

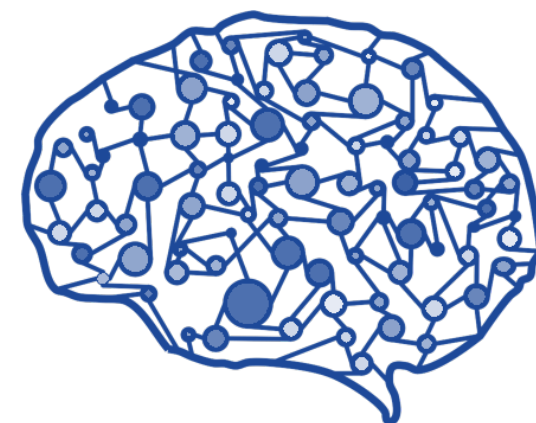


# COC800

## Introdução à Ciência de Dados

### Aprendizado não Supervisionado

- Agrupamentos particionais
- Agrupamentos hierárquicos
- Associação e Apriori
- Modelagem *fuzzy*



## Agrupamentos

Particionais: constroem partições dos dados a cada iteração otimizando algum critério de avaliação

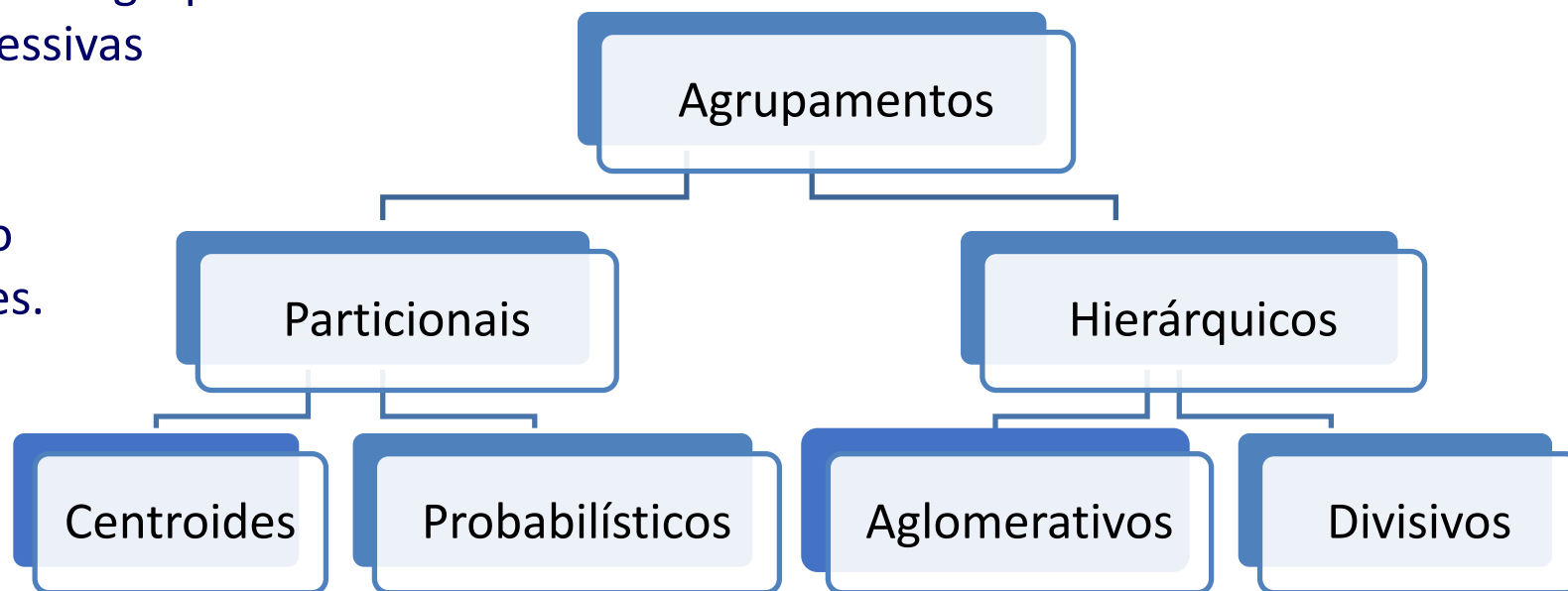
- Centroides: grupos determinados por seus elementos centrais, não necessariamente membros dos grupos;
- Probabilísticos: grupos modelados por distribuições probabilísticas, possibilitando que os elementos a pertinência de elementos a diversos grupos.

Hierárquicos: constroem hierarquias de subconjuntos de registros definidas pelas suas proximidades

- Aglomerativos: *bottom-up*, mescla de grupos
- Divisivos: *top-down*, divisões sucessivas

Essa lista não termina nesses itens.

Esses são apenas alguns dos que serão apresentados nesta sequência de slides.



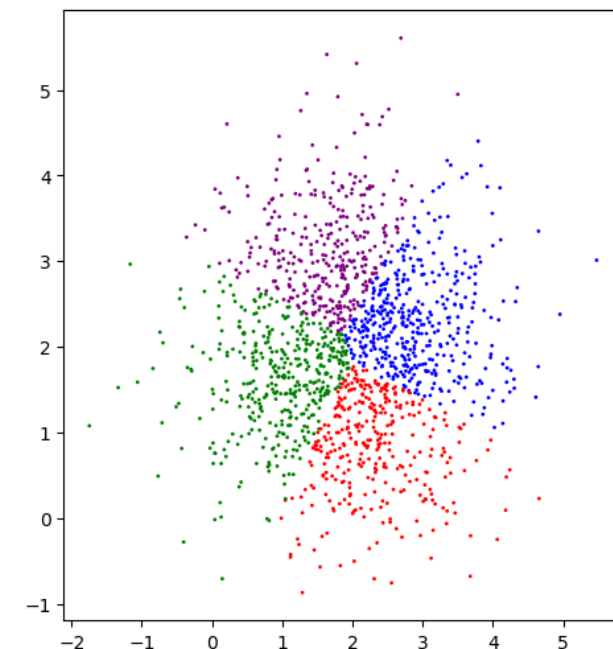
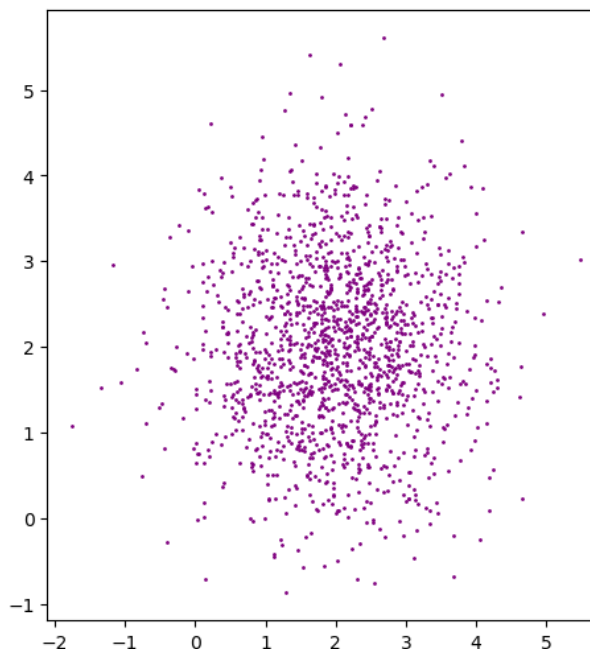
## Agrupamentos pelos centroides: *k-means*

O *k-means* ou k-médias é o algoritmo tradicional da estratégia particional, supondo a existência de grupos convexos e isotrópicos. O objetivo é encontrar as localizações dos centros dos grupos por meio da minimização da inércia ou da soma das distâncias intragrupos, isto é, das distâncias entre os registros e os centros dos seus grupos:

$$\min J(\mathbf{C}) = \sum_{i=1}^N \sum_{\substack{j=1 \\ x_i \in G_j}}^K d(\mathbf{x}_i, \mathbf{c}_j)^2$$

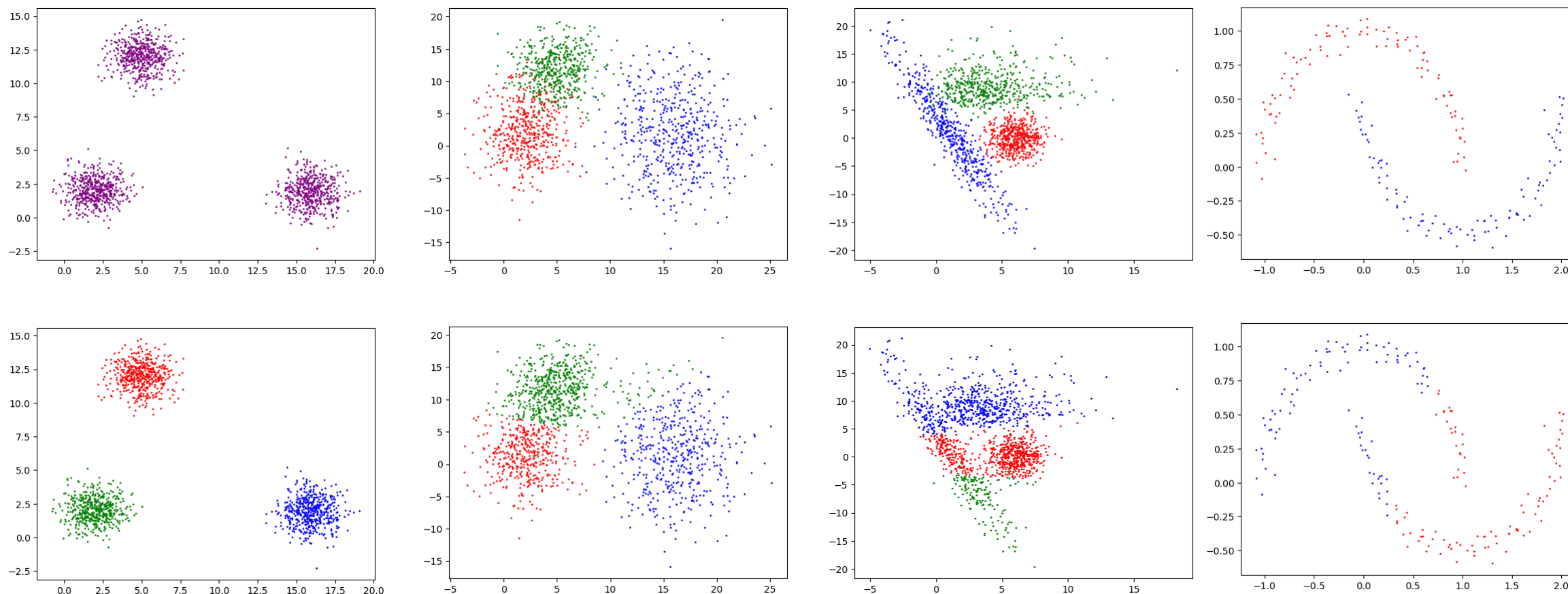
Variabilidade  
Dissimilaridade

- $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_K]$  é a matriz dos centros dos grupos;
- $\mathbf{c}_j = (c_{j1}, \dots, c_{jm}) \in R^m$  é o centro do grupo  $j$ ,  $1 \leq j \leq K$ , geralmente calculado como a média dos seus elementos;
- $\mathbf{x}_i = (x_{i1}, \dots, x_{im}) \in R^m$  é o registro  $i$ ,  $1 \leq i \leq N$ , da base de dados;
- $d(\mathbf{x}_i, \mathbf{c}_j)$  é a função distância entre o registro  $\mathbf{x}_i$  e o centro  $\mathbf{c}_j$  do seu grupo.



