_{\text{atributos}}$
- ▶ Para cada proteína que conhecemos (através de experimentos) nós definimos:

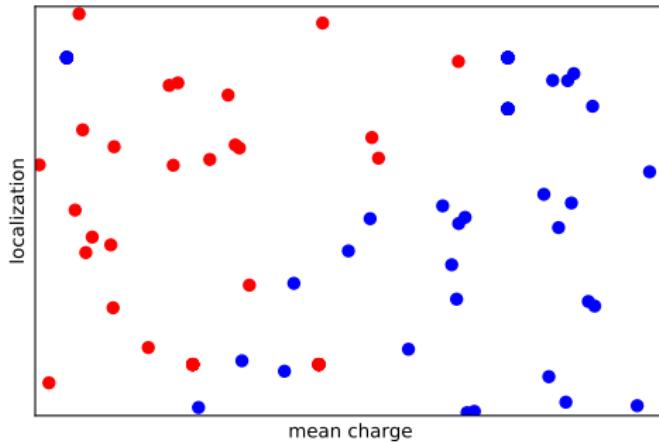
$$\textit{label} = \begin{cases} \textcolor{red}{\text{red}}, & \text{se a proteína é exportada} \\ \textcolor{blue}{\text{blue}}, & \text{caso contrário} \end{cases}$$

Desenhand o classificador

Desenhandando o classificador

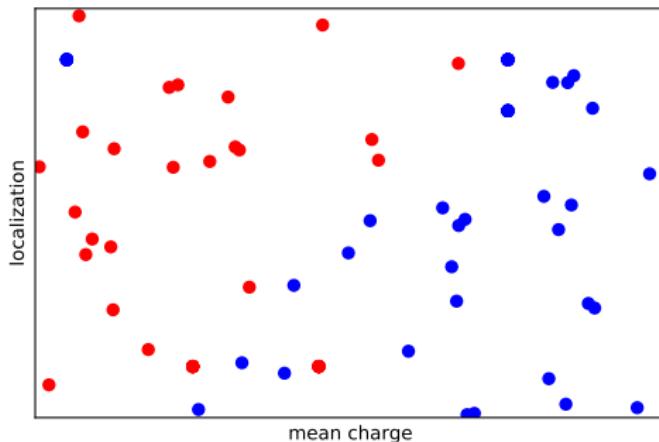


Desenhandando o classificador



- ▶ Como realizar a classificação?

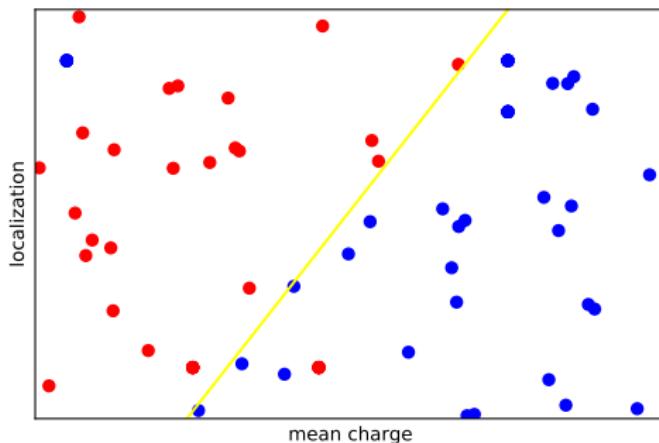
Desenhando o classificador



- ▶ Como realizar a classificação?
- ▶ Podemos tentar encontrar uma função:

$$h(\textit{localization}, \textit{charge}) = \textit{label}$$

Desenhando o classificador



- ▶ Como realizar a classificação?
- ▶ Podemos tentar encontrar uma função:

$$h(\textit{localization}, \textit{charge}) = \textit{label}$$

O que usar em Aprendizado de Máquina?

O que usar em Aprendizado de Máquina?



O que usar em Aprendizado de Máquina?



O que usar em Aprendizado de Máquina?



colab.research.google.com

Parte prática 1: classificadores

Abra em seu COLAB o arquivo:

`ic.unicamp.br/~msreis/Atividade-Pratica-1.ipynb`

E quando o problema é MUITO não-linearmente separável?

E quando o problema é MUITO não-linearmente separável?

- ▶ Transformações não-lineares, SVMs com o truque do kernel

E quando o problema é MUITO não-linearmente separável?

- ▶ Transformações não-lineares, SVMs com o truque do kernel
- ▶ Redes neurais (“Deep learning” está na moda!)

E quando o problema é MUITO não-linearmente separável?

- ▶ Transformações não-lineares, SVMs com o truque do kernel
- ▶ Redes neurais (“Deep learning” está na moda!)
- ▶ k-vizinhos, etc.

E quando o problema é MUITO não-linearmente separável?

- ▶ Transformações não-lineares, SVMs com o truque do kernel
- ▶ Redes neurais (“Deep learning” está na moda!)
- ▶ k-vizinhos, etc.
- ▶ Todavia, o problema pode ainda ter poucos ou nenhum rótulo!

Resumo de hoje:

Conceitos iniciais (definições, paradigmas, etc.)

Aprendizado Supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

Aprendizado Não-supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

E agora?

Sessão de perguntas e respostas (Q&A)

Aprendizado Não-supervisionado

Aprendizado Não-supervisionado

- Em Aprendizado Supervisionado, temos um vetor x de atributos e queremos associá-lo a uma variável resposta

Aprendizado Não-supervisionado

- ▶ Em Aprendizado Supervisionado, temos um vetor x de atributos e queremos associá-lo a uma variável resposta
- ▶ Todavia, em **Aprendizado Não-supervisionado**, avaliamos apenas a distribuição de valores de x

Aprendizado Não-supervisionado

- ▶ Em Aprendizado Supervisionado, temos um vetor x de atributos e queremos associá-lo a uma variável resposta
- ▶ Todavia, em **Aprendizado Não-supervisionado**, avaliamos apenas a distribuição de valores de x
- ▶ Qual é a relevância de se olhar para dados não rotulados?

Razões para Aprendizado Não-supervisionado

Razões para Aprendizado Não-supervisionado

1. Coletar e rotular padrões pode ser custoso em “big data”

Razões para Aprendizado Não-supervisionado

1. Coletar e rotular padrões pode ser custoso em “big data”
2. Padrões podem mudar durante o processo estudado

Razões para Aprendizado Não-supervisionado

1. Coletar e rotular padrões pode ser custoso em “big data”
2. Padrões podem mudar durante o processo estudado
3. Aprendizado Não-supervisionado pode ser utilizado como pré-processamento para o Supervisionado

Razões para Aprendizado Não-supervisionado

1. Coletar e rotular padrões pode ser custoso em “big data”
 2. Padrões podem mudar durante o processo estudado
 3. Aprendizado Não-supervisionado pode ser utilizado como pré-processamento para o Supervisionado
- Exemplo desse tipo de Aprendizado: agrupamento (clustering)

Agrupamento (clustering)

Agrupamento (clustering)

- ▶ Método não-supervisionado em que pontos são agrupados segundo um critério de similaridade

Agrupamento (clustering)

- ▶ Método não-supervisionado em que pontos são agrupados segundo um critério de similaridade
- ▶ Exemplo: agrupar pontos que estão próximos entre si segundo a distância Euclidiana entre eles

