



REDES NEURAIS E DEEP LEARNING

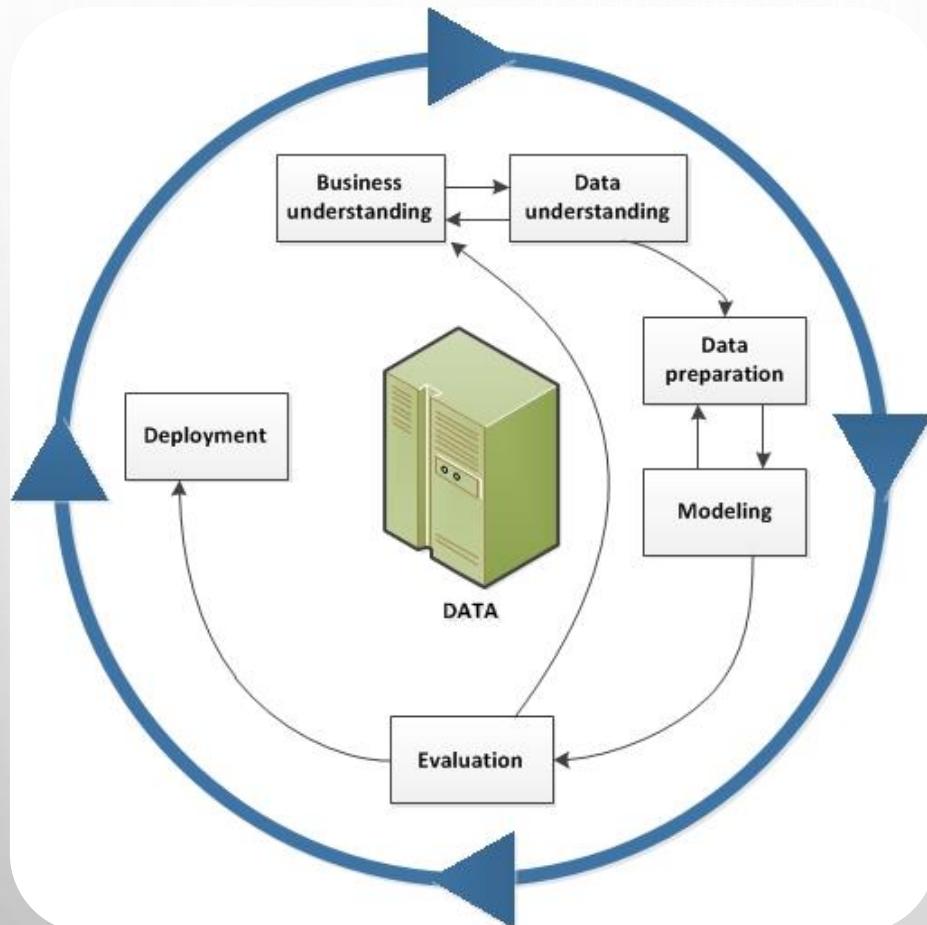
AGRUPAMENTO

DIEGO RODRIGUES DSC
INFNET

Bloco	Matéria	Calendário	Avaliação
Treinamento Clássico	Introdução	06/10	
	Classificação	08, *13	
	Régressão	27, *29	
	Agrupamento 	03/11, *05	
Redes Profundas	Séries Temporais	10, *12	<Modelo Clássico>
	Deep Feed Forward	17, *19	
	Visão Computacional	24, *26	
	Autoencoders	01/12, *03	<Modelo Profundo>
Treinamento Moderno	Transfer Learning	08, *10	
	Sequências	15, *17	<Modelo Avançado>
	Modelos Generativos	<COMBINAR>	

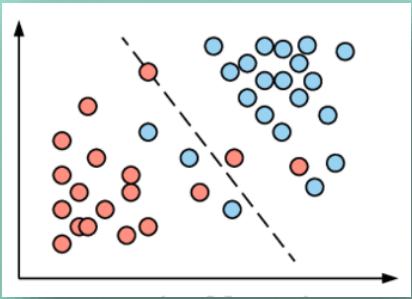
PARTE 1 : TEORIA

CROSS INDUSTRY PROCESS FOR DATA MINING (CRISP-DM)

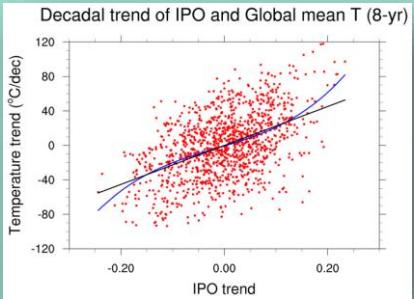


BUSINESS UNDERSTANDING

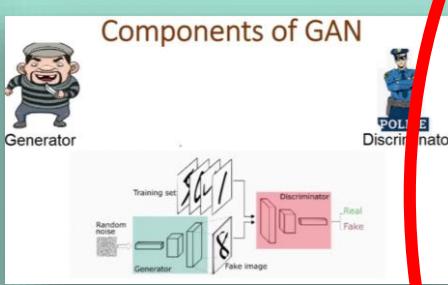
APRENDIZADO SUPERVISIONADO



CLASSIFICAÇÃO

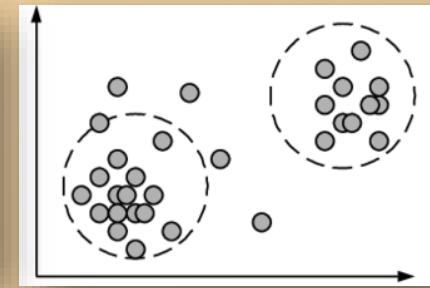


REGRESSÃO



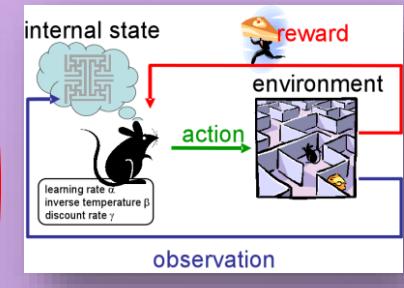
GERATIVO

APRENDIZADO NÃO-SUPERVISIONADO



AGRUPAMENTO

APRENDIZADO POR REFORÇO



REFORÇO

Aprendizado Não-Supervisionado

Não existe um **conhecimento “a priori” dos grupos** contidos nos dados. Algoritmos de agrupamento dependem fortemente de uma definição de “**distância**” ou “**similaridade**” entre as observações.