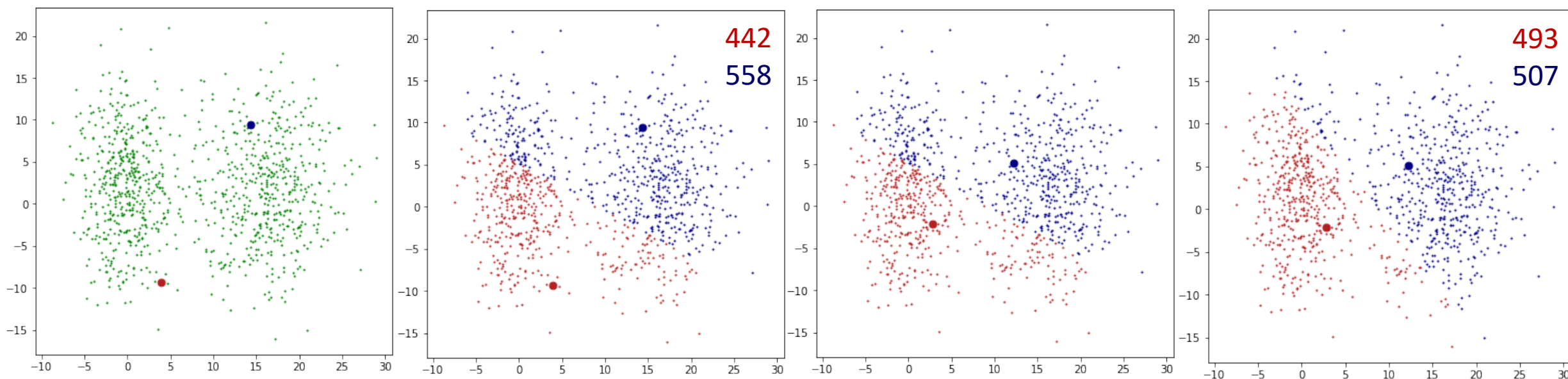
## Agrupamentos pelos centroides: *k-means*

O desempenho é bom quando a suposição se mostra verdadeira. Caso contrário, os resultados podem ser bastante ruins. Apesar disso, ele se destaca por ser bastante simples e convergir rapidamente.



## Agrupamentos pelos centroides: *k-means*

Inicialmente, os centros são alocados aleatoriamente ou arbitrariamente e as distâncias dos dados aos centros são calculadas.



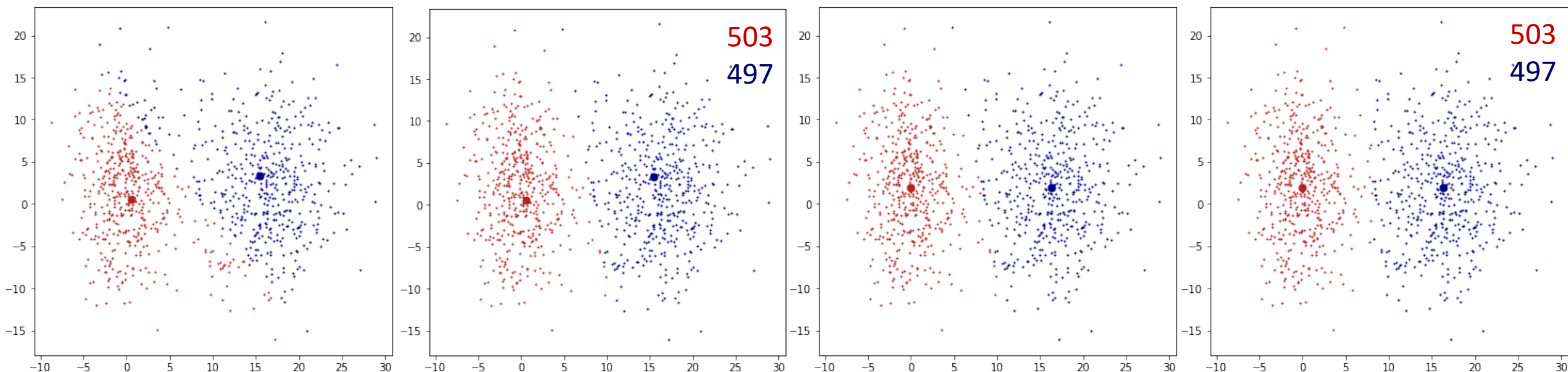
Em seguida, os dados são alocados aos centros mais próximos, formando os grupos.

Após as alocações, as posições dos centros são ajustadas pelas médias das coordenadas dos seus registros.

Em seguida, são efetuadas realocações dos registros aos grupos considerando as novas localizações dos centros.

## Agrupamentos pelos centroides: *k-means*

Esses dois passos (alocação e reposicionamento) são repetidos até que não haja mais mudanças significativas nas localizações dos centros.



Algoritmos particionais pelos centroides devem ser executados diversas vezes, com diferentes sementes aleatórias e quantidades de grupos, necessitando das medidas de validação de grupos para indicação do particionamento mais adequado.



## Agrupamentos pelos centroides: *k-means*

O *k-means* é sensível à alocação inicial dos centroides, pois estes podem não estar bem distribuídos. Então, pode-se:

- Executar diversas execuções com diferentes inicializações e selecionar o agrupamento que minimiza a função de dissimilaridade;
- Usar alguma heurística de inicialização dos centroides (registros, particionamento aleatório, afastados).

O *k-means* é sensível a *outliers*, pois estes atrairão os centroides para regiões do espaço com pouca representatividade. Assim, outros algoritmos particionais pelos centroides foram propostos para produzir centros mais robustos às presenças dos *outliers*. Algumas alternativas são:

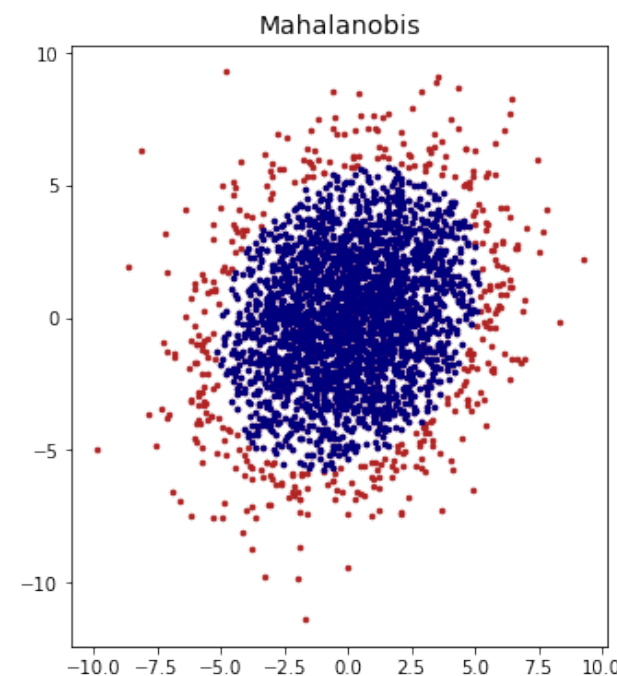
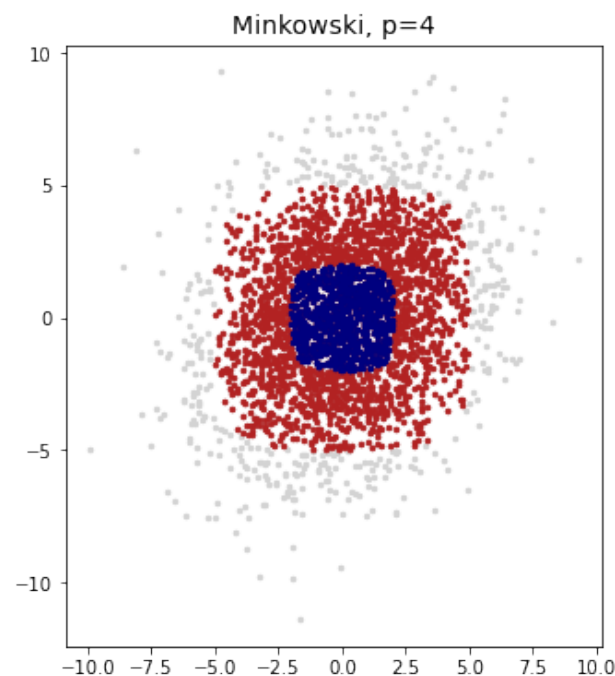
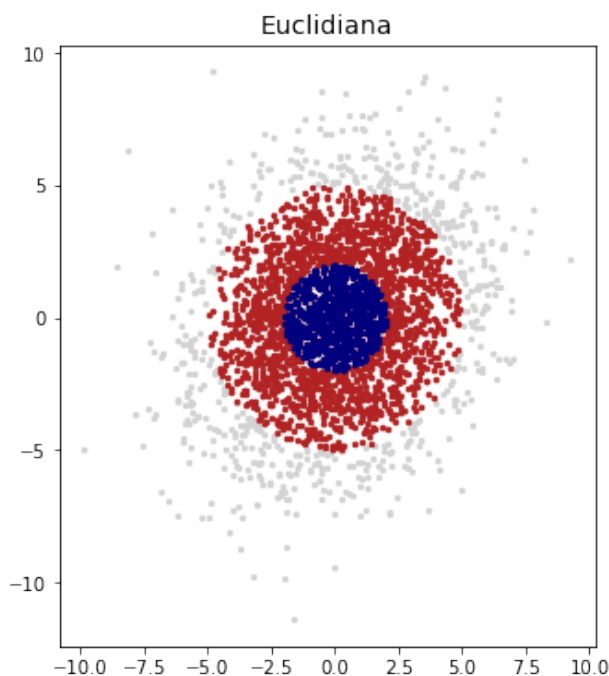
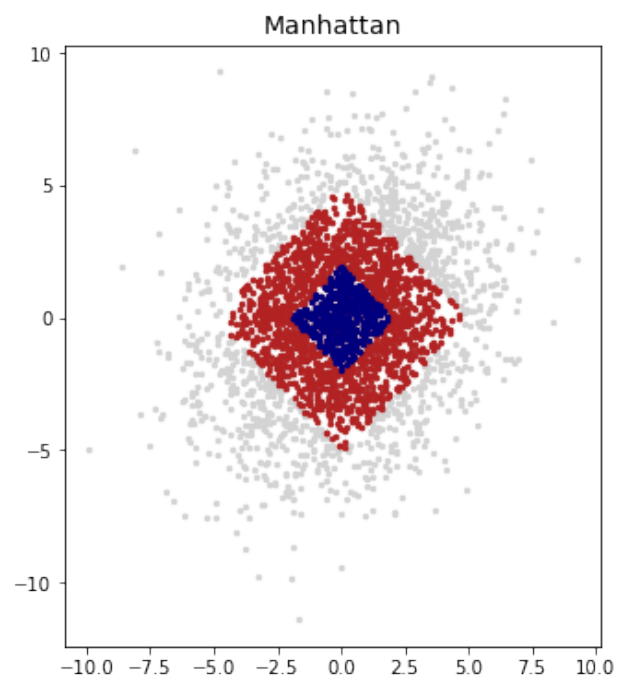
- *k-medians* ou *k-medianas*, em que os centros dos grupos são estimativas das medianas geométricas dos seus elementos;
- *k-medoids* ou *k-medoides*, em que os centros dos grupos são registros que minimizam a função de dissimilaridade dentro de seus grupos.

Em ambos os casos, a mudança de centro ocorre ou de modo heurístico (*k-medians*) ou de modo aleatório (*k-medoids*).



## Agrupamentos pelos centroides: *k-means*

Em algoritmos que utilizam distâncias, como os particionais pelos centroides, a função distância impacta a alocação dos dados aos grupos. Para o *k-means* e para o *k-medians*, as adequadas são a Euclidiana e a Manhattan, respectivamente.