Agrupamento (clustering)

- ▶ Método não-supervisionado em que pontos são agrupados segundo um critério de similaridade
- ▶ Exemplo: agrupar pontos que estão próximos entre si segundo a distância Euclidiana entre eles
- ▶ Dois métodos relevantes: k-médias e agrupamento hierárquico

k-médias

k-médias

- ▶ Divide um conjunto de N pontos em k grupos disjuntos, onde cada grupo j é representado pela média μ_j de seus pontos

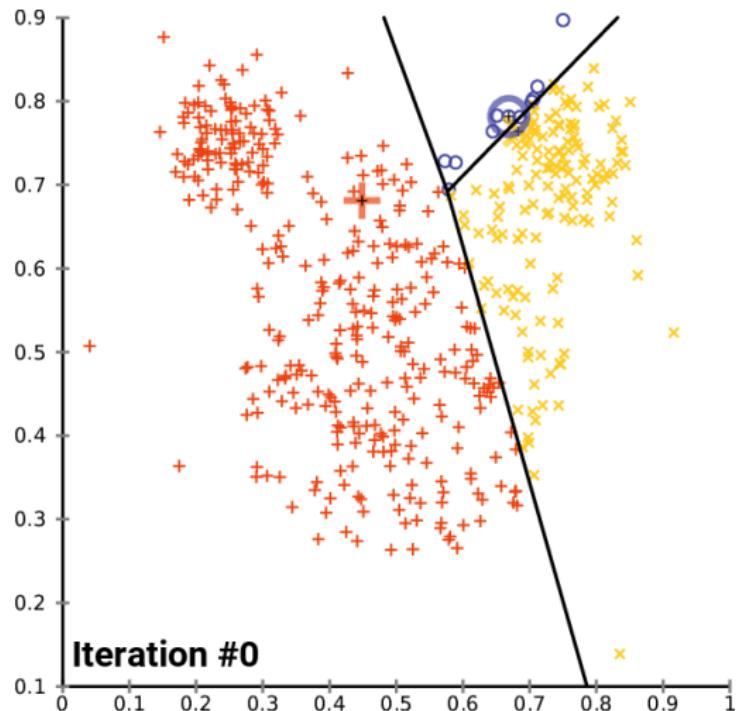
k-médias

- ▶ Divide um conjunto de N pontos em k grupos disjuntos, onde cada grupo j é representado pela média μ_j de seus pontos

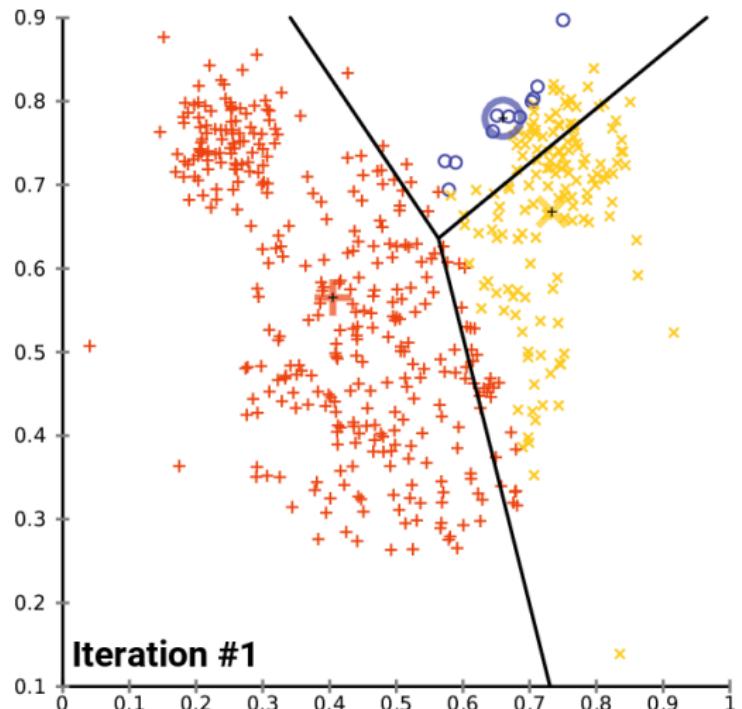
$$\text{minimizar} \sum_{j=1}^k \sum_{x \in C_j} \|x - \mu_j\|^2,$$