Agrupamento (Clustering)

Um bebê consegue **agrupar objetos por cor, tamanho, formato** e muitos outros atributos que ele pode observar nos objetos.

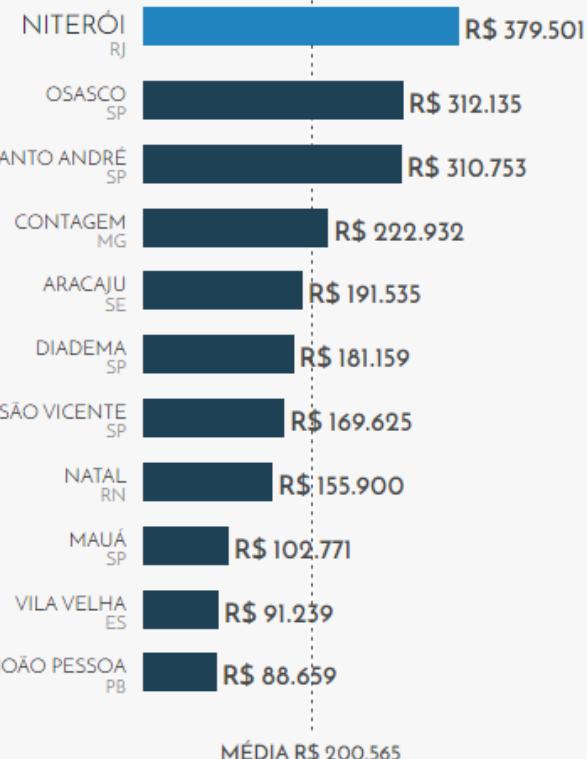
Diferentes maneiras de organizar os objetos são diferentes **estruturas de agrupamentos** existentes em uma amostra de dados.

Um **modelo de agrupamento** é usado para **identificar grupos**, ou estruturas de agrupamentos, nos dados.



De quantas maneiras estes blocos podem ser organizados em grupos?

REPRESENTAÇÃO: COMO ENCONTRAR OS 10 MUNICÍPIOS MAIS SIMILARES A NITERÓI?



VARIÁVEIS QUE FORMAM O GRUPO
COMPARAÇÃO

■ Seu município ■ Média do grupo

Domicílios urbanos - (QTD)

169.162	169.822
---------	---------

Características do Entorno

79,34%	73,59%
--------	--------

Domicílios subnormais - (QTD)

24.286	21.725
--------	--------

Renda média domiciliar

R\$ 4.687	R\$ 2.503
-----------	-----------

Saneamento básico - (QTD)

133.750	136.548
---------	---------

*No gráfico ao lado, é possível comparar o município selecionado com os 10 outros municípios brasileiros de perfil mais semelhante para cada item de receita.

Para cada um destes, foi definido o conjunto de variáveis que mais afetam seu resultado – por exemplo, frota de veículos influencia fortemente o valor total de IPVA.

Por meio dos valores dessas variáveis, chega-se aos 10 municípios mais comparáveis com o selecionado.

Veja acima as variáveis que foram utilizadas para o componente de receita definido.

Clique em cada variável acima para entender sua importância.



VARIÁVEIS QUE FORMAM O GRUPO
COMPARAÇÃO

■ Seu município ■ Média do grupo

Domicílios urbanos - (QTD)

169.162	169.822
---------	---------

Características do Entorno

79,34%	73,59%
--------	--------

Domicílios subnormais - (QTD)

24.286	21.725
--------	--------

Renda média domiciliar

R\$ 4.687	R\$ 2.503
-----------	-----------

Saneamento básico - (QTD)

133.750	136.548
---------	---------

*No gráfico ao lado, é possível comparar o município selecionado com os 10 outros municípios brasileiros de perfil mais semelhante para cada item de receita.

Para cada um destes, foi definido o conjunto de variáveis que mais afetam seu resultado – por exemplo, frota de veículos influencia fortemente o valor total de IPVA.

Por meio dos valores dessas variáveis, chega-se aos 10 municípios mais comparáveis com o selecionado.

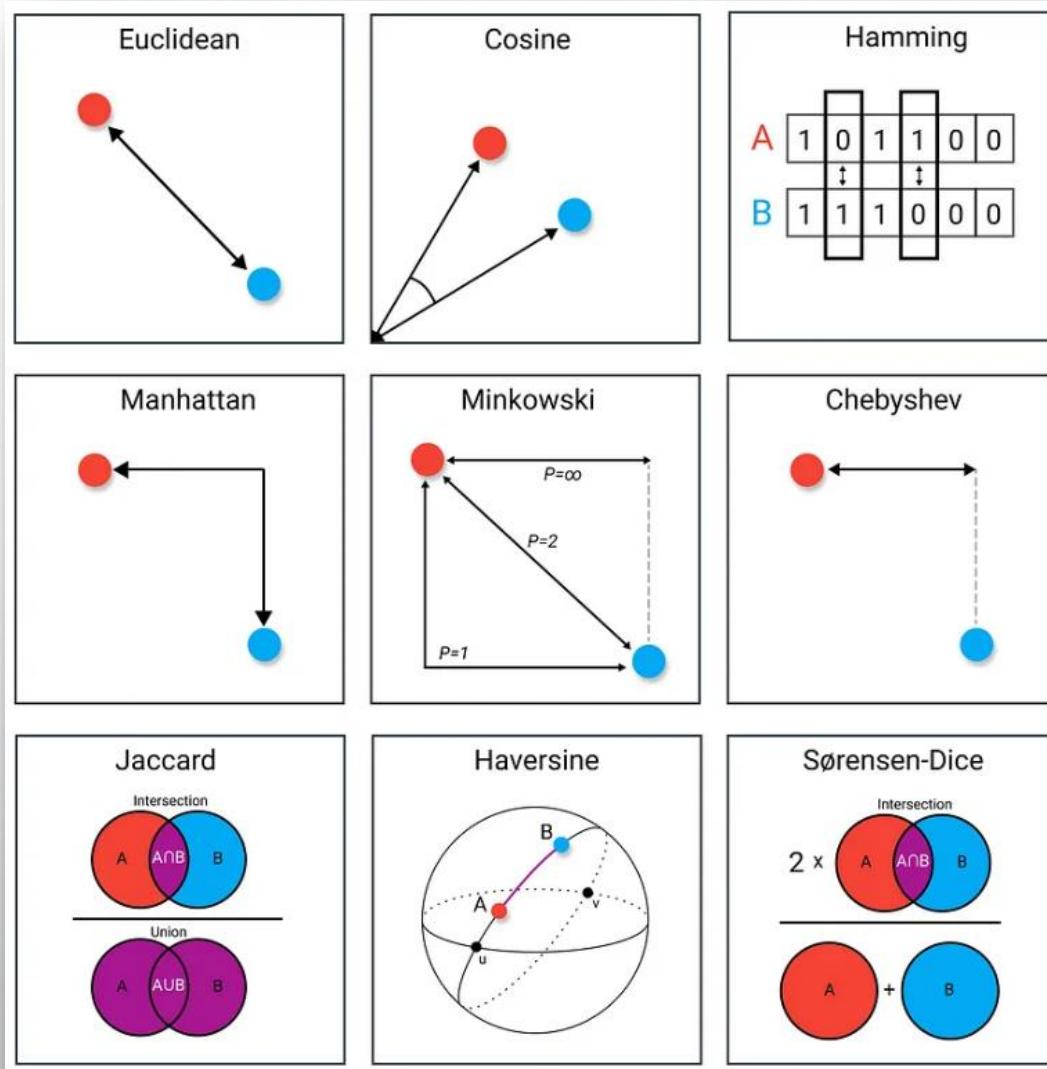
Veja acima as variáveis que foram utilizadas para o componente de receita definido.