$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_1 = \sum_{i=1}^n |u_i - v_i|$$

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_2 = \sqrt{\sum_{i=1}^n (u_i - v_i)^2}$$

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_p = \sqrt[p]{\sum_{i=1}^n |u_i - v_i|^p}$$

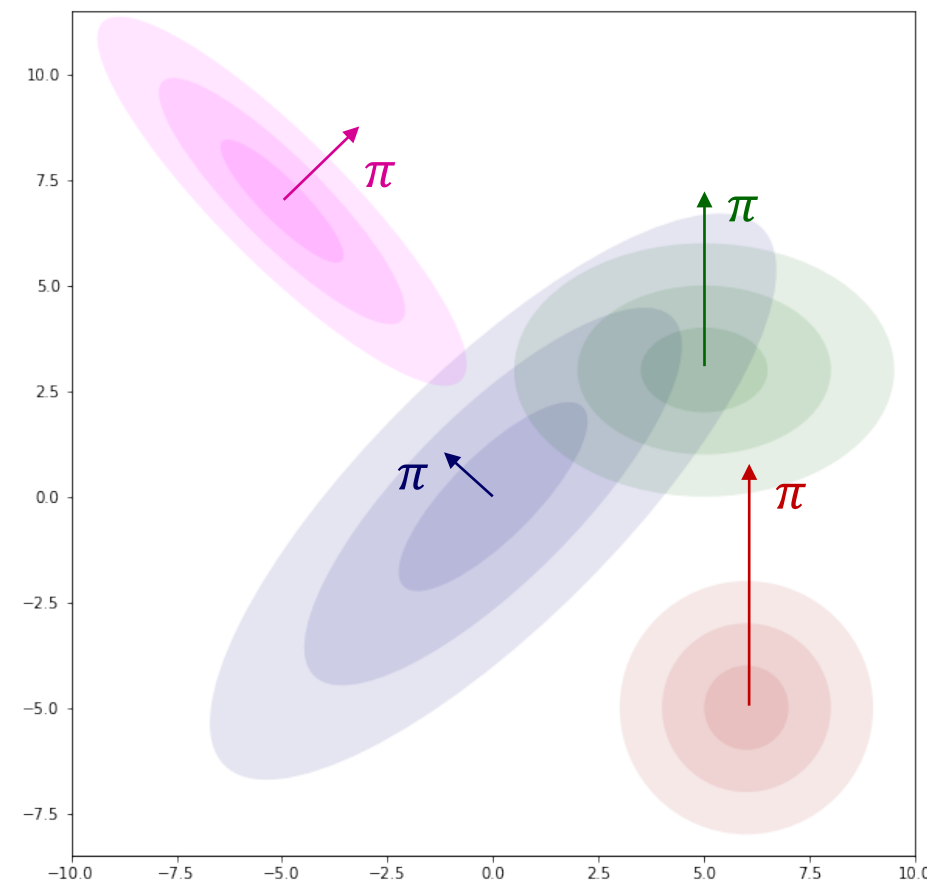
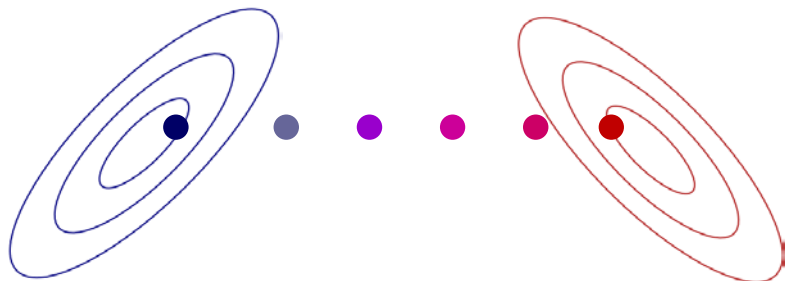
$$d(\mathbf{u}, \mathbf{v}) = \sqrt{(\mathbf{u} - \mathbf{v})\Sigma^{-1}(\mathbf{u} - \mathbf{v})^T}$$

## Agrupamentos probabilísticos: mistura gaussiana

Diferentemente do particionamento pelos centroides, em que estes são utilizados para representar os grupos e determinar inclusões ou alocações de novos elementos, o agrupamento probabilístico caracteriza os grupos por distribuições de probabilidades e as usa para novas alocações.

Uma das modelagens mais populares é a **mistura gaussiana**, em que uma coleção de gaussianas, cada uma descrevendo a localização e a dispersão de cada grupo, são combinadas para produzir a distribuição de probabilidade global.

Uma mistura é um modelo probabilístico que procura capturar e representar possíveis subconjuntos populacionais implícitos nos dados. A distribuição resultante da mistura sugere a distribuição populacional a partir dos dados observados.



## Agrupamentos probabilísticos: mistura gaussiana

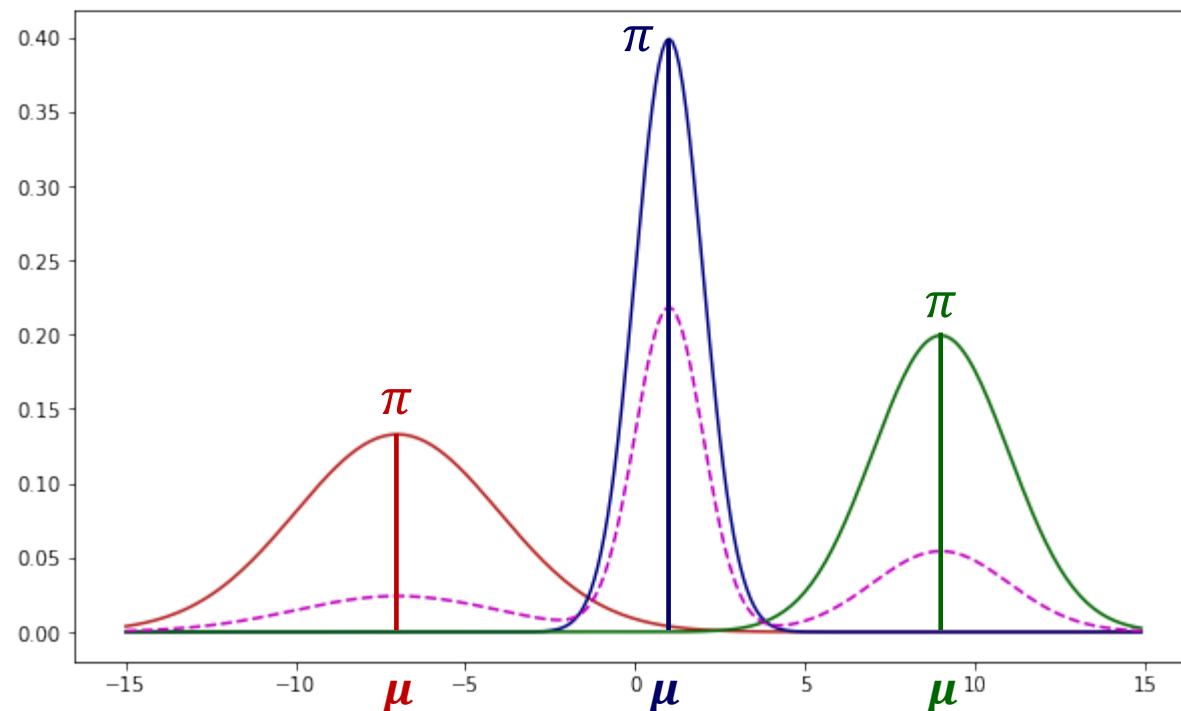
Cada gaussiana é caracterizada por peso ( $\pi$ ), vetor média ( $\mu$ ) e matriz de covariâncias ( $\Sigma$ ). O peso atribuído a cada gaussiana é associado à frequência relativa da cardinalidade do grupo, ou, em outras palavras, proporcional à probabilidade da sua média, de modo que a soma de todos os pesos resulte em 1.

Qual a probabilidade de um objeto  $i$  pertencer a um grupo  $j$ ?

$$p(x_i \in \text{grupo}_j) = \pi_j$$

Supondo que o objeto  $i$  pertença a um grupo  $j$ , qual a probabilidade de se encontrar outro objeto como  $i$  em  $j$ ?

$$p(x_i | x_i \in \text{grupo}_j) = \mathcal{N}(x_i | \mu_j, \Sigma_j)$$



Essas duas probabilidades são utilizadas para caracterizar a incerteza sobre a alocação de  $i$  em  $j$ .

## Agrupamentos probabilísticos: mistura gaussiana

Assim, para cada grupo  $j$ , o aprendizado especificará seu peso  $\pi_j$  e sua gaussiana (por seus parâmetros  $\boldsymbol{\mu}_j$  e  $\boldsymbol{\Sigma}_j$ ). O algoritmo mais comum para encontrar a coleção de  $\{(\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \forall j\}$  que melhor se ajuste aos dados observados é o de expectativa-maximização ou *expectation-maximization* (EM).

O EM procura identificar a coleção  $\{(\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \forall j\}$  que maximize a probabilidade de observação dos dados (de treinamento)  $P(\mathbf{X} \mid \pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ :

$$p(\mathbf{x}_i) = \sum_{j=1}^k \pi_j \cdot \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{probabilidade de observação do registro } \mathbf{x}_i \text{ considerando } (\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

$$P(\mathbf{X} \mid \pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \prod_{i=1}^n p(\mathbf{x}_i) \quad \text{probabilidade de observação da base de dados } \mathbf{X} \text{ considerando } (\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

Maximizar  $P(\mathbf{X} \mid \pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$  em relação a  $(\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$  significa caminhar na direção do seu gradiente e igualá-lo a zero (ponto de máximo), produzindo as estimativas de máxima verossimilhança dos parâmetros.

## Agrupamentos probabilísticos: mistura gaussiana

Como o produto presente em  $P(\mathbf{X} \mid \pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$  resulta em valores muito pequenos, na prática usa-se o logaritmo desta função, o que simplifica o cálculo de gradientes:

$$\ln \left( P(\mathbf{X} \mid \pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \right) = \ln \prod_{i=1}^n p(\mathbf{x}_i) = \sum_{i=1}^n \ln(p(\mathbf{x}_i)) = \sum_{i=1}^n \ln \left( \frac{\exp \left( -\frac{(\mathbf{x}_i - \boldsymbol{\mu}_j)^T \cdot \boldsymbol{\Sigma}_j^{-1} \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)}{2} \right)}{\sqrt{(2\pi)^d \cdot |\boldsymbol{\Sigma}_j|}} \right)$$

sendo  $d$  a dimensão do problema.

Sabendo-se que maximizar uma função gaussiana ou o seu logaritmo produz o mesmo resultado, ou seja, a média, escolhe-se a estratégia de cálculo mais simples: usar o logaritmo.

Portanto, derivando-se a função de log-verossimilhança  $\ln \left( P(\mathbf{X} \mid \pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \right)$  em relação a cada parâmetro de  $(\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ , com a restrição  $\sum_{j=1}^k \pi_j = 1$ , e igualando estas derivadas a zero, serão produzidas as expressões de atualização dos valores  $(\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$  que irão maximizá-la, assim como o farão para a função de verossimilhança  $P(\mathbf{X} \mid \pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ .

## Agrupamentos probabilísticos: mistura gaussiana

Considerando a alocação de dados nos grupos dada por  $\gamma(\mathbf{x}_i \in \text{grupo}_j) \forall i, j$ :

(expectation)

$$\gamma(\mathbf{x}_i \in \text{grupo}_j) = \boxed{P(\mathbf{x}_i \in \text{grupo}_j | \mathbf{x}_i) = \frac{p(\mathbf{x}_i | \mathbf{x}_i \in \text{grupo}_j) \cdot p(\text{grupo}_j)}{p(\mathbf{x}_i)}} = \frac{\mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \cdot \pi_j}{p(\mathbf{x}_i)}$$

Teorema de Bayes

$$= \frac{\pi_j \cdot \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{p(\mathbf{x}_i)}$$

e a cardinalidade de cada grupo dada por  $N_j = \sum_{i=1}^N \gamma(\mathbf{x}_i \in \text{grupo}_j)$ , as expressões de atualização dos parâmetros são:

(maximization)

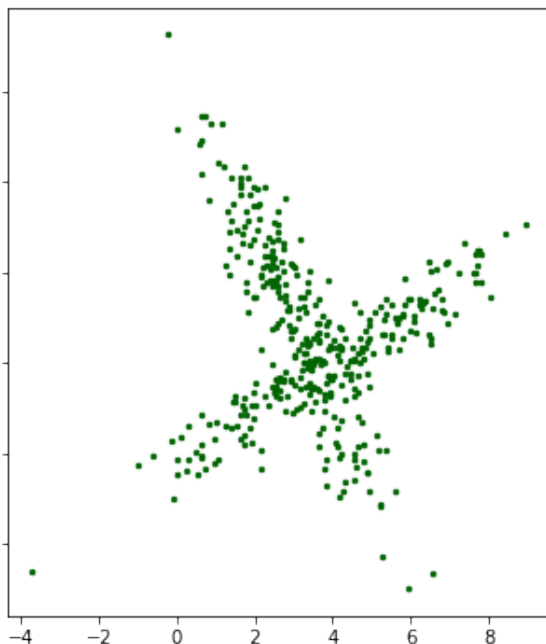
$$\boldsymbol{\mu}_j' = \frac{1}{N_j} \sum_{i=1}^N \gamma(\mathbf{x}_i \in \text{grupo}_j) \cdot \mathbf{x}_i \quad \boldsymbol{\Sigma}_j' = \frac{1}{N_j} \sum_{i=1}^N \gamma(\mathbf{x}_i \in \text{grupo}_j) \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j) \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \quad \pi_j' = \frac{N_j}{N}$$

$(\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$  e  $\gamma(\mathbf{x}_i \in \text{grupo}_j)$  são interdependentes e o algoritmo promove sucessivas atualizações de ambos até que algum critério de parada seja alcançado.