onde C_j é o conjunto dos pontos no grupo j

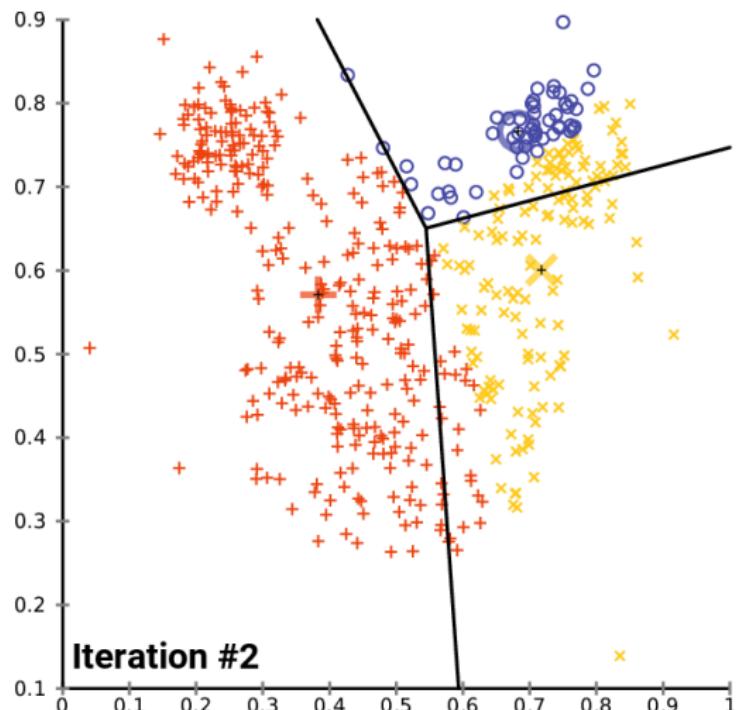
k-médias



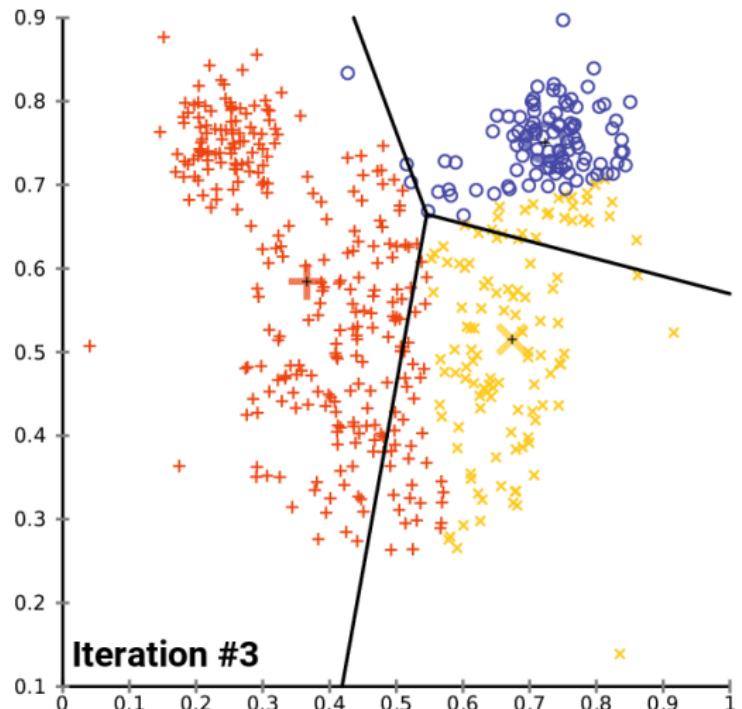
k-médias



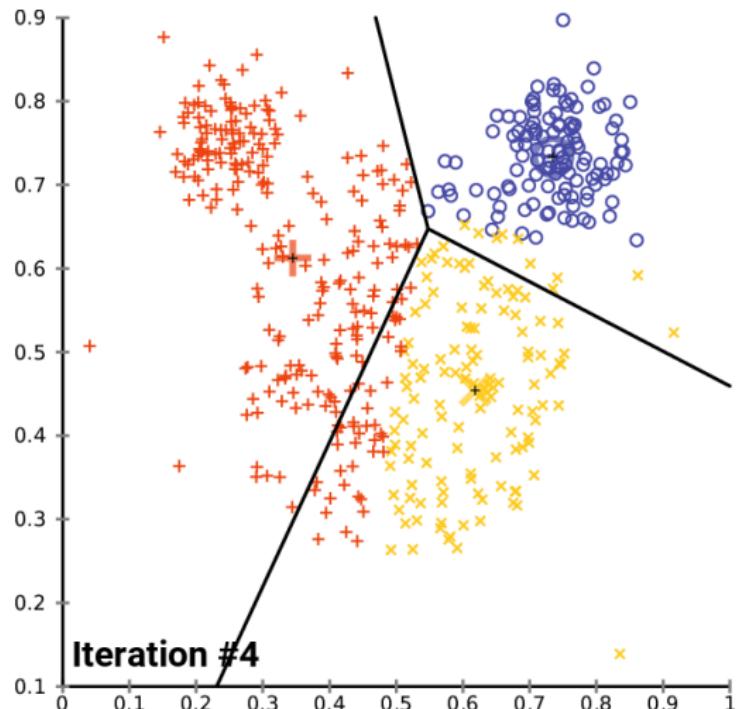
k-médias



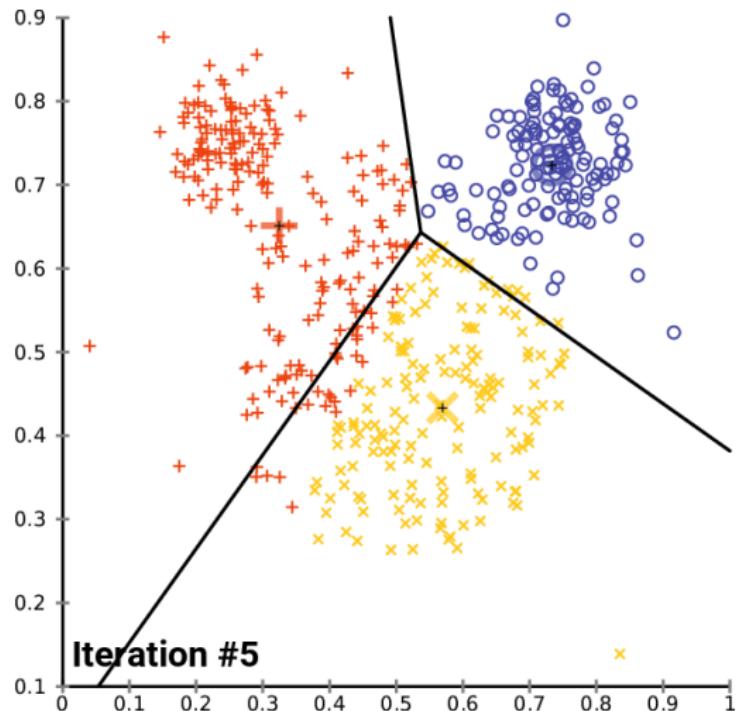
k-médias



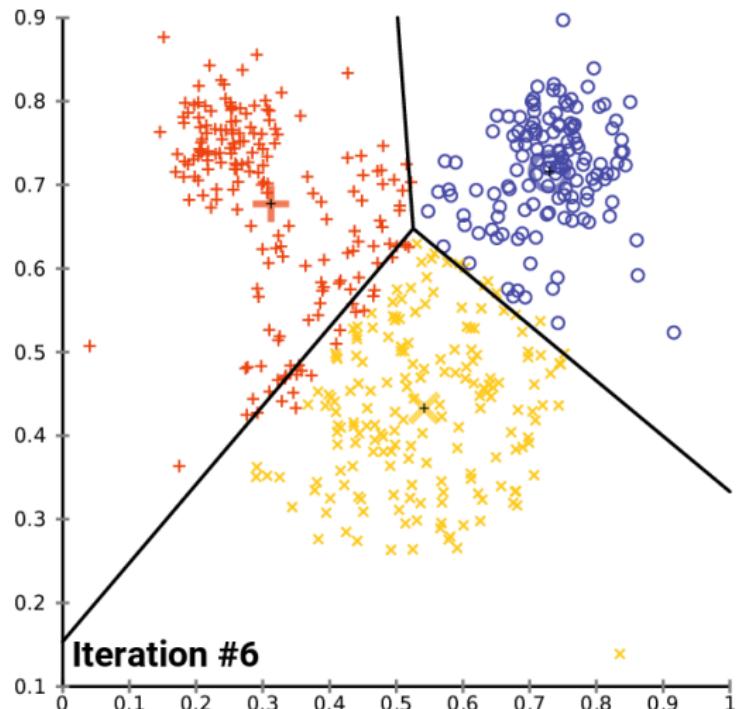
k-médias



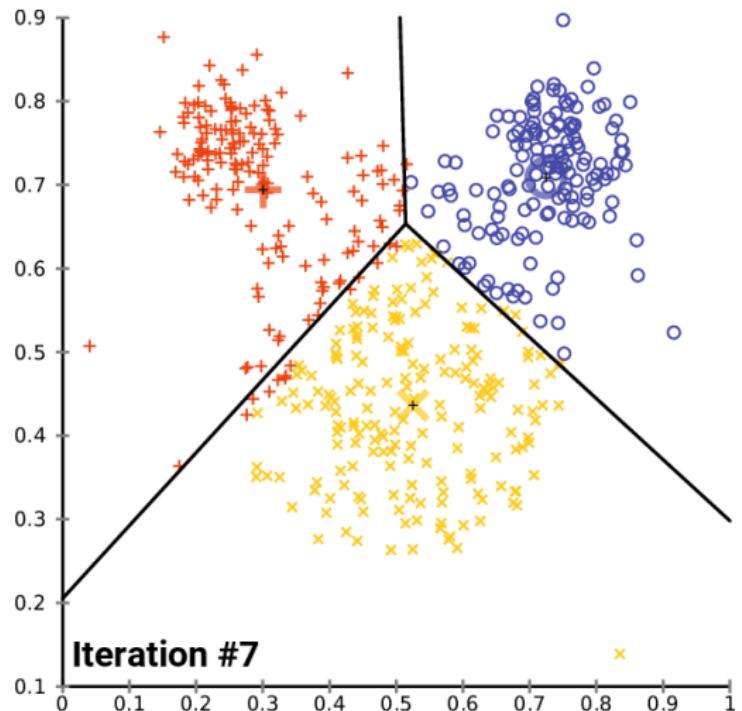
k-médias



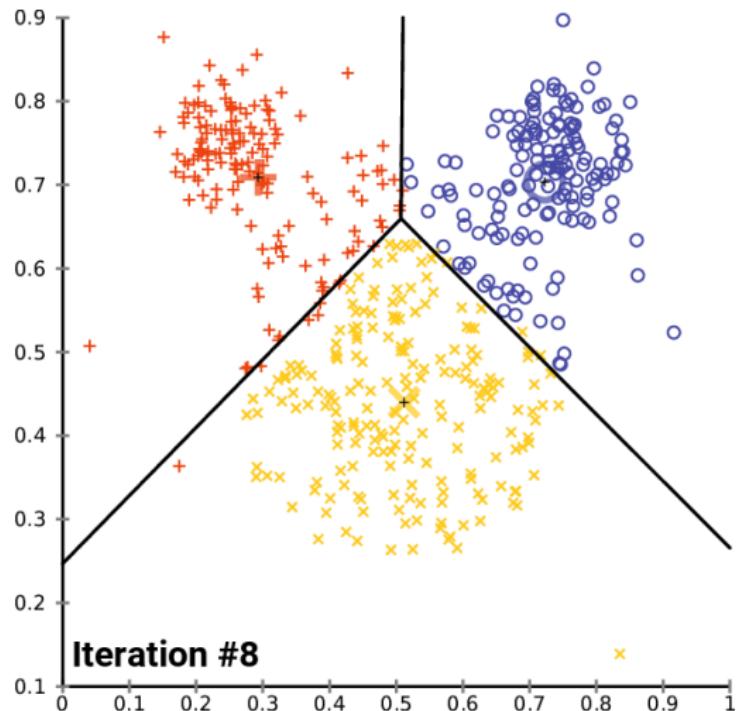
k-médias



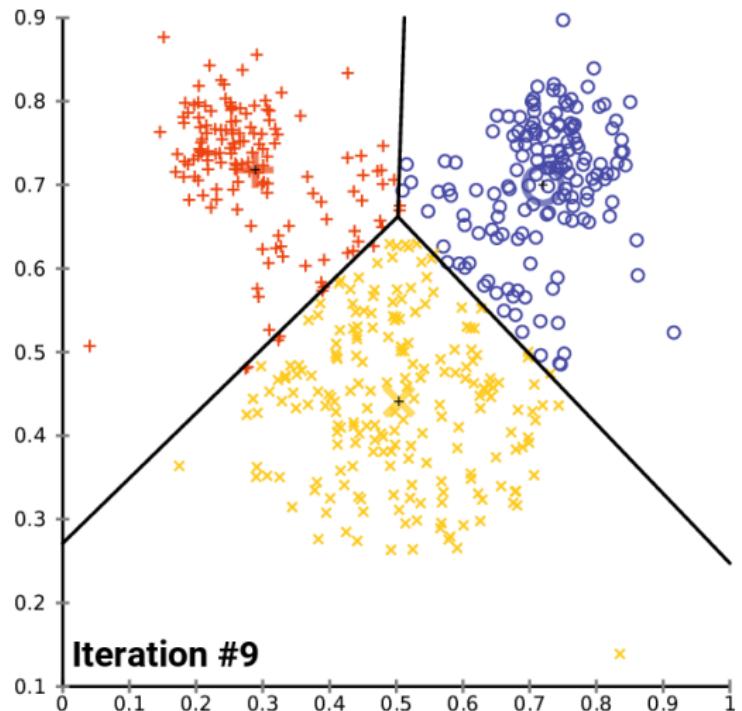
k-médias



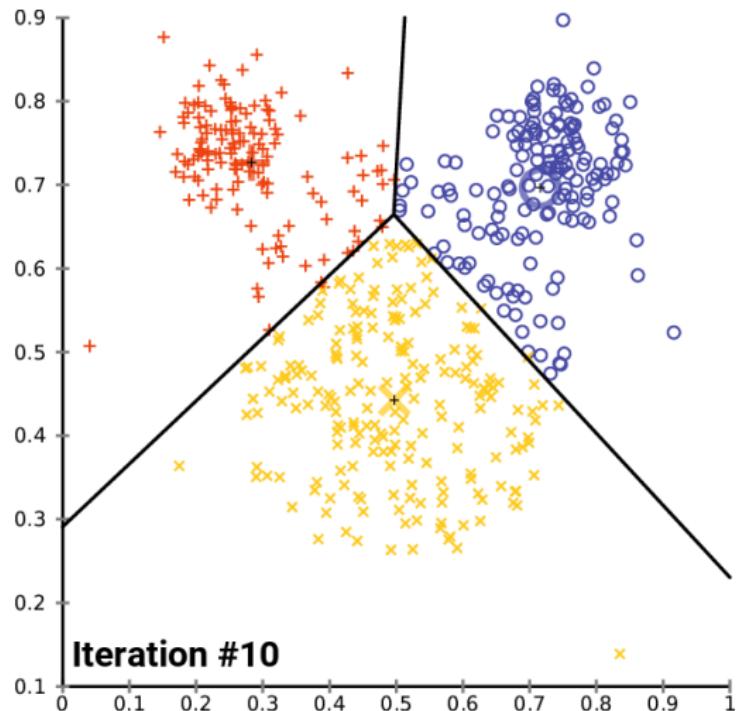
k-médias



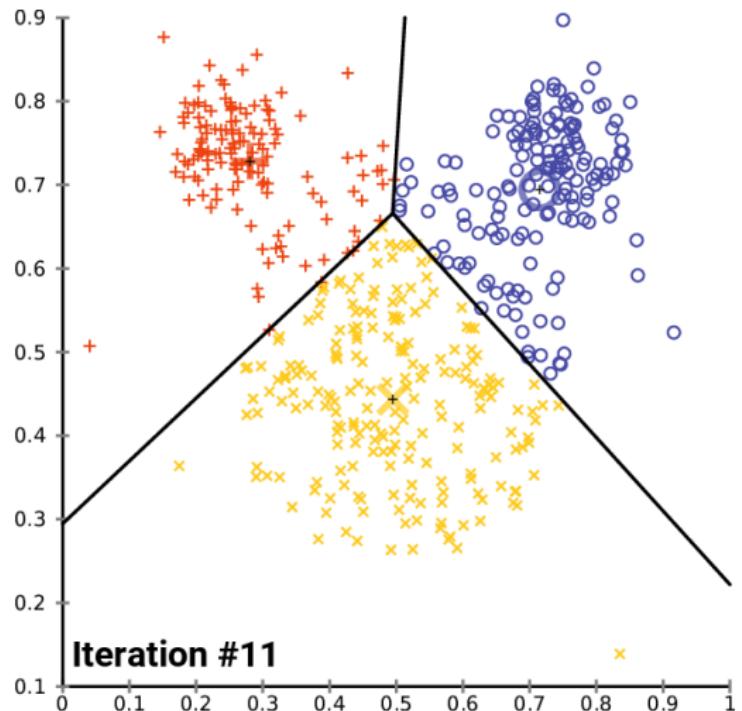
k-médias



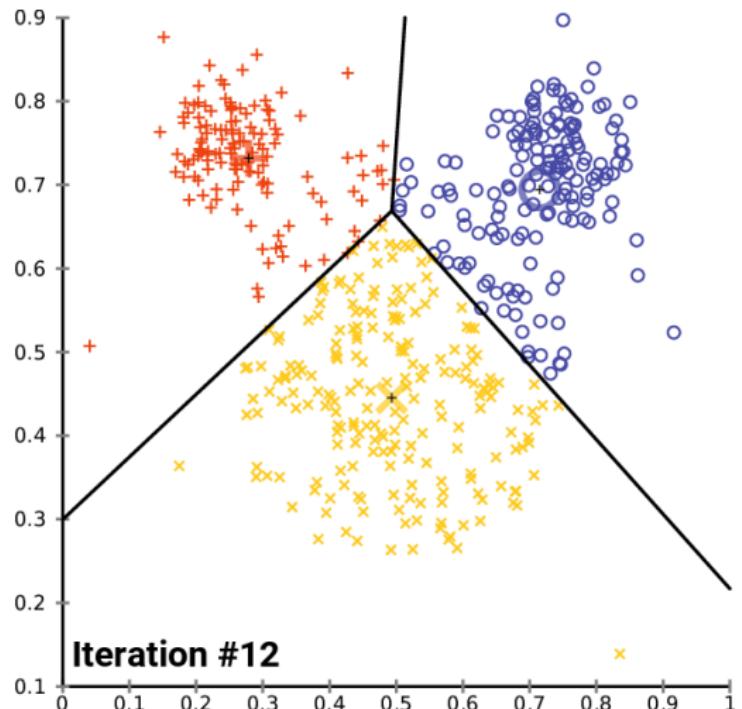
k-médias



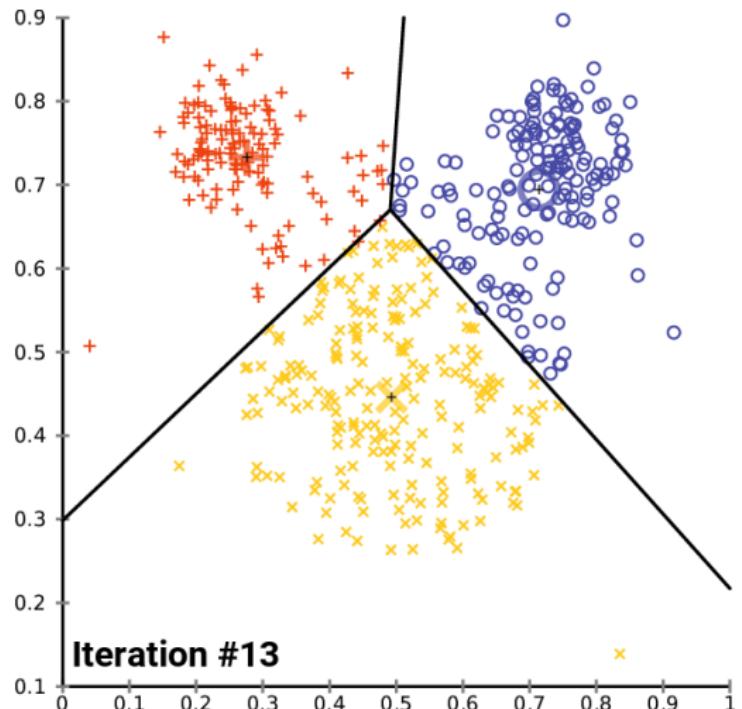