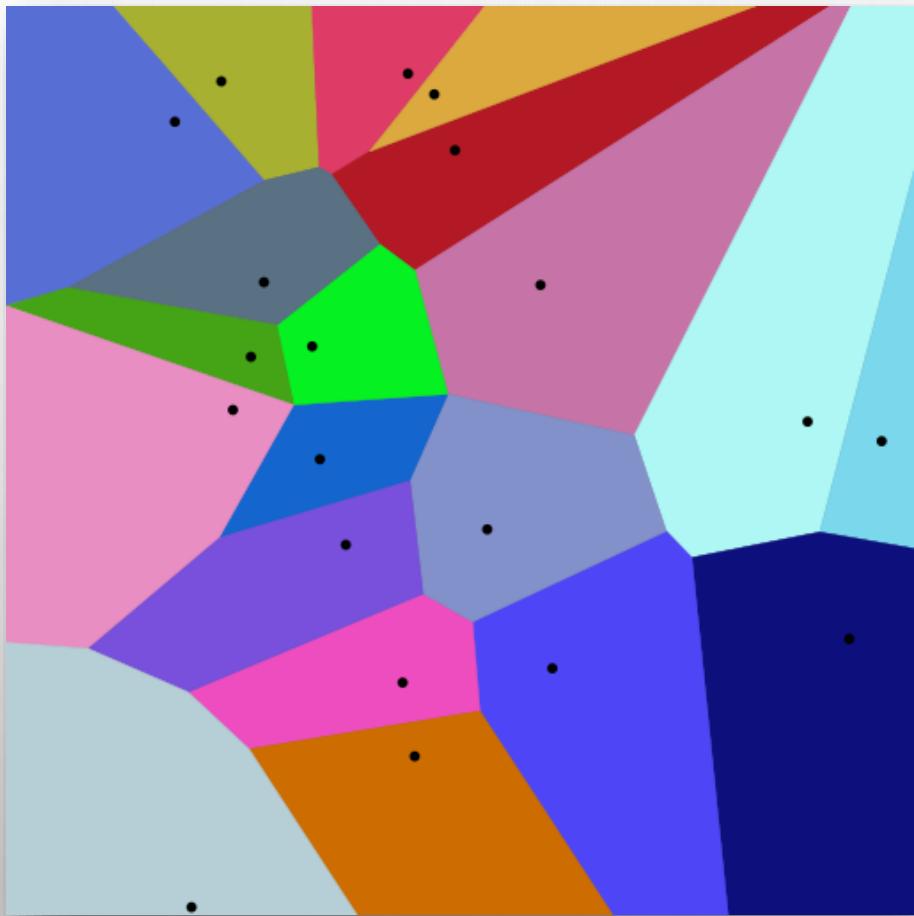
Clique em cada variável acima para entender sua importância.