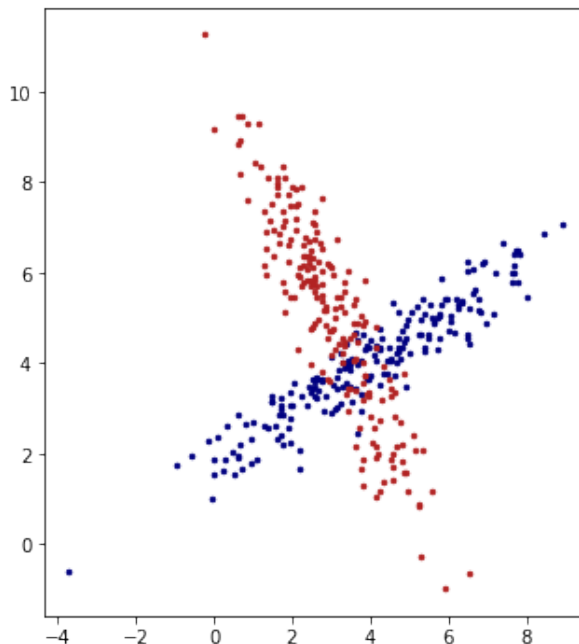
## Agrupamentos probabilísticos: mistura gaussiana

Exemplo:

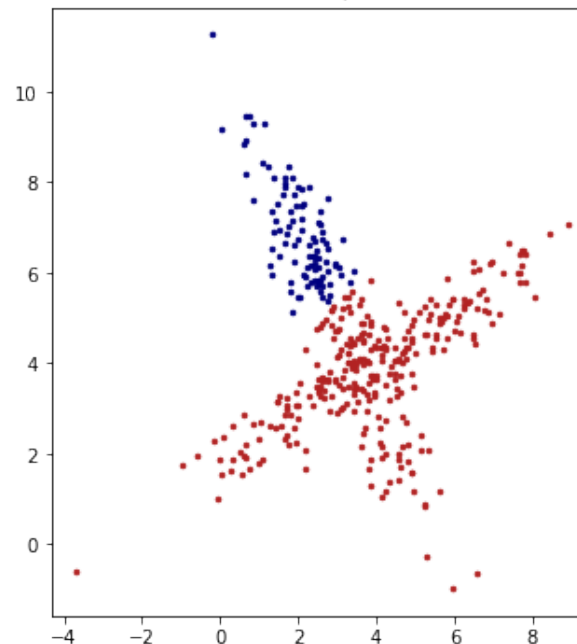
Dados não rotulados



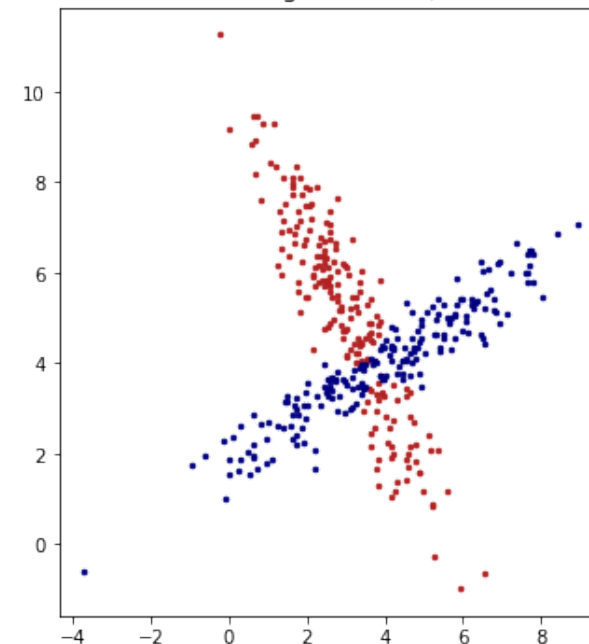
Dados rotulados



K-means, k=2



Misturas gaussianas, k=2



200 pontos em cada distribuição

$$\mu_1 = (4,0; 4,0)$$

$$\mu_2 = (3,0; 5,0)$$

$$\mu_1 = (2,0; 6,9)$$

$$\mu_2 = (3,9; 3,8)$$

Azuis: 97

Vermelhos: 303

$$\mu_1 = (4,0; 4,0)$$

$$\mu_2 = (3,0; 5,1)$$

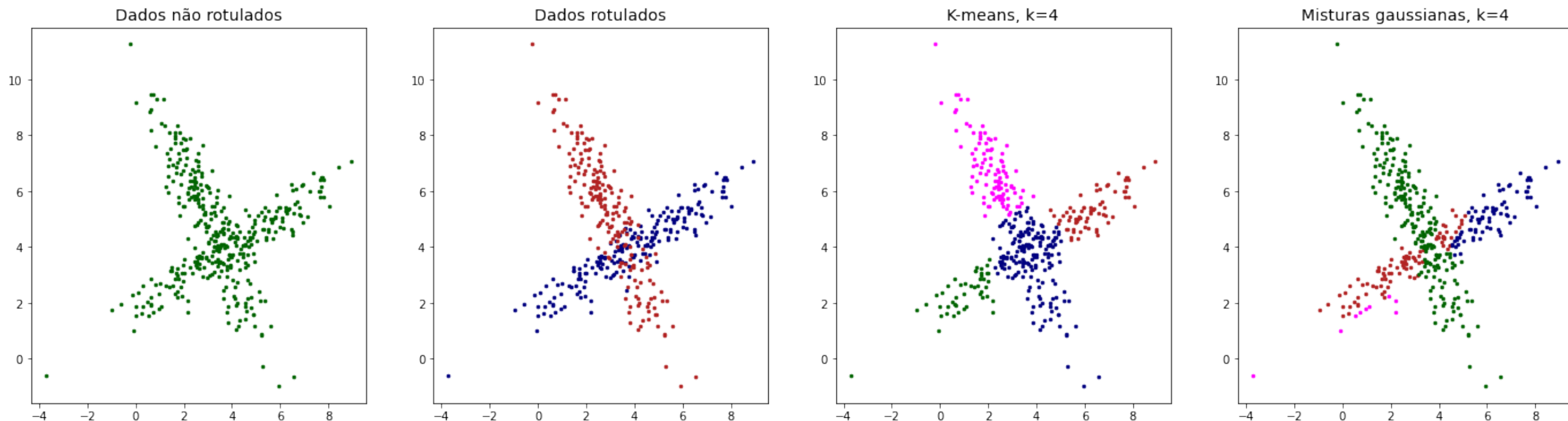
Azuis: 200

Vermelhos: 200



## Agrupamentos probabilísticos: mistura gaussiana

Exemplo:

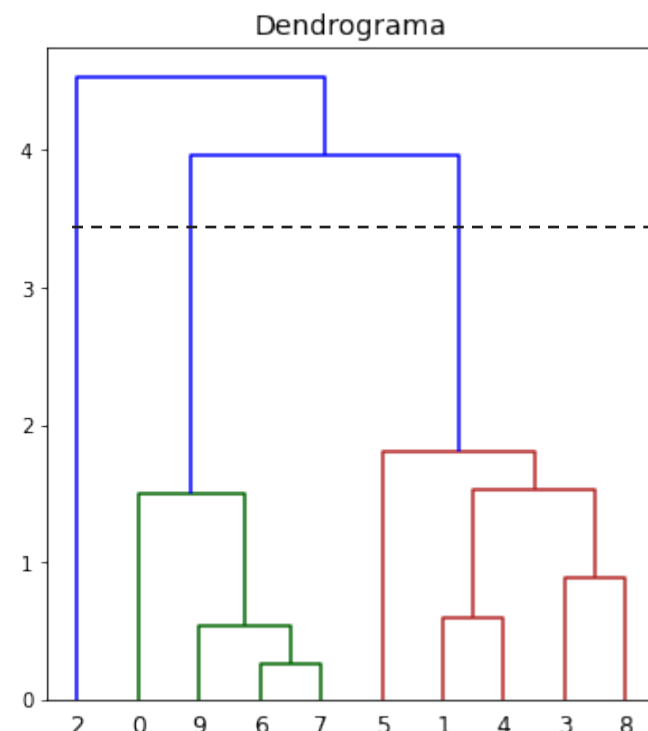
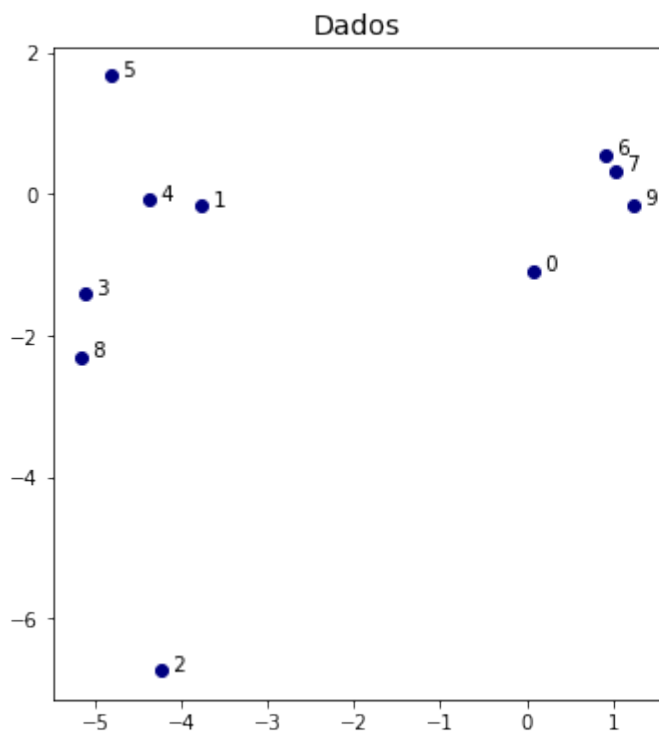


Raramente será possível visualizar os grupos para escolher o resultado mais interessante, o que reforça a importância de análise de **diversas** medidas de validação de grupos. Entretanto, elas não são robustas e a decisão final frequentemente caberá ao especialista dos dados.

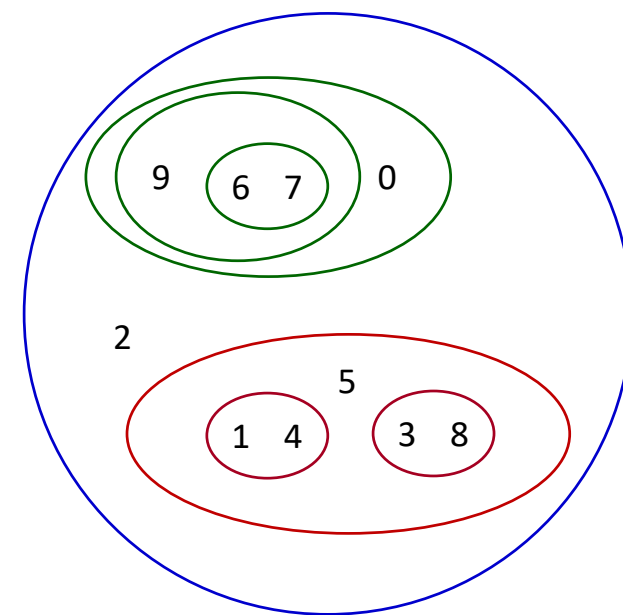
Índice	DB	CH	S
K-means, 2	0,838	213,632	0,395
K-means, 4	<b>0,693</b>	<b>373,026</b>	<b>0,472</b>
MisturaG, 2	2,827	33,933	<b>0,149</b>
MisturaG, 4	<b>1,543</b>	<b>77,495</b>	0,145

## Agrupamentos aglomerativos

Em agrupamentos hierárquicos, a hipótese é de que registros próximos têm uma relação mais forte do que registros mais distantes, criando-se então uma conexão entre eles e, conseqüentemente, um grupo. O raciocínio se estende aos grupos, que são também conectados. A representação dessas conexões é realizada com os **dendrogramas**.