k-médias



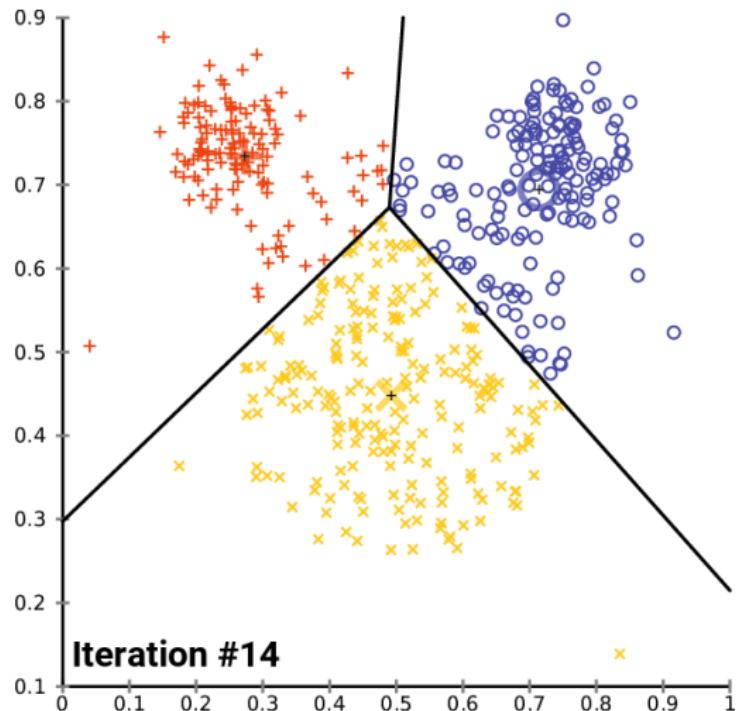
k-médias



k-médias



k-médias



Agrupamento hierárquico

Agrupamento hierárquico

- Em aplicações biológicas, é comum que dados estejam organizados de forma hierárquica

Agrupamento hierárquico

- ▶ Em aplicações biológicas, é comum que dados estejam organizados de forma hierárquica
- ▶ Exemplo: em taxonomia temos reino, filo, classe, ordem...

Agrupamento hierárquico

- ▶ Em aplicações biológicas, é comum que dados estejam organizados de forma hierárquica
- ▶ Exemplo: em taxonomia temos reino, filo, classe, ordem...
- ▶ Em *agrupamento hierárquico* buscamos representações que capturem essa hierarquia

Formalização do agrupamento hierárquico *aglomerativo*

Formalização do agrupamento hierárquico *aglomerativo*

- ▶ Considere uma sequência de partições de N pontos em grupos

Formalização do agrupamento hierárquico *aglomerativo*

- ▶ Considere uma sequência de partições de N pontos em grupos
- ▶ Primeira partição contém N grupos, cada um com um ponto

Formalização do agrupamento hierárquico *aglomerativo*

- ▶ Considere uma sequência de partições de N pontos em grupos
- ▶ Primeira partição contém N grupos, cada um com um ponto
- ▶ Segunda partição contém $N - 1$ grupos
(um grupo com 2 pontos, todos os outros com 1)