MODELING

DISTÂNCIA

ALGUMAS MÉTRICAS DE DISTÂNCIA





Distância Euclideana

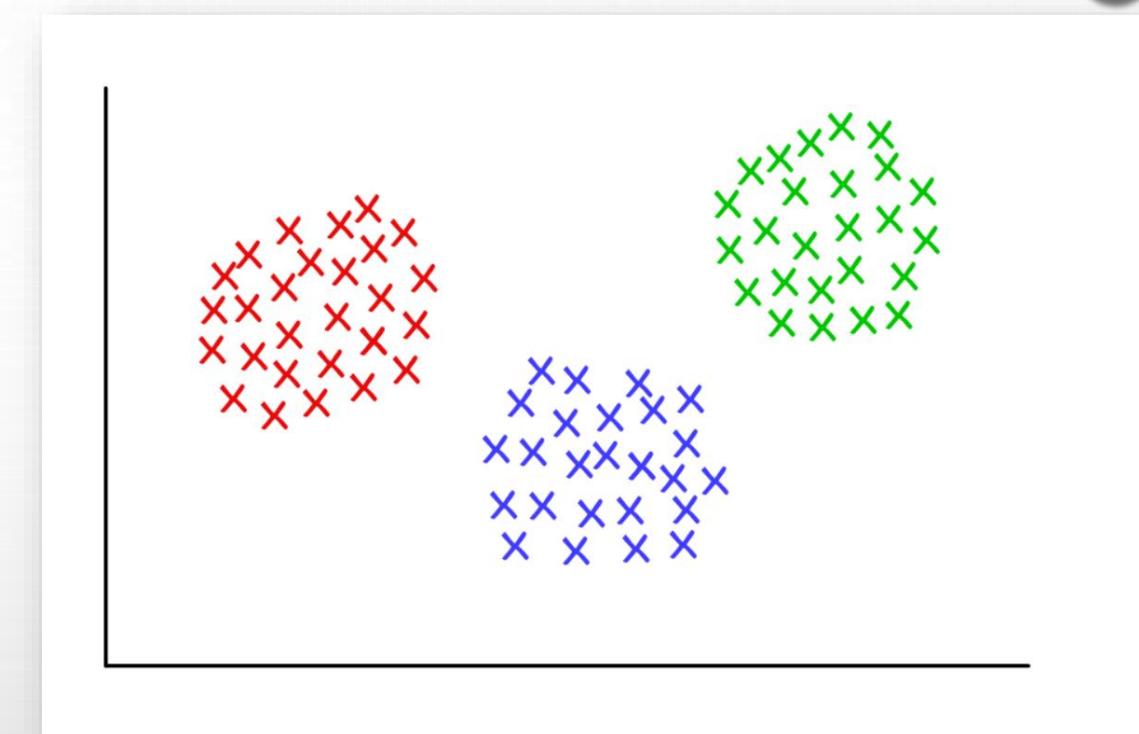


Distância de Manhattan

ALGORITMOS

AGRUPAMENTO: ALGORITMOS

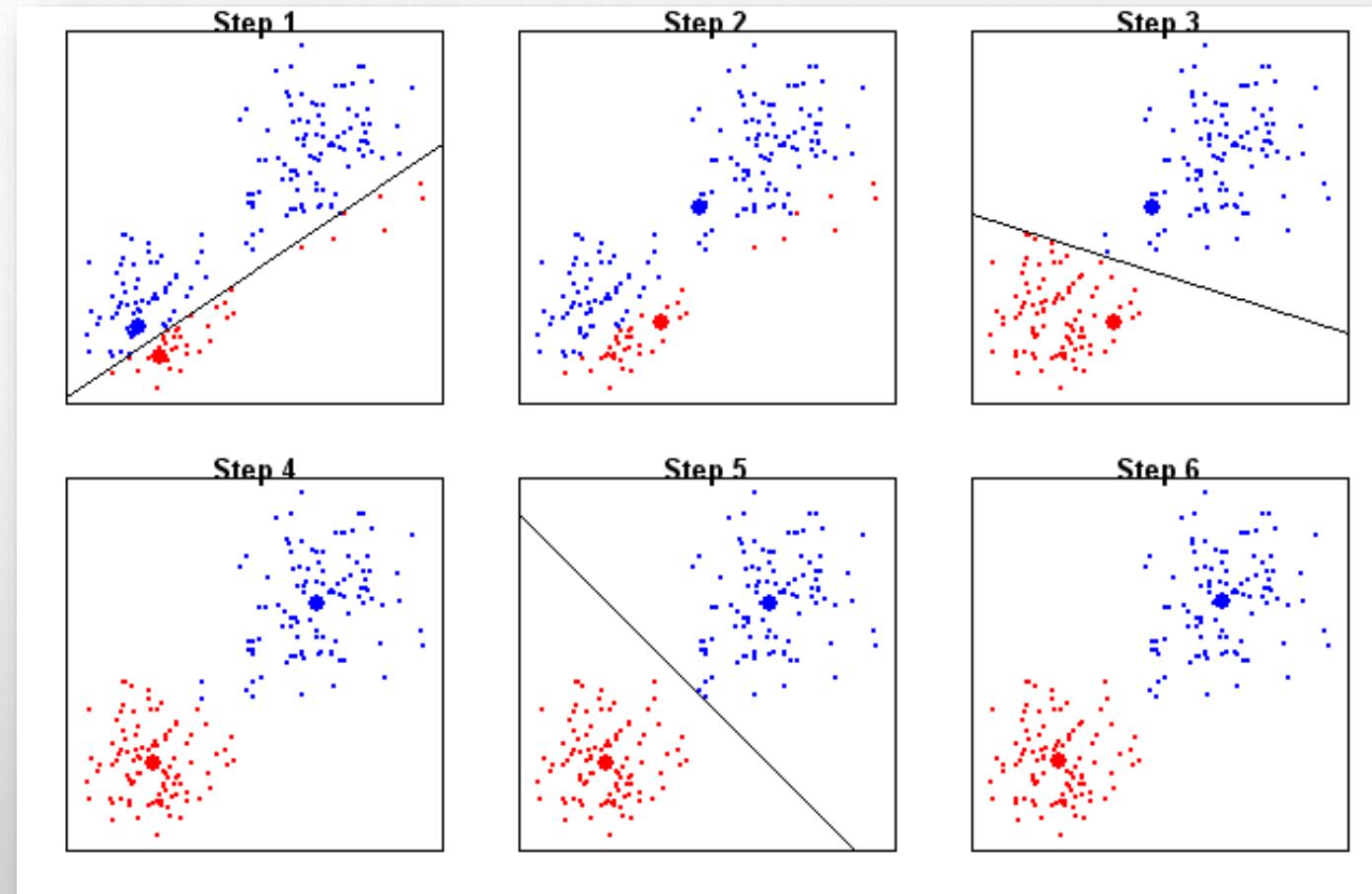
- 1)K-Means
- 2)Hieráquico
- 3)DBSCAN
- 4)Mapa Auto-Organizável



Além da escolha do algoritmo, os resultados do agrupamento dependem diretamente dos atributos e da **métrica escolhida para definir similaridade entre os objetos**.