4.529  
3.957  
1.811  
1.529  
1.496  
0.886  
0.596  
0.534  
0.255

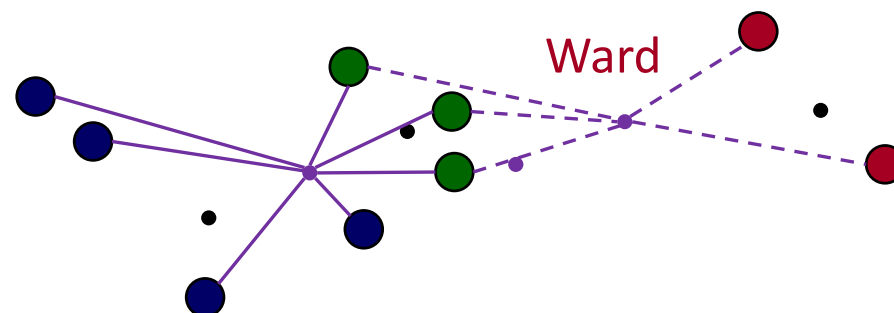
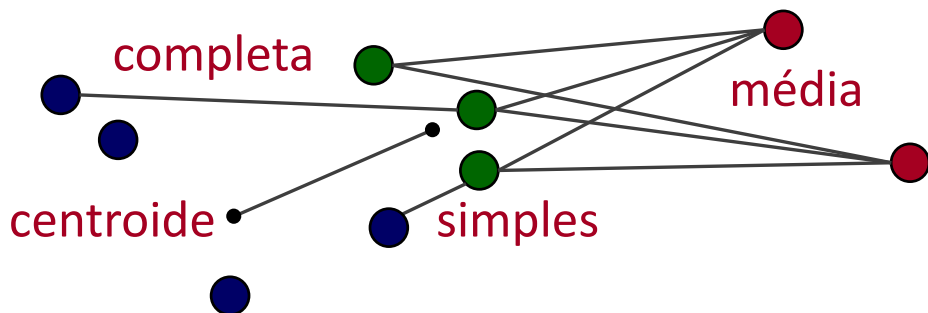


## Agrupamentos aglomerativos

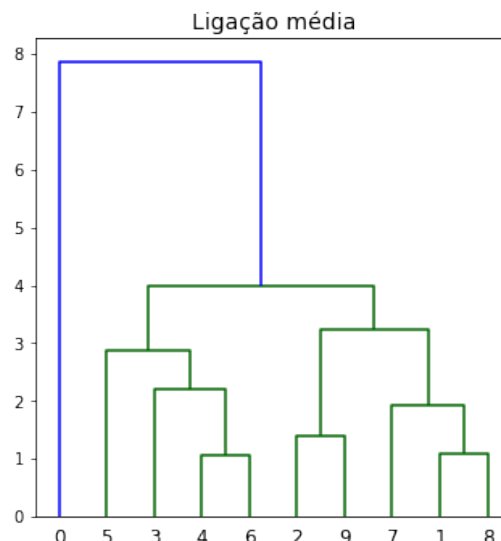
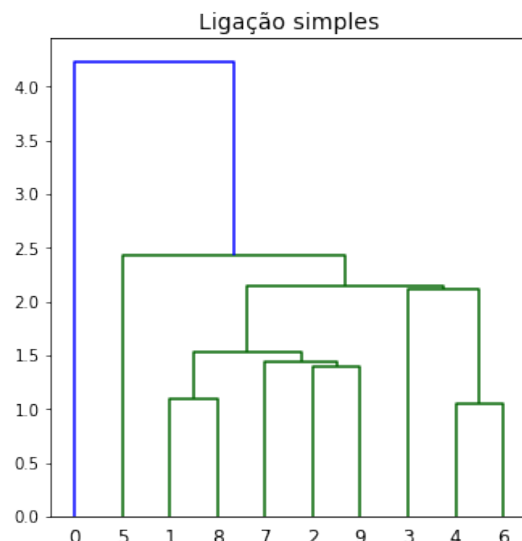
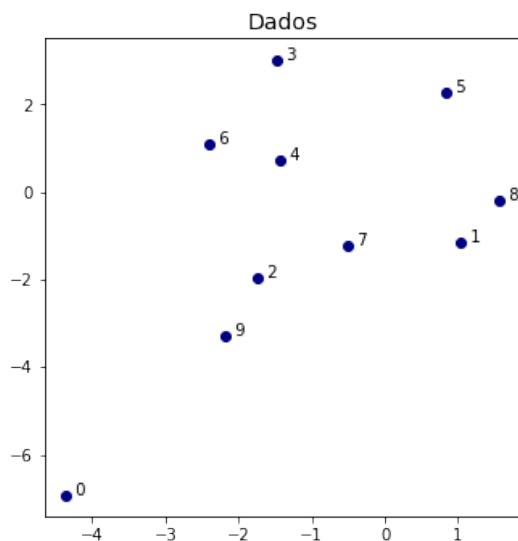
No início, cada registro é um grupo. A cada iteração, unir o par de grupos mais próximos, segundo a medida de similaridade utilizada e o critério adotado, até que todos os dados estejam em um único grupo.

Critérios de formação de grupos:

- Ligação simples (*single linkage*): os que tiverem par de elementos mais próximos, um de cada grupo;
- Ligação média (*average linkage*): a partir da menor distância média entre os pares de elementos, um de cada grupo;
- Ligação completa (*complete linkage*): a partir da menor distância máxima entre os pares de elementos, um de cada grupo;
- Centróide: os que tiverem menor distância entre os seus centróides;
- Ward: os que minimizam a soma de diferenças quadradas intragrupos



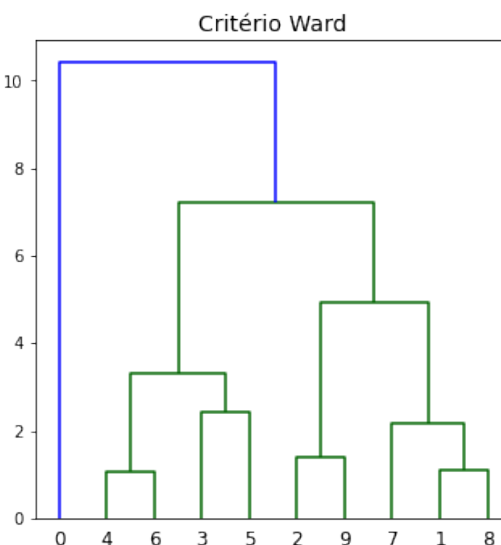
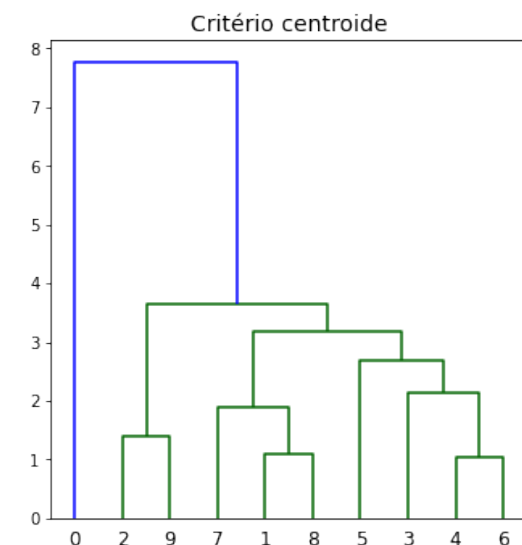
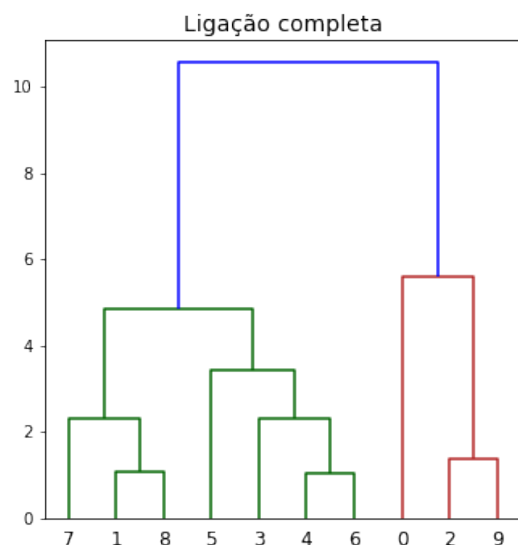
## Agrupamentos aglomerativos



$$d(G_i, G_j) = \min_{u \in G_i, v \in G_j} d(u, v)$$

$$d(G_i, G_j) = \frac{1}{|G_i| \cdot |G_j|} \cdot \sum_{u \in G_i, v \in G_j} d(u, v)$$

$$d(G_i, G_j) = \max_{u \in G_i, v \in G_j} d(u, v)$$



$$d(G_i, G_j) = d(c_i, c_j)$$

$$d(G_i \cup G_j, G_k) = \frac{(|G_i| + |G_k|)}{|G_i| + |G_j| + |G_k|} \cdot d(c_i, c_k) + \frac{(|G_j| + |G_k|)}{|G_i| + |G_j| + |G_k|} \cdot d(c_j, c_k) - \frac{|G_k|}{|G_i| + |G_j| + |G_k|} \cdot d(c_i, c_j)$$

## Agrupamentos divisivos

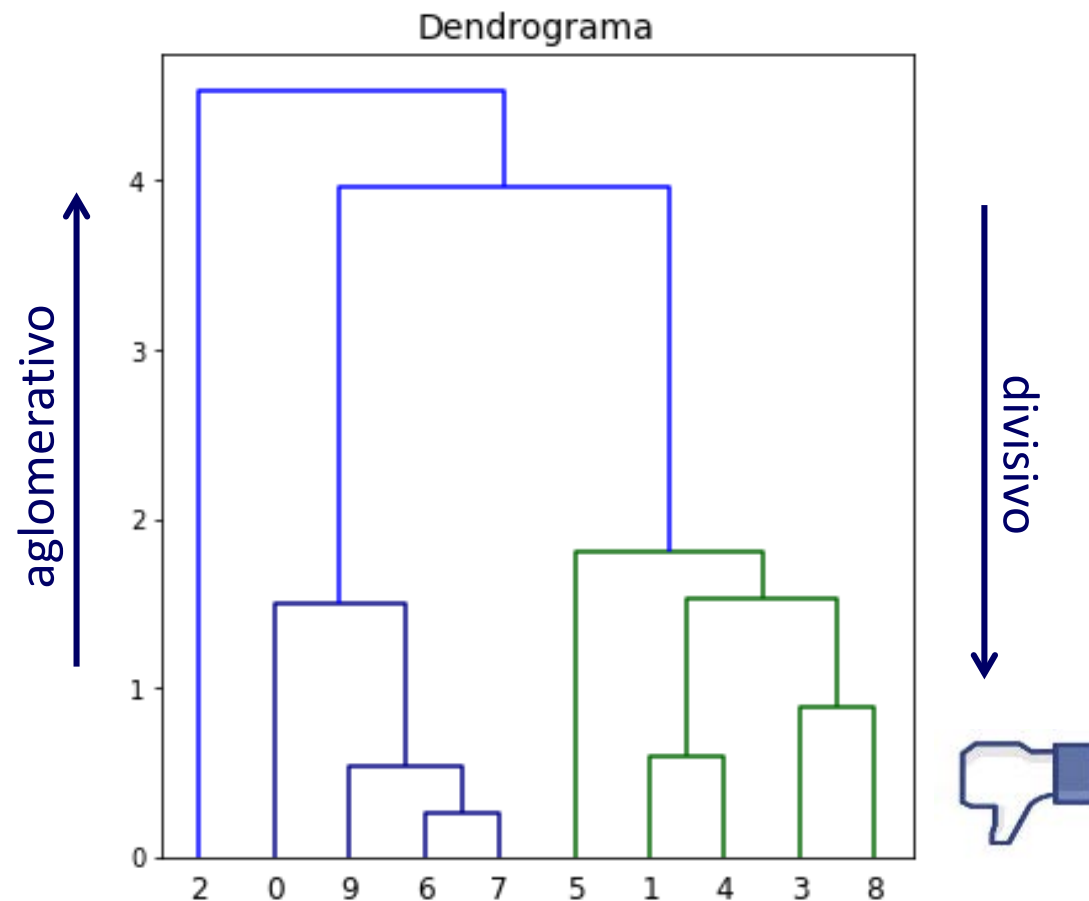
Inicialmente, os dados estão em um único grupo e o registro mais afastado dos demais é excluído do grupo, formando outro.