Formalização do agrupamento hierárquico *aglomerativo*

- ▶ Considere uma sequência de partições de N pontos em grupos
- ▶ Primeira partição contém N grupos, cada um com um ponto
- ▶ Segunda partição contém $N - 1$ grupos
(um grupo com 2 pontos, todos os outros com 1)
- ▶ A k -ésima partição contém $N - k + 1$ grupos

Exemplo: agrupando x_1, \dots, x_8

$k = 1:$ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

Exemplo: agrupando x_1, \dots, x_8

$k = 1:$ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

$k = 2:$ $x_1, x_2, x_3, x_4, x_5, \{x_6, x_7\}, x_8$

Exemplo: agrupando x_1, \dots, x_8

$k = 1:$ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

$k = 2:$ $x_1, x_2, x_3, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 3:$ $x_1, \{x_2, x_3\}, x_4, x_5, \{x_6, x_7\}, x_8$

Exemplo: agrupando x_1, \dots, x_8

$k = 1:$ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

$k = 2:$ $x_1, x_2, x_3, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 3:$ $x_1, \{x_2, x_3\}, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 4:$ $x_1, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}, x_8$

Exemplo: agrupando x_1, \dots, x_8

$k = 1:$ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

$k = 2:$ $x_1, x_2, x_3, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 3:$ $x_1, \{x_2, x_3\}, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 4:$ $x_1, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}, x_8$

$k = 5:$ $x_1, \{x_2, x_3\}, \{\{x_4, x_5\}, \{x_6, x_7\}\}, x_8$

Exemplo: agrupando x_1, \dots, x_8

$k = 1:$ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

$k = 2:$ $x_1, x_2, x_3, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 3:$ $x_1, \{x_2, x_3\}, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 4:$ $x_1, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}, x_8$

$k = 5:$ $x_1, \{x_2, x_3\}, \{\{x_4, x_5\}, \{x_6, x_7\}\}, x_8$

$k = 6:$ $\{x_1, \{x_2, x_3\}\}, \{\{x_4, x_5\}, \{x_6, x_7\}\}, x_8$

Exemplo: agrupando x_1, \dots, x_8

$k = 1:$ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

$k = 2:$ $x_1, x_2, x_3, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 3:$ $x_1, \{x_2, x_3\}, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 4:$ $x_1, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}, x_8$

$k = 5:$ $x_1, \{x_2, x_3\}, \{\{x_4, x_5\}, \{x_6, x_7\}\}, x_8$

$k = 6:$ $\{x_1, \{x_2, x_3\}\}, \{\{x_4, x_5\}, \{x_6, x_7\}\}, x_8$

$k = 7:$ $\{x_1, \{x_2, x_3\}\}, \{\{\{x_4, x_5\}, \{x_6, x_7\}\}\}, x_8$

Exemplo: agrupando x_1, \dots, x_8

$k = 1:$ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

$k = 2:$ $x_1, x_2, x_3, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 3:$ $x_1, \{x_2, x_3\}, x_4, x_5, \{x_6, x_7\}, x_8$

$k = 4:$ $x_1, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}, x_8$

$k = 5:$ $x_1, \{x_2, x_3\}, \{\{x_4, x_5\}, \{x_6, x_7\}\}, x_8$

$k = 6:$ $\{x_1, \{x_2, x_3\}\}, \{\{x_4, x_5\}, \{x_6, x_7\}\}, x_8$

$k = 7:$ $\{x_1, \{x_2, x_3\}\}, \{\{\{x_4, x_5\}, \{x_6, x_7\}\}\}, x_8$

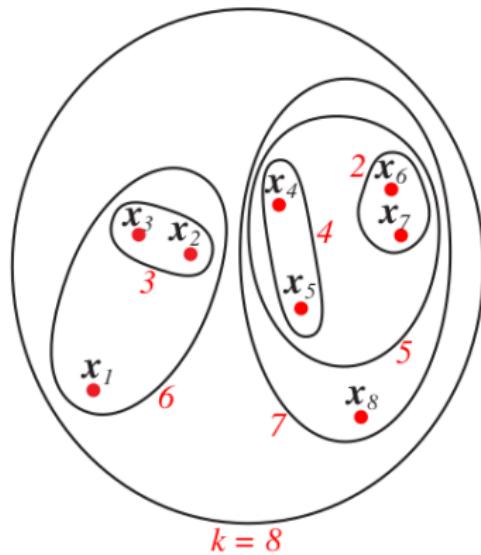
$k = 8:$ $\{\{x_1, \{x_2, x_3\}\}\}, \{\{\{x_4, x_5\}, \{x_6, x_7\}\}\}, x_8\}$

Parte prática 2: agrupamentos

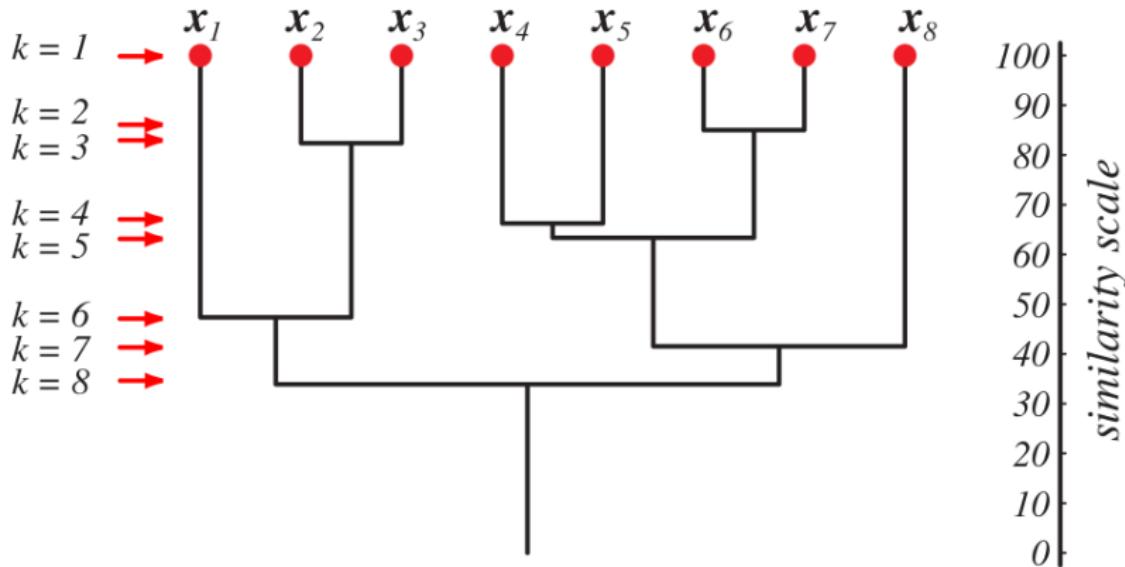
Abra em seu COLAB o arquivo:

`ic.unicamp.br/~msreis/Atividade-Pratica-2.ipynb`

Representação por diagrama de Venn



Dendrogramas



Agrupando $N \times d$ matrizes

- ▶ Organizando N pontos d -dimensionais em uma matriz $N \times d$, podemos realizar dois agrupamentos hierárquicos