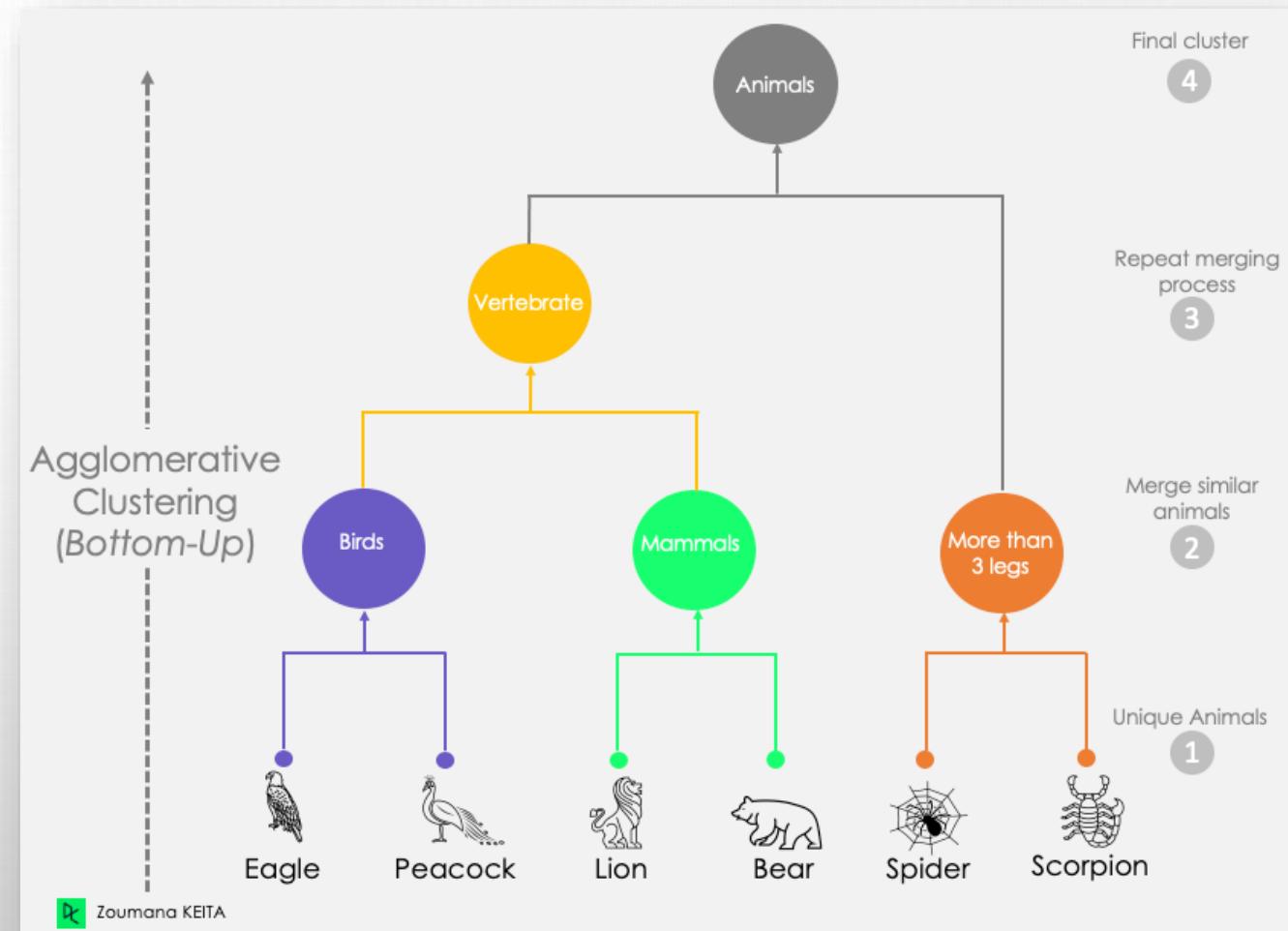
PARTICIONAMENTO : K-MEANS

K-means (ou K-médias) parte de **K centroides** (centros de agrupamento) e através de iteração, **recalcula os centroides** até que particione os dados em **K grupos**.



AGRUPAMENTO HIERÁRQUICO

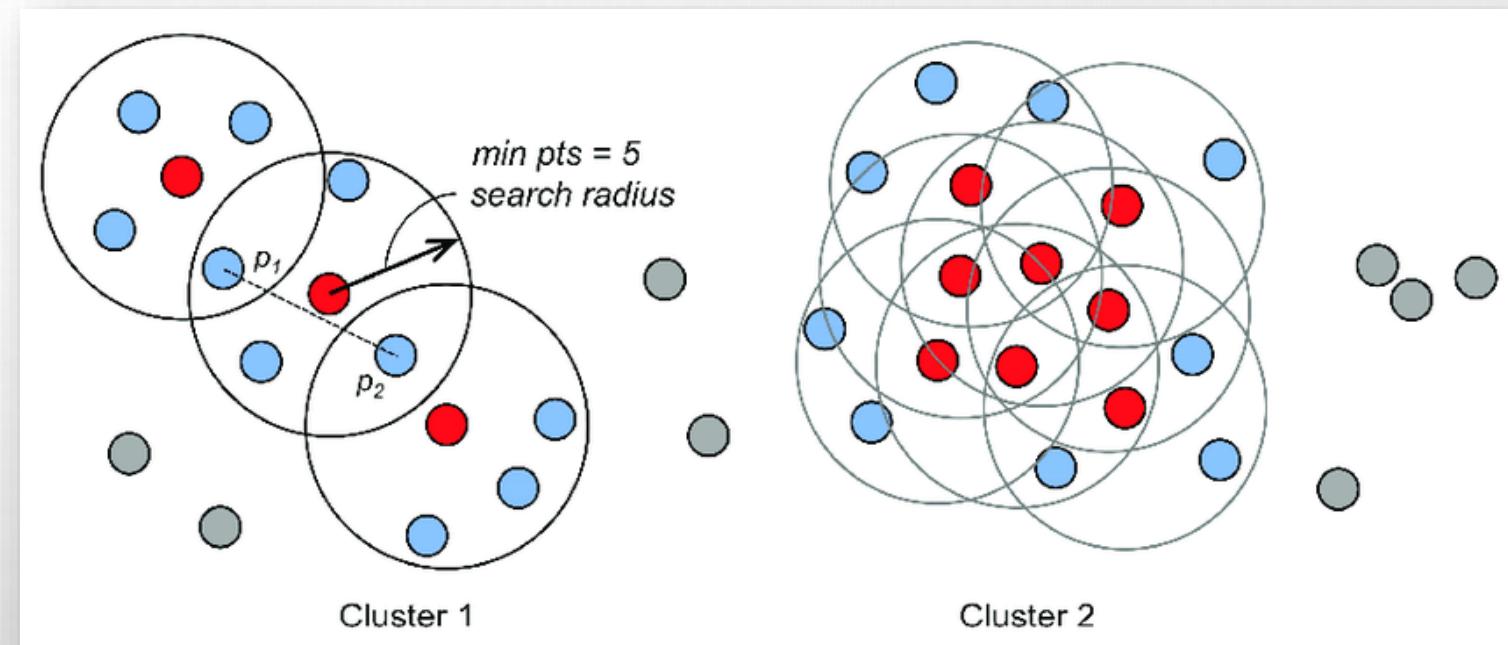
Esse algoritmo depende de mais uma definição de “**Linkage**” para decidir como **calcular similaridade entre um grupo e um indivíduo**, e permite que diferentes arranjos de grupos sejam detectados, variando o **limiar de distância**.