A seguir, os dados são distribuídos nos dois grupos de acordo com a proximidade a cada grupo.

O processo continua com o grupo com maior diâmetro, isto é, com a maior distância entre dois elementos quaisquer.

Após cada divisão, sempre haverá a redistribuição dos registros nos grupos encontrados.

É bastante lento para grandes bases de dados, mais do que os aglomerativos, inclusive. Algumas tentativas são realizadas usando-se o *k-means* para fazer as divisões.



## Visão geral

É a tarefa de descoberta de relacionamentos relevantes entre objetos ou variáveis de uma base de dados, sejam elas transacionais, relacionais ou não estruturadas. A relevância é determinada em função da frequência do relacionamento.

O resultado da associação é uma base de regras que sugerem relações de causa e consequência entre os objetos estudados. Como já apresentado anteriormente, a estrutura geral de uma regra de associação com apenas um objeto na conclusão é:

antecedente

SE objeto<sub>1</sub> E ... E objeto<sub>N</sub>, ENTÃO objeto<sub>j</sub> em que  $j \notin \{1, \dots, N\}$

consequente

Se as variáveis não forem discretas, então será necessário discretizá-las pois as frequências dos seus valores precisam ser detectáveis, algo que raramente ou nunca ocorreria em uma variável contínua.

SE  $\text{valor}_{1i} \leq \text{variável}_1 \leq \text{valor}_{1f}$  E ... E  $\text{valor}_{ni} \leq \text{variável}_n \leq \text{valor}_{nf}$ ,  
ENTÃO  $\text{valor}_{ji} \leq \text{variável}_j \leq \text{valor}_{jf}$ ,  $j \notin \{1, \dots, n\}$

## Métricas básicas e independência

Cobertura: Informa a frequência , absoluta ou relativa, do antecedente nos dados

Suporte: Informa a frequência, absoluta ou relativa, da ocorrência simultânea dos objetos

Confiança:  $[0,1]$ : Informa a proporção de registros relacionados ao antecedente em que o objeto do consequente foi observado.

$$\text{Confiança} = \frac{\text{Suporte}}{\text{Suporte}_{\text{Antecedente}}} = \frac{\text{Suporte}}{\text{Cobertura}}$$

Há independência entre os objetos do antecedente e do consequente se é observada a seguinte igualdade:

$$\text{Suporte}_{\text{Regra}} = \overbrace{\text{Suporte}_{\text{Antecedente}} \times \text{Suporte}_{\text{Consequente}}}^{\text{cobertura}}$$

suporte esperado

Se  $\text{Suporte}_{\text{Regra}} < \text{Suporte}_{\text{Esperado}}$  ,  
a dependência é **negativa**.

Se  $\text{Suporte}_{\text{Regra}} > \text{Suporte}_{\text{Esperado}}$  ,  
a dependência é **positiva**.



## Métricas básicas e independência

### Exemplo

Dados
Ana, Antônio, João, José, Maria, Sandro
Ana, Maria, Túlio
Ana, Antônio, José, Maria, Salvador, Túlio
Ana, Antônio, Elvis, Salvador, Sandro, Túlio
João, Maria, Sandro, Túlio
Ana, Antônio, João, Maria, Sandro, Túlio
Ana, Antônio, João, Túlio
Ana, Antônio, João, José, Maria, Salvador

1. As presenças de Maria e João reduzem a expectativa de presença de Ana;
2. A presença de Ana aumenta a expectativa de presença de Antônio;
3. A presença de Sandro não interfere na presença de Túlio;
4. A presença de José pouco interfere na presença de João.

Regras	Suporte	Cobertura	Confiança	Suporte Conseq.	Suporte Esperado	
João $\Rightarrow$ José	2 ou 0,250	5 ou 0,625	0,400	0,375	0,234	+
Maria, João $\Rightarrow$ Ana	3 ou 0,375	4 ou 0,500	0,750	0,875	0,438	-
Ana $\Rightarrow$ Antônio	6 ou 0,750	7 ou 0,875	0,857	0,750	0,656	+
Elvis, Salvador $\Rightarrow$ Sandro	1 ou 0,125	1 ou 0,125	1,000	0,500	0,063	+
Sandro $\Rightarrow$ Túlio	3 ou 0,375	4 ou 0,500	0,750	0,750	0,375	=
Antônio, Túlio $\Rightarrow$ Ana	4 ou 0,500	4 ou 0,500	1,000	0,875	0,438	+
Elvis $\Rightarrow$ Maria	0	1 ou 0,125	0,000	0,750	0,094	-
João, José, Maria $\Rightarrow$ Ana	2 ou 0,250	2 ou 0,250	1,000	0,875	0,219	+
Salvador $\Rightarrow$ Sandro	1 ou 0,125	2 ou 0,250	0,500	0,500	0,125	=
José $\Rightarrow$ João	2 ou 0,250	3 ou 0,375	0,667	0,625	0,234	+

As dependências positivas são as mais interessantes e, para elas, são definidas as medidas de interesse (*lift*, novidade, convicção) das regras de associação.

## Métricas de interesse

*Lift*:  $[0, +\infty[$ : mede o aumento da ocorrência do consequente dado o antecedente (em outras palavras, a força da associação além do esperado). Bom quando  $> 1$ .

$$Lift = \frac{P(Consequente|Antecedente)}{P(Consequente)} = \frac{Confiança}{Suporte_{Consequente}} = \frac{Suporte}{Suporte_{Esperado}}$$

*Leverage* ou *Novidade*:  $[-1, +1]$ : mede a diferença entre os suportes real e esperado. Bom quando  $> 0$ .

$$Novidade = Suporte - Suporte_{Esperado}$$

*Convicção*:  $[0, +\infty[$ : mede o grau de implicação da regra. Bom quando  $> 1$ .

$$Convicção = \frac{\overbrace{P(Antecedente) \cdot P(\overline{Consequente})}^{\text{falhas esperadas supondo indepedência}}}{\underbrace{P(Antecedente \cap \overline{Consequente})}_{\text{falhas observadas nos dados}}} = \frac{1 - Suporte_{Consequente}}{1 - Confiança}$$

Quando convicção  $> 1$ , a falha da regra ocorre com uma frequência muito inferior ao que seria esperado ao acaso, pela suposição de independência.

## Métricas de interesse

Regras com dependência +	Cobertura	Sup.Conseq.	Sup.Esperado	Suporte	Confiança	Lift	Novidade	Convicção
João $\Rightarrow$ José	0,625	0,375	0,234	0,250	0,400	1,067	0,016	1,042
Ana $\Rightarrow$ Antônio	0,875	0,750	0,656	0,750	0,857	1,143	0,094	1,748
Elvis, Salvador $\Rightarrow$ Sandro	0,125	0,500	0,063	0,125	1,000	2,000	0,063	-
Antônio, Túlio $\Rightarrow$ Ana	0,500	0,875	0,438	0,500	1,000	1,143	0,063	-
João, José, Maria $\Rightarrow$ Ana	0,250	0,875	0,219	0,250	1,000	1,143	0,031	-
José $\Rightarrow$ João	0,375	0,625	0,234	0,250	0,667	1,067	0,016	1,126

Regras com dependência -	Cobertura	Sup.Conseq.	Sup.Esperado	Suporte	Confiança	Lift	Novidade	Convicção
Maria, João $\Rightarrow$ Ana	0,500	0,875	0,438	0,375	0,750	0,857	-0,063	0,500
Sandro $\Rightarrow$ Túlio	0,500	0,750	0,375	0,375	0,750	1,000	0,000	1,000
Elvis $\Rightarrow$ Maria	0,125	0,750	0,094	0,000	0,000	0,000	-0,094	0,250
Salvador $\Rightarrow$ Sandro	0,250	0,500	0,125	0,125	0,500	1,000	0,000	1,000

## Algoritmo Apriori

É o algoritmo clássico da associação. Ele pode ser resumido em duas etapas:

- Descobrir os conjuntos de objetos frequentes;
- Gerar as regras a partir desses conjuntos.

Os objetos são chamados de **itens** (frequentes) e os conjuntos de itens são chamados de **conjuntos frequentes** (*itemsets*). Um suporte mínimo (**minsup**) é imposto aos conjuntos frequentes para a eliminação de objetos ou associações esporádicas.

A etapa inicial avalia apenas os conjuntos frequentes unitários. A cada iteração, novos conjuntos frequentes são formados pela união dos conjuntos da etapa anterior desde que haja somente a inclusão de um único item. Se não for possível a descoberta de mais conjuntos frequentes, o algoritmo para.

O estabelecimento de novos conjuntos frequentes obedece às seguintes propriedades de conjuntos:

- Monotonicidade: todo subconjunto não vazio de um conjunto frequente também é frequente;
- Anti-monotonicidade: todo superconjunto de um conjunto não frequente também não é frequente.

De cada conjunto frequente encontrado, as regras são formadas com um ou mais itens no conseqüente e os demais no antecedente. Regras com confianças inferiores a um valor mínimo (**minconf**) são descartadas.

## Algoritmo Apriori

Exemplo: Apriori com minsup=4

Dados
Ana, Antônio, João, José, Maria, Sandro
Ana, Maria, Túlio
Ana, Antônio, José, Maria, Salvador, Túlio
Ana, Antônio, Elvis, Salvador, Sandro, Túlio
João, Maria, Sandro, Túlio
Ana, Antônio, João, Maria, Sandro, Túlio
Ana, Antônio, João, Túlio
Ana, Antônio, João, José, Maria, Salvador

1

1-Itemsets	Suporte
Ana	7
Antônio	6
Elvis	1
João	5
José	3
Maria	6
Salvador	1
Sandro	4
Túlio	6

2

2-Itemsets	Suporte
Ana, Antônio	6
Ana, João	4
Ana, Maria	5
Ana, Sandro	3
Ana, Túlio	5
Antônio, João	4
Antônio, Maria	4
Antônio, Sandro	3
Antônio, Túlio	4
João, Maria	4
João, Sandro	3
João, Túlio	3
Maria, Sandro	3
Maria, Túlio	4
Sandro, Túlio	3

A etapa inicial avalia apenas os conjuntos frequentes unitários.

Objetos raros são descartados e permanecem os que atingirem o suporte mínimo.

A cada iteração, novos conjuntos frequentes são formados pela união dos conjuntos da etapa anterior desde que haja somente a inclusão de um único item.

## Algoritmo Apriori

Exemplo: Apriori com minsup=4

Dados
Ana, Antônio, João, José, Maria, Sandro
Ana, Maria, Túlio
Ana, Antônio, José, Maria, Salvador, Túlio
Ana, Antônio, Elvis, Salvador, Sandro, Túlio
João, Maria, Sandro, Túlio
Ana, Antônio, João, Maria, Sandro, Túlio
Ana, Antônio, João, Túlio
Ana, Antônio, João, José, Maria, Salvador

3

3-Itemsets	Suporte
Ana, Antônio, João	4
Ana, Antônio, Maria	4
Ana, Antônio, Túlio	4
Ana, João, Maria	3
Ana, João, Túlio	2
Ana, Maria, Túlio	3
Antônio, João, Maria	3
Antônio, João, Túlio	2
Antônio, Maria, Túlio	2
João, Maria, Túlio	2

4

4-Itemsets	Suporte
Ana, Antônio, João, Maria	3
Ana, Antônio, João, Túlio	2
Ana, Antônio, Maria, Túlio	2

Se não for possível a descoberta de mais conjuntos frequentes, o algoritmo para.