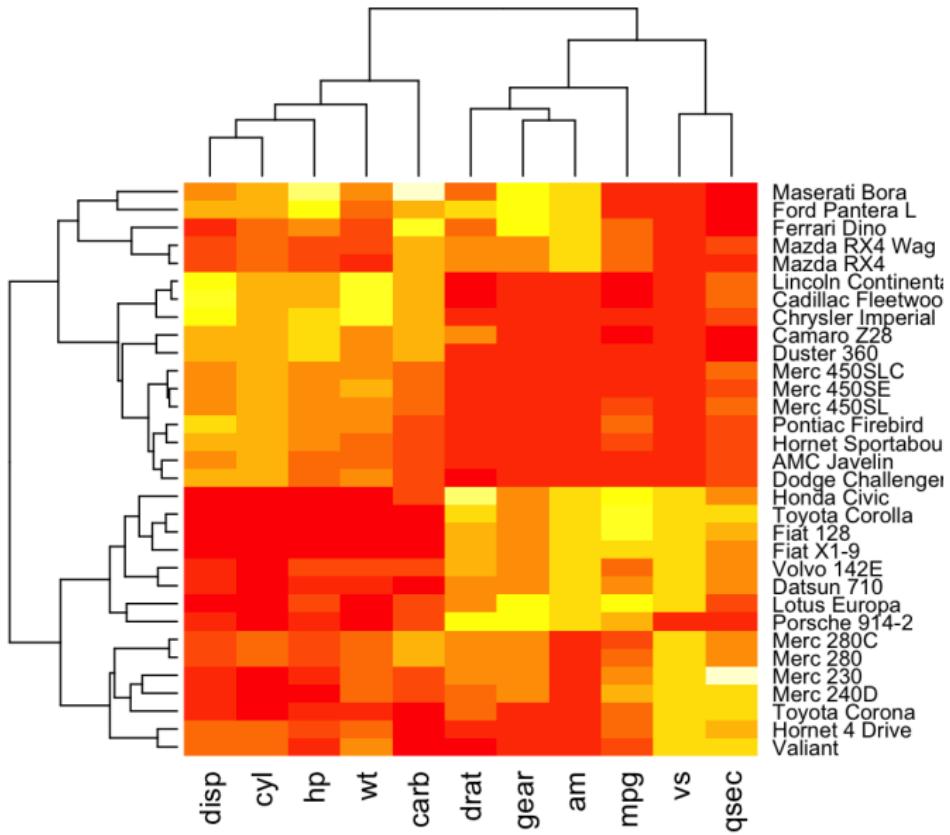
Agrupando $N \times d$ matrizes

- ▶ Organizando N pontos d -dimensionais em uma matriz $N \times d$, podemos realizar dois agrupamentos hierárquicos
- ▶ Essa matriz também é conhecida como *matriz de ocorrência*

Agrupando $N \times d$ matrizes

- ▶ Organizando N pontos d -dimensionais em uma matriz $N \times d$, podemos realizar dois agrupamentos hierárquicos
- ▶ Essa matriz também é conhecida como *matriz de ocorrência*
- ▶ O resultado obtido pode ser visualizado através de um mapa de calor (*heatmap*)

Heatmaps



Comunidade e conjuntos de dados (datasets)

Comunidade e conjuntos de dados (datasets)

kaggle

Comunidade e conjuntos de dados (datasets)

kaggle



Resumo de hoje:

Conceitos iniciais (definições, paradigmas, etc.)

Aprendizado Supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

Aprendizado Não-supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

E agora?

Sessão de perguntas e respostas (Q&A)

E agora?

E agora?

1. Explorar as abordagens disponíveis no `sklearn`

E agora?

1. Explorar as abordagens disponíveis no `sklearn`
2. Ler livros-texto de qualidade (muitos gratuitos disponíveis)

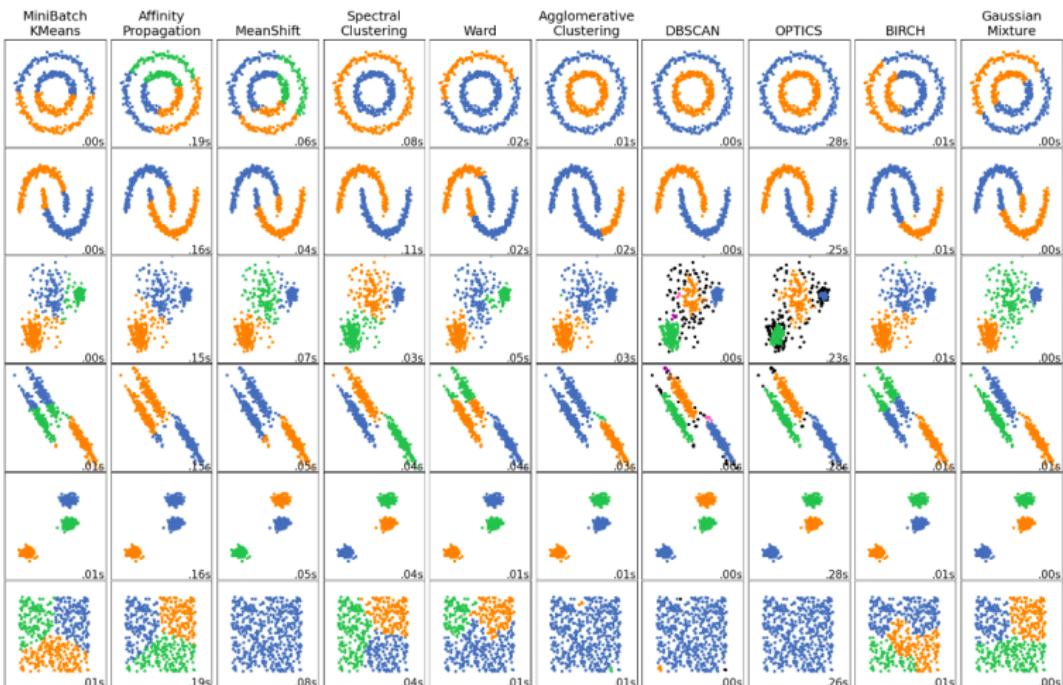
E agora?

1. Explorar as abordagens disponíveis no `sklearn`
2. Ler livros-texto de qualidade (muitos gratuitos disponíveis)
3. Acompanhar a literatura relevante e recente da área

E agora?

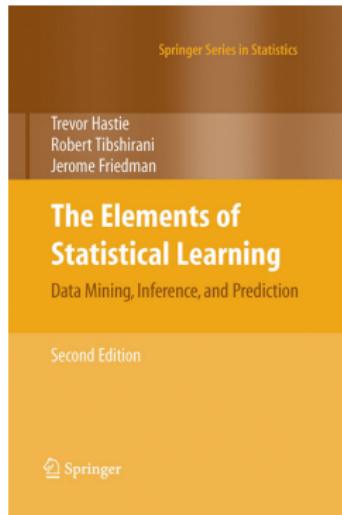
1. Explorar as abordagens disponíveis no `sklearn`
2. Ler livros-texto de qualidade (muitos gratuitos disponíveis)
3. Acompanhar a literatura relevante e recente da área
4. Inscrever-se em cursos e/ou disciplinas na área