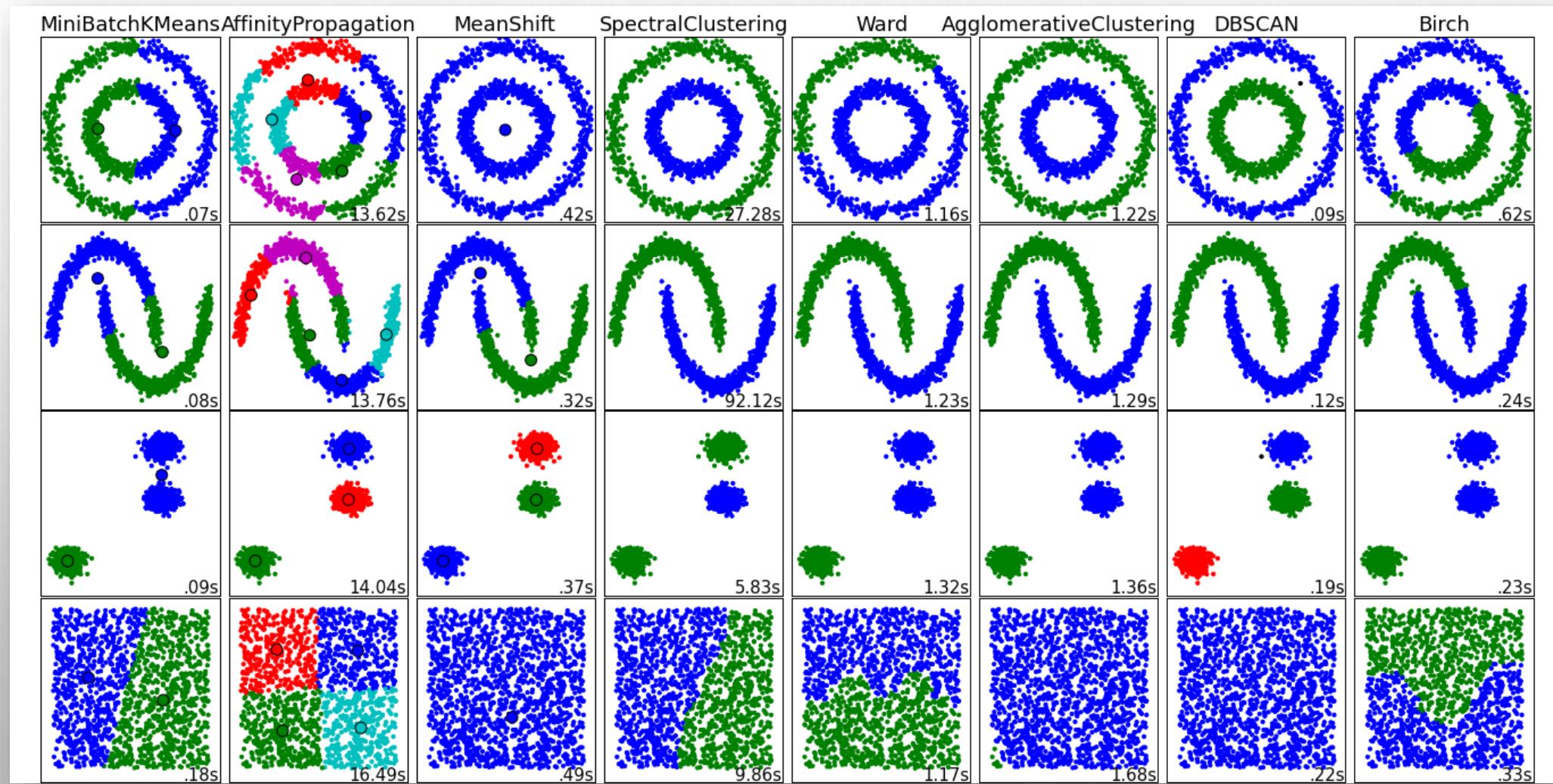
DBSCAN : DENSIDADE

Density-Based Spatial Clustering of Applications with Noise

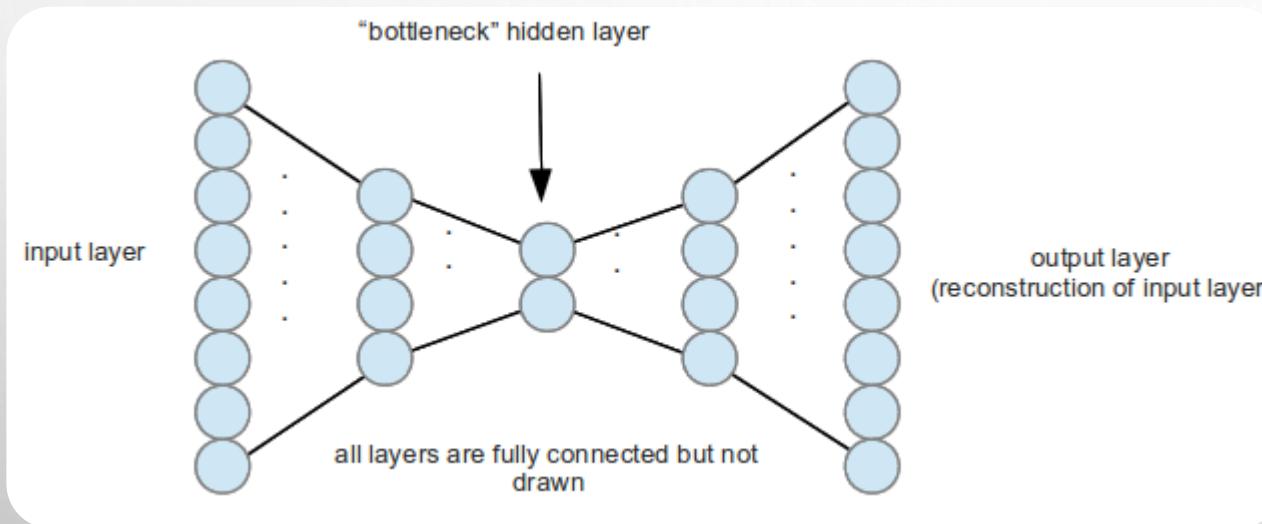
Baseado em uma **similaridade mínima** e **quantidade de vizinhos** para ser considerado um **ponto central**, DBSCAN agrupa **pontos que tenham vizinhos comuns**.