Itemsets	Suporte
Ana, Antônio	6
Ana, João	4
Ana, Maria	5
Ana, Túlio	5
Antônio, João	4
Antônio, Maria	4
Antônio, Túlio	4
João, Maria	4
Maria, Túlio	4
Ana, Antônio, João	4
Ana, Antônio, Maria	4
Ana, Antônio, Túlio	4

## Algoritmo Apriori

Exemplo: Apriori com minsup=4

Itemsets	Suporte
Ana, Antônio	6
Ana, João	4
Ana, Maria	5
Ana, Túlio	5
Antônio, João	4
Antônio, Maria	4
Antônio, Túlio	4
João, Maria	4
Maria, Túlio	4
Ana, Antônio, João	4
Ana, Antônio, Maria	4
Ana, Antônio, Túlio	4

De cada conjunto frequente encontrado, regras são formadas com um (escolha) dos itens no consequente e os demais no antecedente.

Ana  $\Rightarrow$  Antônio

Ana  $\Rightarrow$  João

Ana  $\Rightarrow$  Maria

Ana  $\Rightarrow$  Túlio

Antônio  $\Rightarrow$  Ana

João  $\Rightarrow$  Ana

Maria  $\Rightarrow$  Ana

Túlio  $\Rightarrow$  Ana

Antônio  $\Rightarrow$  João

Antônio  $\Rightarrow$  Maria

Antônio  $\Rightarrow$  Túlio

João  $\Rightarrow$  Antônio

Maria  $\Rightarrow$  Antônio

Túlio  $\Rightarrow$  Antônio

João  $\Rightarrow$  Maria

Maria  $\Rightarrow$  João

Túlio  $\Rightarrow$  Maria

Maria  $\Rightarrow$  Túlio

Ana e Antônio  $\Rightarrow$  João

Ana e João  $\Rightarrow$  Antônio

Antônio e João  $\Rightarrow$  Ana

Ana e Antônio  $\Rightarrow$  Maria

Ana e Maria  $\Rightarrow$  Antônio

Antônio e Maria  $\Rightarrow$  Ana

Ana e Antônio  $\Rightarrow$  Túlio

Ana e Túlio  $\Rightarrow$  Antônio

Antônio e Túlio  $\Rightarrow$  Ana

Regras com confianças inferiores a um valor mínimo são descartadas.  
Considerando minconf = 0,70:

27 regras encontradas

15 regras mantidas

12 regras descartadas



## Conjuntos e lógica *fuzzy*

São expansões da teoria de conjuntos e da lógica clássicas (booleanas) em que são possíveis as pertinências parciais a conjuntos e o raciocínio com incertezas subjetivas. Batizada em 1965 por Lotfi Zadeh, já era estudada desde 1930 por Jan Łukasiewicz e Alfred Tarski.

Desde então, a teoria *fuzzy* tem sido aplicada com sucesso em tomadas de decisão com relevantes e variadas incertezas e quando a representação do conhecimento é subjetiva. Notadamente, em aplicações de controle, sempre foi bastante utilizada devido à facilidade de incorporar e representar o conhecimento de especialistas.

Outra característica interessante é que ela pode ser facilmente usada por outros algoritmos, ampliando as possibilidades de modelagem.



booleana

INFORMATION AND CONTROL **8**, 338–353 (1965)

### Fuzzy Sets\*

L. A. ZADEH

*Department of Electrical Engineering and Electronics Research Laboratory,  
University of California, Berkeley, California*

A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one. The notions of inclusion, union, intersection, complement, relation, convexity, etc., are extended to such sets, and various properties of these notions in the context of fuzzy sets are established. In particular, a separation theorem for convex fuzzy sets is proved without requiring that the fuzzy sets be disjoint.



fuzzy

## Conjuntos e lógica *fuzzy*

Na teoria clássica, dados um conjunto  $A$  e o seu universo  $U$ ,  $A$  pode ser caracterizado pela enumeração dos seus elementos (quando possível), por uma sentença aberta no universo (uma afirmação sobre alguma propriedade dos seus elementos) ou por sua **função característica**:

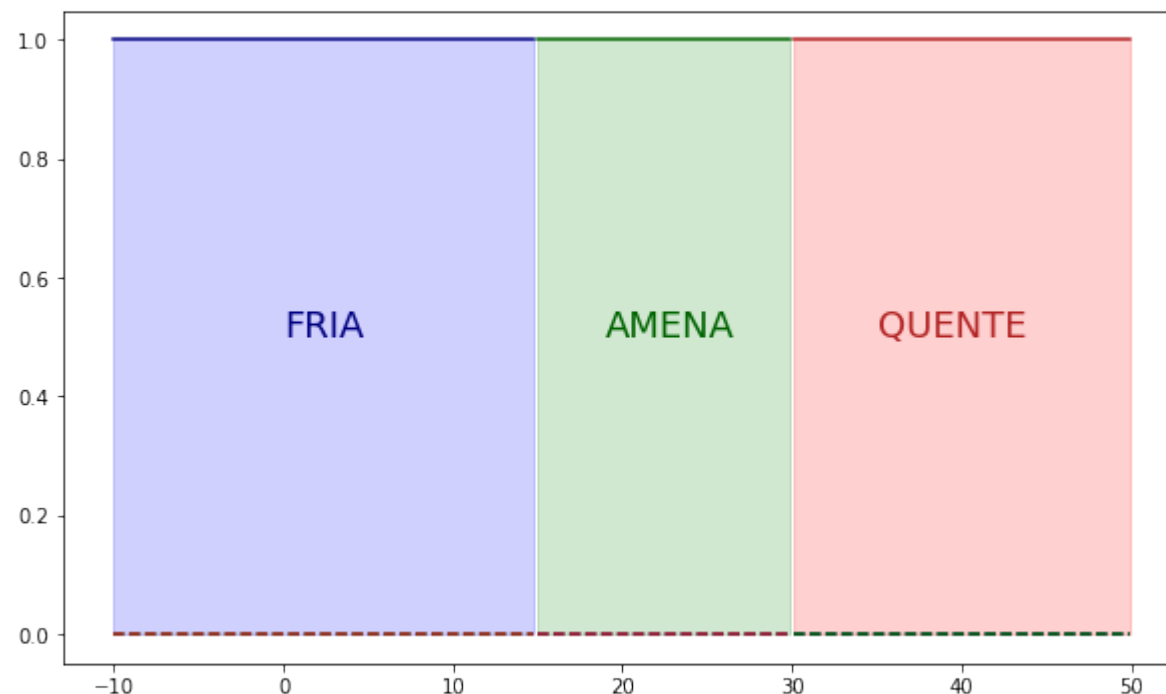
$$f_A: U \rightarrow \{0,1\} \qquad f_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

Exemplo:

$$f_{FRIA}(x) = \begin{cases} 1, & x < 15 \\ 0, & x \geq 15 \end{cases}$$

$$f_{AMENA}(x) = \begin{cases} 0, & x < 15 \\ 1, & 15 \leq x \leq 30 \\ 0, & x > 30 \end{cases}$$

$$f_{QUENTE}(x) = \begin{cases} 1, & x > 30 \\ 0, & x \leq 30 \end{cases}$$



## Conjuntos e lógica *fuzzy*

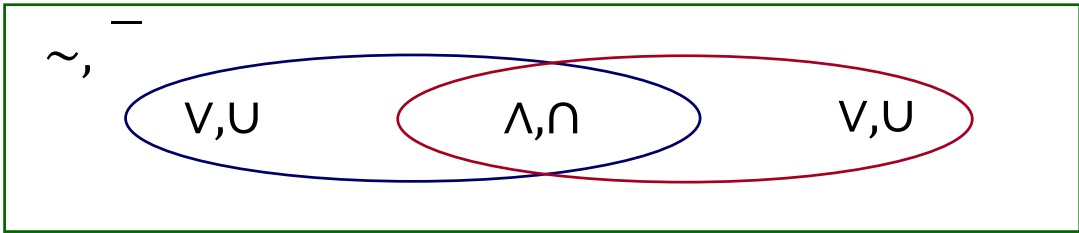
Na teoria clássica, um objeto não pode pertencer a um conjunto e ao complemento deste simultaneamente. Em outras palavras, a função característica do conjunto complementar ao conjunto A é dada por

$$f_{\bar{A}}: U \rightarrow \{0,1\} \qquad f_{\bar{A}}(x) = 1 - f_A(x)$$

Na teoria clássica, uma afirmação necessariamente será verdadeira (V) ou falsa (F), não havendo outra possibilidade. As afirmações são combinadas por operadores lógicos, associados a operações entre conjuntos.

$p$	$q$	$\sim p$	$p \wedge q$	$p \vee q$	$p \underline{\vee} q$	$p \leftrightarrow q$	$p \rightarrow q$
V	V	F	V	V	F	V	V
V	F	F	F	V	V	F	F
F	V	V	F	V	V	F	V
F	F	V	F	F	F	V	V

Operação lógica	Operação de conjunto
Negação	Complementação
Conjunção	Interseção
Disjunção	União



## Conjuntos e lógica *fuzzy*

Na teoria *fuzzy*, dados um conjunto  $A$  e o seu universo  $U$ , a **função de pertinência** de  $A$  informa o quanto um elemento pertence a ele:

$$\mu_A: U \rightarrow [0,1] \quad \mu_A(x)$$

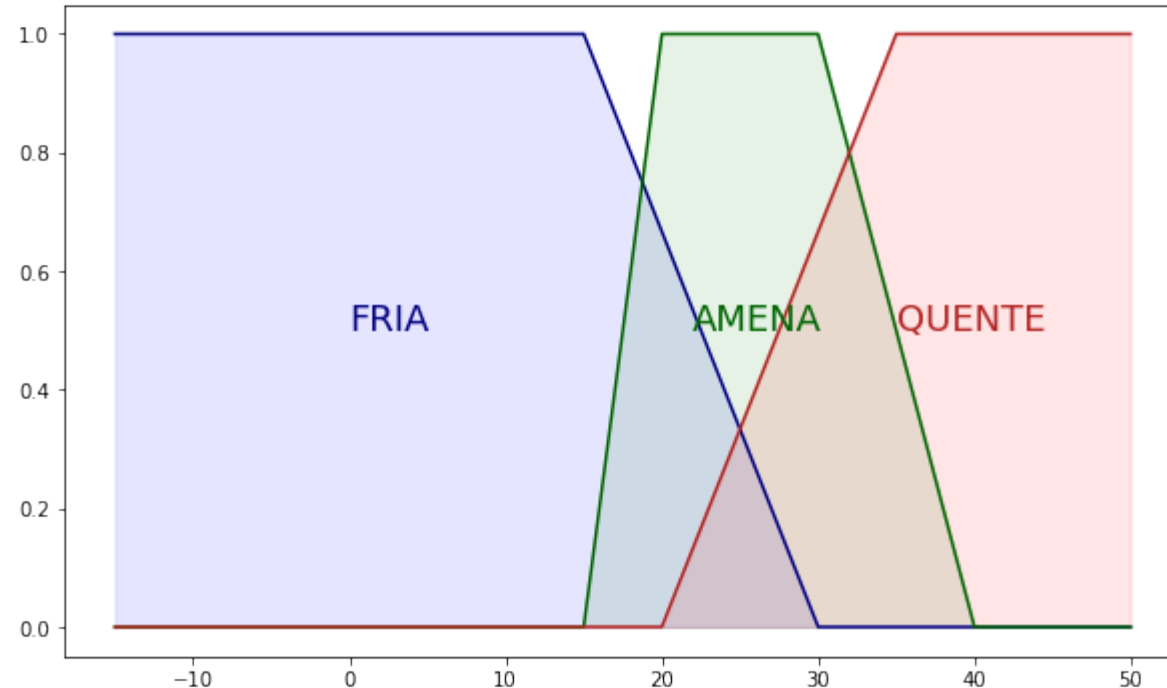
Observe que a imagem da função não é mais binária, mas um subconjunto do intervalo real  $[0,1]$ . Isso significa que os elementos podem ter pertinências parciais aos conjuntos.