2.3.1. Overview of clustering methods



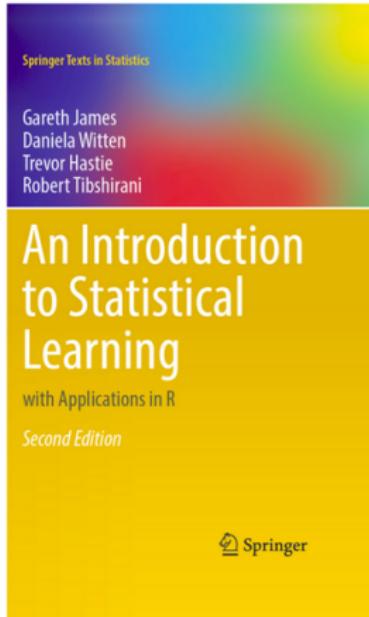
A comparison of the clustering algorithms in scikit-learn

Livros-texto gratuitos



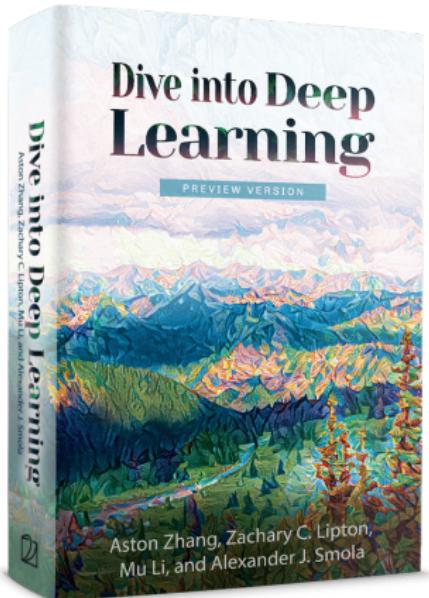
Free version online: hastie.su.domains/ElemStatLearn

Livros-texto gratuitos



Free version online: www.statlearning.com

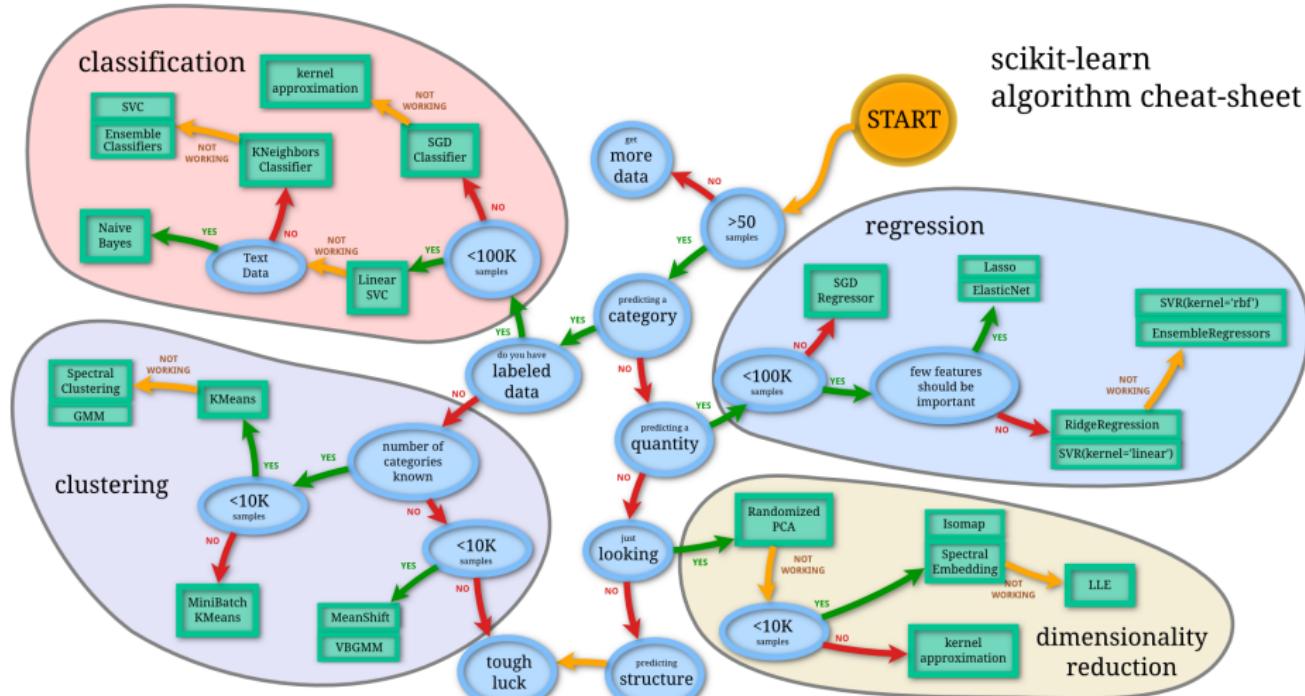
Livros-texto gratuitos



Free version online: d2l.ai

Scikit-learn user guide e tutoriais

scikit-learn algorithm cheat-sheet



Conferências mais importantes em IA/ML

Conferências mais importantes em IA/ML

1. Conference on Neural Information Processing Systems
(NeurIPS, antiga NIPS)

Conferências mais importantes em IA/ML

1. Conference on Neural Information Processing Systems
(NeurIPS, antiga NIPS)

Conferências mais importantes em IA/ML

1. Conference on Neural Information Processing Systems
(NeurIPS, antiga NIPS)
2. International Conference on Machine Learning (ICML)

Conferências mais importantes em IA/ML

1. Conference on Neural Information Processing Systems (NeurIPS, antiga NIPS)
2. International Conference on Machine Learning (ICML)
3. International Conference on Learning Representations

Conferências mais importantes em IA/ML

1. Conference on Neural Information Processing Systems (NeurIPS, antiga NIPS)
2. International Conference on Machine Learning (ICML)
3. International Conference on Learning Representations
4. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)

Cursos em IA na UNICAMP

Cursos em IA na UNICAMP

- ▶ Extensão Mineração de Dados Complexos (100% online)

ic.unicamp.br/~mdc

Cursos em IA na UNICAMP

- ▶ Extensão Mineração de Dados Complexos (100% online)

ic.unicamp.br/~mdc

- ▶ Outras extensões (consultar página do IC)

Cursos em IA na UNICAMP

- ▶ Extensão Mineração de Dados Complexos (100% online)

ic.unicamp.br/~mdc

- ▶ Outras extensões (consultar página do IC)
- ▶ Pós-graduação em Computação (Mestrado e Doutorado)

Cursos em IA na UNICAMP

- ▶ Extensão Mineração de Dados Complexos (100% online)
ic.unicamp.br/~mdc
- ▶ Outras extensões (consultar página do IC)
- ▶ Pós-graduação em Computação (Mestrado e Doutorado)
- ▶ Cursar disciplinas de IA/ML como aluno especial

Resumo de hoje:

Conceitos iniciais (definições, paradigmas, etc.)

Aprendizado Supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

Aprendizado Não-supervisionado

Alguma fundamentação teórica

Parte prática (Colab)

E agora?

Sessão de perguntas e respostas (Q&A)

Sessão de perguntas e respostas (Q&A)

Muito Obrigado! :-)

Nosso grupo de pesquisa está aberto a interessad*s em desenvolver projetos de pós-graduação em Aprendizado de Máquina e/ou Bioinformática! Escrevam para:

msreis@ic.unicamp.br