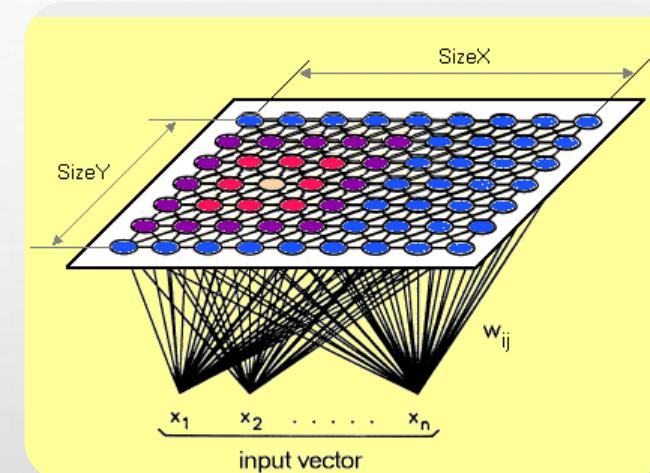
COMPARAÇÃO DE ALGORITMOS : SKLEARN



O “CLUSTERIZADOR UNIVERSAL”



Rede AutoEncoder

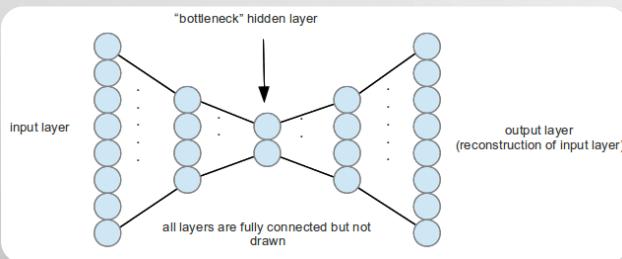


Mapa Auto
Organizável

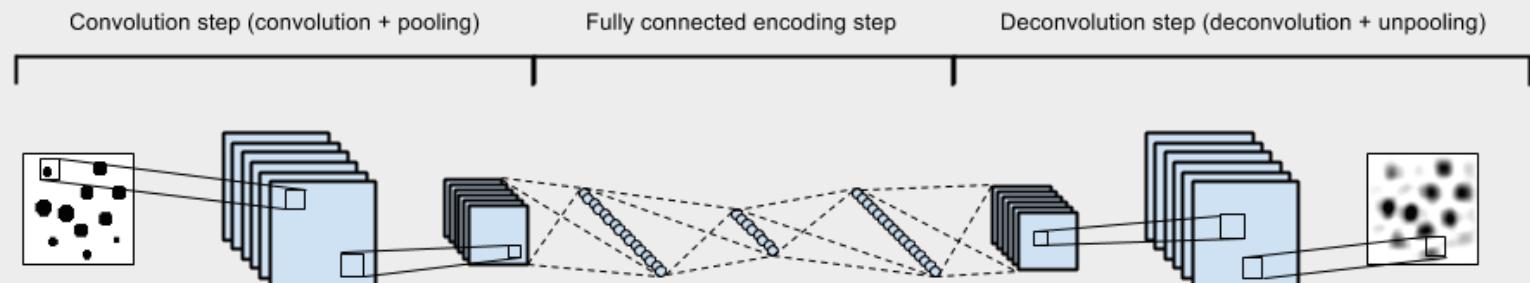
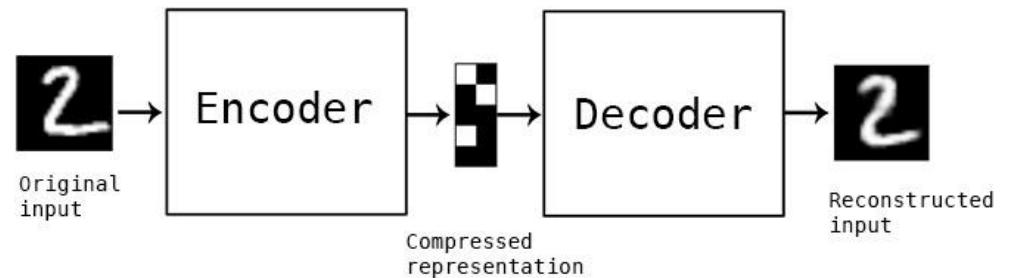
REDE AUTO ENCODER