Exemplo:

$$\mu_{FRIA}(x) = \begin{cases} 1, & x \leq 15 \\ (30 - x)/15, & 15 < x < 30 \\ 0, & x \geq 30 \end{cases}$$

$$\mu_{AMENA}(x) = \begin{cases} (x - 15)/5, & 15 < x < 20 \\ 1, & 20 \leq x \leq 30 \\ (40 - x)/10, & 30 < x < 40 \\ 0, & \text{caso contrário} \end{cases}$$

$$\mu_{QUENTE}(x) = \begin{cases} 1, & x \geq 35 \\ (x - 20)/15, & 20 < x < 35 \\ 0, & x \leq 20 \end{cases}$$



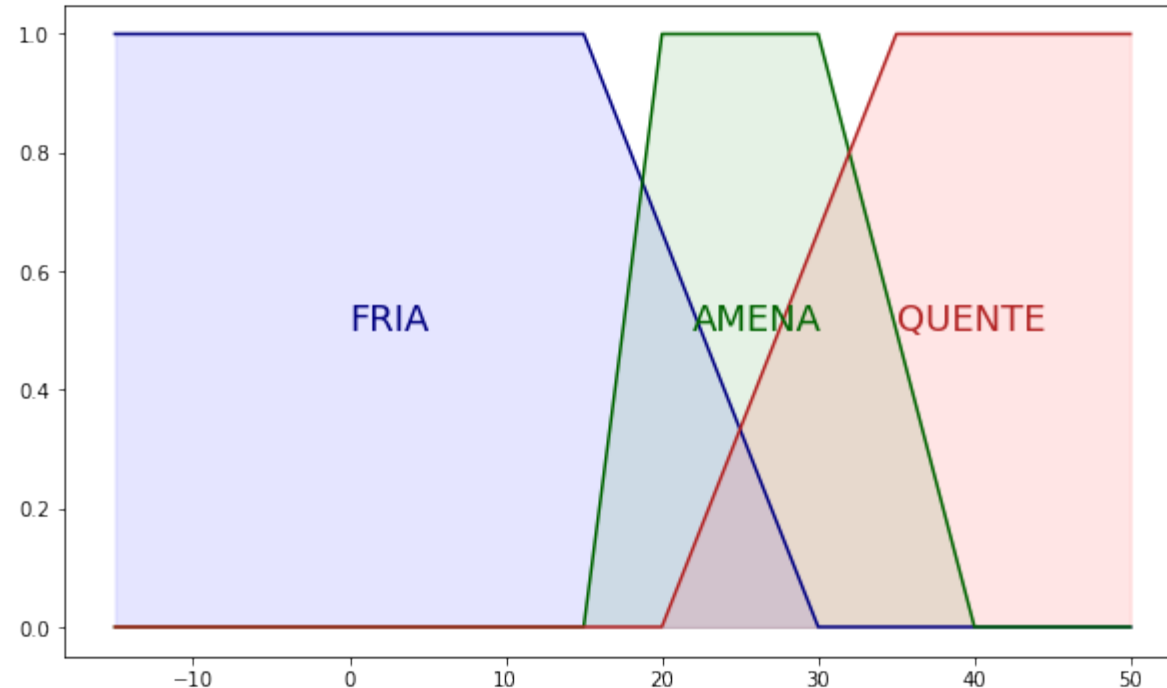
## Conjuntos e lógica *fuzzy*

A partir das pertinências parciais aos conjuntos, obtém-se pertinências não nulas aos complementos, ou seja, os objetos pertencem aos conjuntos e aos seus complementos simultaneamente, cenário contraditório na teoria clássica.

Supondo conjuntos mutuamente exclusivos (na visão clássica) formando uma partição do universo, é possível observar na matemática *fuzzy* que:

- nem sempre a soma das pertinências resulta em 1;
- nem sempre a interseção é vazia, o que caracteriza a incerteza chamada de **ambiguidade**;
- nem sempre a união dos conjuntos resultará no universo.

Além da ambiguidade, outra incerteza é manifestada na escolha dos parâmetros dos conjuntos, sendo chamada de **vaguidade**.



## Agrupamento *fuzzy*

Em agrupamentos *fuzzy* os elementos pertencem parcialmente a diversos grupos, sendo a pertinência inversamente proporcional à distância ao centro do grupo.

O algoritmo tradicional é o *fuzzy c-means*, que pode ser interpretado como a versão *fuzzy* do *k-means*:

- Inicialmente, as pertinências aos  $c$  grupos são arbitradas;

- A partir dessas “atribuições”, os centroides são calculados:

$$\mathbf{c}_k = \frac{\sum_x (\mu_k(\mathbf{x}))^m \cdot \mathbf{x}}{\sum_x (\mu_k(\mathbf{x}))^m}$$

- Após o posicionamentos dos centros, as pertinências são atualizadas:

- Com as novas pertinências, os centroides precisam ser recalculados;

- O processo se repete até que não sejam verificadas mudanças na matriz de pertinências ou na de centros.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|\mathbf{x}_i - \mathbf{c}_j\|}{\|\mathbf{x}_i - \mathbf{c}_k\|} \right)^{\frac{1}{m-1}}}$$

O parâmetro de forma  $m \geq 1$  é chamado de *fuzzifier* e ele indica o grau de interseção entre os grupos: quanto maior o seu valor, menor a pertinência de um elemento a um grupo e, assim, maiores as pertinências dele aos demais grupos. Quando o *fuzzifier* se aproxima de 1, o resultado é o particionamento não *fuzzy* da base de dados, como no *k-means*.

## Associação *fuzzy*

Em associação *fuzzy*, as variáveis quantitativas são discretizadas como na associação não *fuzzy*. Porém, em vez de o valor de uma variável pertencer a um único intervalo da partição, ele pertencerá a todos os intervalos com diferentes graus de pertinência.

Exemplo:

Partição *fuzzy*

$$\mu_{CURTA}(x) = \begin{cases} 1, & x \leq 1 \\ (3 - x)/2, & 1 < x < 3 \\ 0, & x \geq 3 \end{cases} \quad \mu_{MÉDIA}(x) = \begin{cases} x/2, & x \leq 2 \\ 4 - x/2, & x > 2 \end{cases} \quad \mu_{LONGA}(x) = \begin{cases} 0, & x < 1 \\ x - 1/2, & 1 < x < 3 \\ 1, & x \geq 3 \end{cases}$$

Registro	Variáveis		Partição tradicional						Partição <i>fuzzy</i>					
	temperatura	duração	fria	amena	quente	curta	média	longa	fria	amena	quente	curta	média	longa
1	15	0	1	0	0	1	0	0	1,00	0,00	0,00	1,00	0,00	0,00
2	34	3	0	0	1	0	1	0	0,00	0,60	0,93	0,00	0,50	1,00
3	23	2,5	0	1	0	0	1	0	0,47	1,00	0,20	0,00	0,75	0,75
4	40	3	0	0	1	0	1	0	0,00	0,00	1,00	0,00	0,50	1,00
5	45	1.8	0	0	1	0	1	0	0,00	0,00	1,00	0,20	0,90	0,40
6	30	4	0	0	1	0	0	1	0,00	1,00	0,67	0,00	0,00	1,00
7	28	2,5	0	1	0	0	1	0	0,13	1,00	0,53	0,00	0,75	0,75
8	22	2	0	1	0	0	1	0	0,53	1,00	0,13	0,00	1,00	0,50
9	20	1,4	0	1	0	0	1	0	0,67	1,00	0,00	0,60	0,70	0,20
10	19	0,5	0	1	0	1	0	0	0,73	0,80	0,00	1,00	0,25	0,00

Partição tradicional

$$f_{CURTA}(x) = \begin{cases} 1, & x < 1 \\ 0, & x \geq 1 \end{cases} \quad f_{MÉDIA}(x) = \begin{cases} 0, & x < 1 \\ 1, & 1 \leq x \leq 3 \\ 0, & x > 3 \end{cases} \quad f_{LONGA}(x) = \begin{cases} 1, & x > 3 \\ 0, & x \leq 3 \end{cases}$$



## Associação *fuzzy*

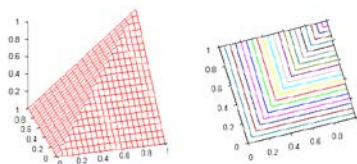
Para a identificação dos itens frequentes, as operações de conjuntos são necessárias e na associação *fuzzy* não é diferente, mas agora em suas versões *fuzzy*:

$$|A| = \sum_x \mu_A(x), \quad \mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x))$$

$T$  é uma norma triangular, isto é, um mapeamento (comutativo, associativo, monotônico, 1 como elemento neutro e 0 como absorvente) que generaliza as operações de interseção (em conjuntos) e de conjunção (em lógica). Algumas das mais comuns são:

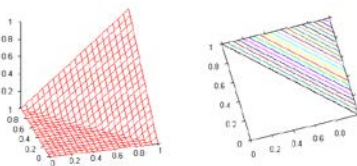
$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$$

mínimo



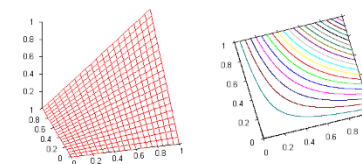
$$\mu_{A \cap B}(x) = \max(0, \mu_A(x) + \mu_B(x) - 1)$$

Łukasiewicz



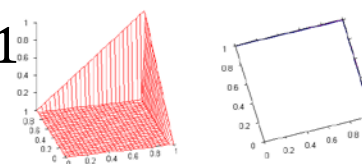
$$\mu_{A \cap B}(x) = \mu_A(x) \cdot \mu_B(x)$$

produto



$$\mu_{A \cap B}(x) = \begin{cases} \mu_A(x), & \text{se } \mu_B(x) = 1 \\ \mu_B(x), & \text{se } \mu_A(x) = 1 \\ 0, & \text{c.c.} \end{cases}$$

drástica



## Associação *fuzzy*

Exemplo:

Registro	Partição <i>fuzzy</i>						Conjunções ou interseções - mínimo			
	fria	amena	quente	curta	média	longa	fria-curta	amena-longa	amena-média	quente-média
1	1,00	0,00	0,00	1,00	0,00	0,00	1,00	0,00	0,00	0,00
2	0,00	0,60	0,93	0,00	0,50	1,00	0,00	0,60	0,50	0,50
3	0,47	1,00	0,20	0,00	0,75	0,75	0,00	0,75	0,75	0,20
4	0,00	0,00	1,00	0,00	0,50	1,00	0,00	0,00	0,00	0,50
5	0,00	0,00	1,00	0,20	0,90	0,40	0,00	0,00	0,00	0,90
6	0,00	1,00	0,67	0,00	0,00	1,00	0,00	1,00	0,00	0,00
7	0,13	1,00	0,53	0,00	0,75	0,75	0,00	0,75	0,75	0,53
8	0,53	1,00	0,13	0,00	1,00	0,50	0,00	0,50	1,00	0,13
9	0,67	1,00	0,00	0,60	0,70	0,20	0,60	0,20	0,70	0,00
10	0,73	0,80	0,00	1,00	0,25	0,00	0,73	0,00	0,25	0,00
Somas	3,53	6,40	4,47	2,80	5,35	5,60	2,33	3,80	3,95	2,77

Regras		Métricas tradicionais			Métricas <i>fuzzy</i> - mínimo		
Antecedente	Consequente	Cobertura	Suporte	Confiança	Cobertura	Suporte	Confiança
temp é fria	duração é curta	0,10	0,10	1,00	3,53	2,33	0,66
temp é amena	duração é longa	0,50	0,00	0,00	6,40	3,80	0,59
temp é amena	duração é média	0,50	0,40	0,80	6,40	3,95	0,62
temp é quente	duração é média	0,40	0,30	0,75	4,47	2,77	0,62
duração é média	temp é quente	0,70	0,30	0,43	5,35	2,77	0,52