Encontra uma representação de menor dimensionalidade do dado

Encontra um hiperespaço reduzido contendo toda a informação



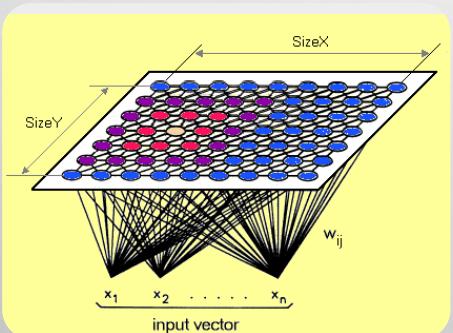
Rede AutoEncoder



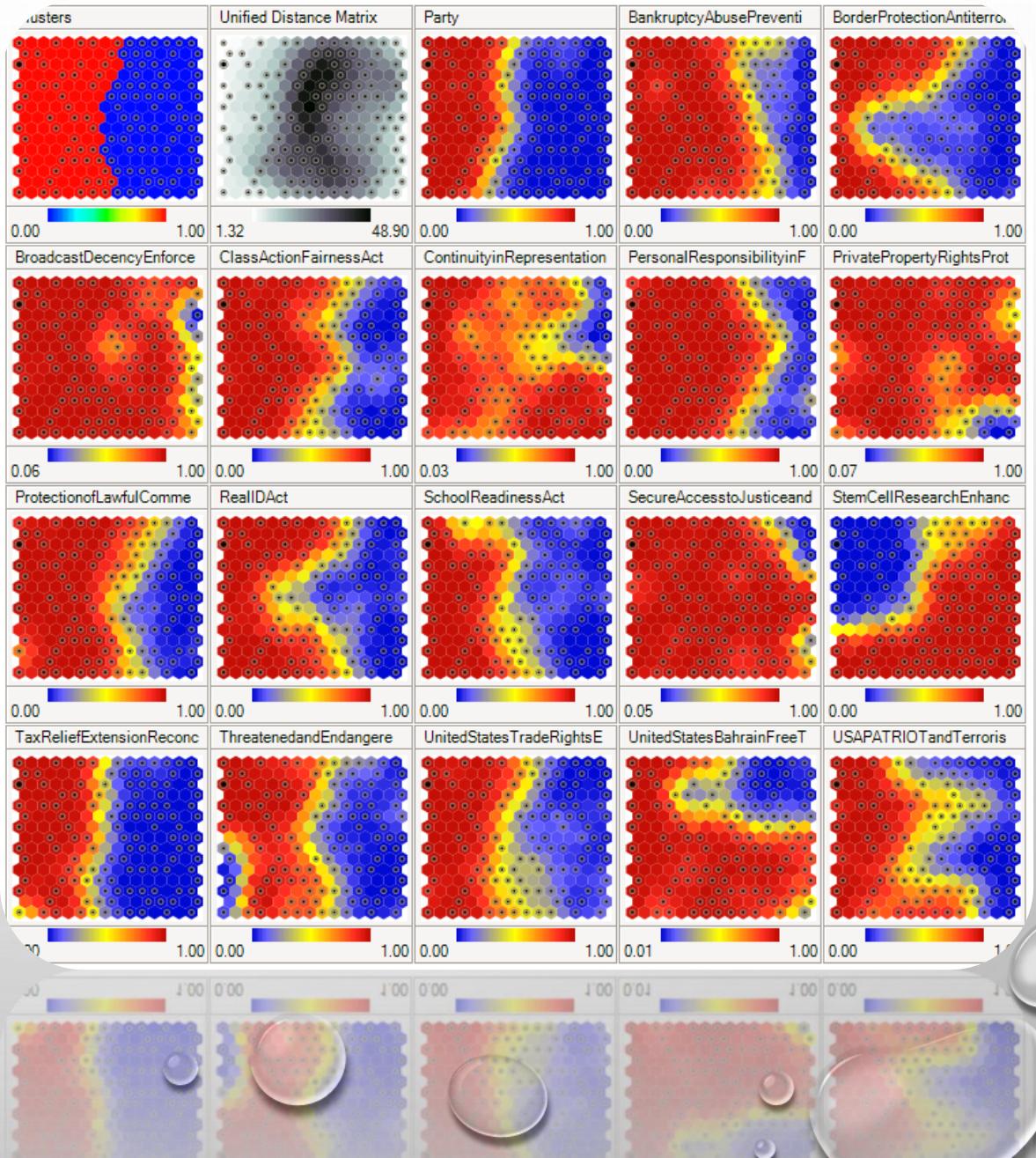
MAPA AUTO ORGANIZÁVEL

Transforma uma
entrada
multidimensional
em um mapa
bidimensional

Cada neurônio
serve como
“centróide” de
uma pequena
região do espaço



Mapa de Kohonen



PARTE 2 : PRÁTICA

AMBIENTE PYTHON



4. Variáveis Aleatórias



1. Editor de Código



2. Gestor de Ambiente



5. Visualização



6. Machine Learning



3. Ambiente Python do Projeto



3. Notebook Dinâmico

PROBLEMA DE NEGÓCIO

Características das flores

Largura & comprimento da pétala

Largura & comprimento da sépala



Iris Setosa



Iris Versicolor



Iris Virginica

REPRESENTAÇÃO



Iris Setosa



Iris Versicolor



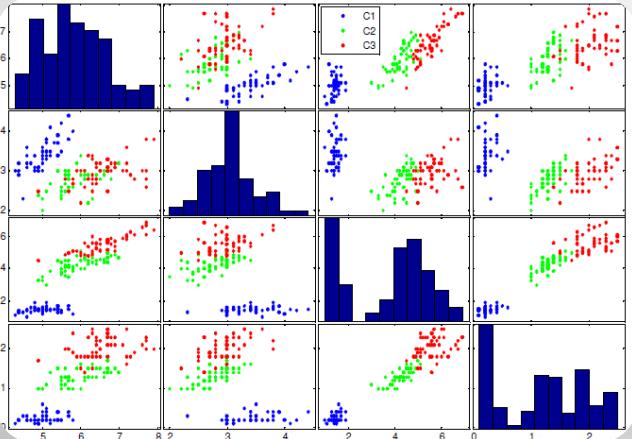
Iris Virginica

Características das flores

Largura & comprimento da pétala

Largura & comprimento da sépala

Espaço de
atributos com
4 dimensões!



<http://archive.ics.uci.edu/ml/datasets/Iris>

MODELAGEM

MAPA AUTO ORGANIZÁVEL

- REPRESENTAÇÃO: 4 ATRIBUTOS > 2 DIMENSÕES NO MAPA
- HIPERPARÂMETROS: NÚMERO DE NEURÔNIOS DO MAPA 10X10.
- TREINAMENTO: BASE DE TREINO COMPLETA.
 - MSE



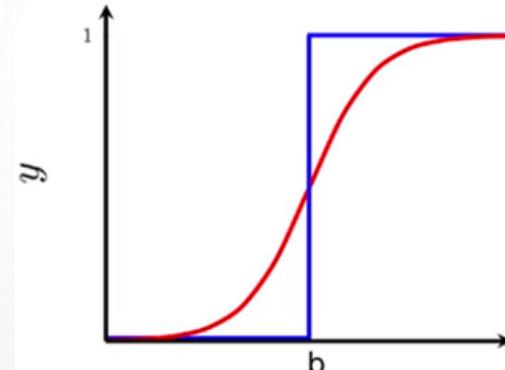
Iris Setosa



Iris Versicolor



Iris Virginica



$$y = \frac{1}{1+e^{-(w^T x + b)}}$$

$$\cdot \sum_{i=1}^n w_i x_i$$

$$\cdot \sum_{i=1}^n \text{fun}_i(x_i)$$

p

REGRESSÃO IRIS

**PRÓXIMA AULA:
MAPA AUTO
ORGANIZÁVEL IRIS**