



Universidade
Federal de Juiz
de Fora

Modelos de Mistura Gaussiana (GMM)

Fundamentos e Aplicações

Claudia Fonseca

Paulo Sérgio de Castro Nascimento

Patrícia Oliveira Silva

Juiz de Fora, 8 de outubro de 2021



Universidade
Federal de Juiz
de Fora

Conceitos iniciais de *clusterização*

Considerações iniciais

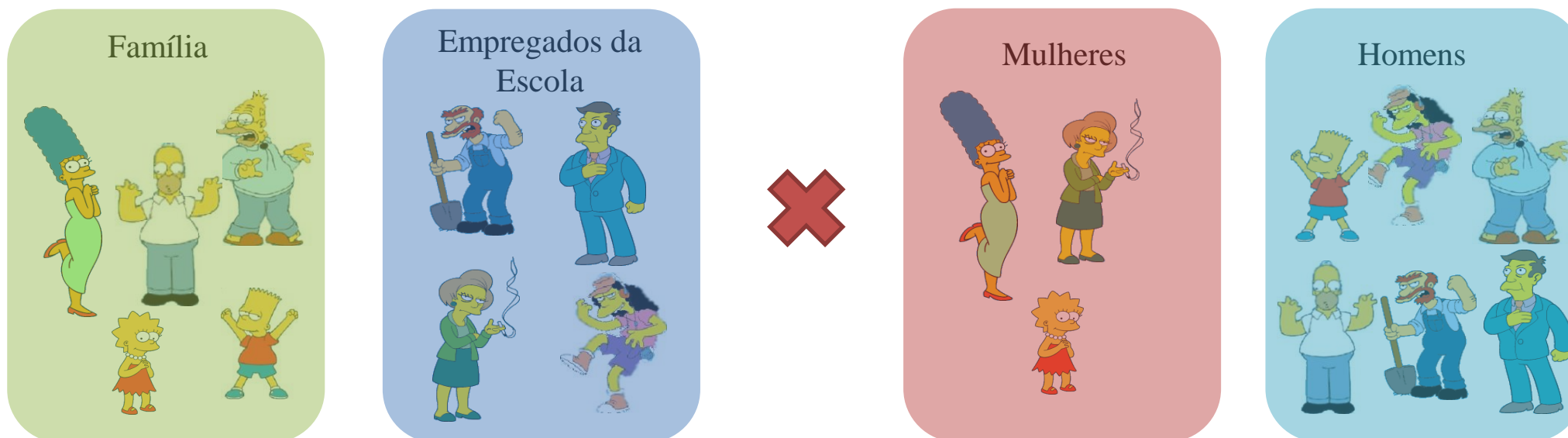
- Clusterização ou agrupamento é a tarefa de encontrar grupos onde os elementos sejam similares entre si.
- No entanto, existem situações nas quais não sabemos a maneira apropriada de **agrupar** uma coleção de objetos de acordo com suas “similaridades”;
- Frequentemente não sabemos se existe algum **agrupamento natural** dos objetos segundo um conjunto de características que descrevem esses objetos.

Agrupamentos

- O que é um agrupamento natural entre os seguintes objetos?

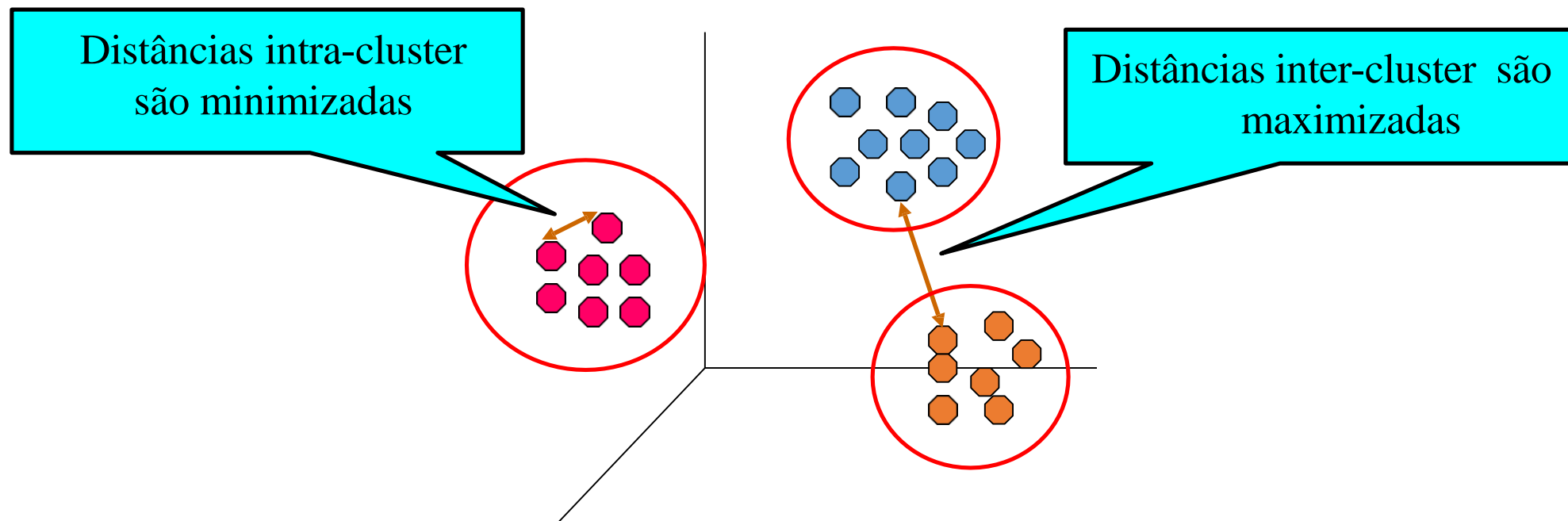


- Grupo é um conceito subjetivo:



Uma definição para agrupamento de dados

“Finding groups of objects such that the objects in a group are similar (or related) to one another and different from (or unrelated to) the objects in other groups.” (Tan et al., 2006)



Algoritmos de *clusterização*

- Algoritmos de *clustering* induzem *clusters*;
- Os *clusters* a serem induzidos dependem de uma série de fatores, além dos dados propriamente ditos, por exemplo:
 - Medidas de similaridade ou dissimilaridade;
 - Índices de avaliação;
 - Parâmetros definidos pelo usuário, etc.
- No ***Machine Learning*** (Aprendizado de Máquina):
 - Projetista define o que o computador pode aprender;
 - Existem diversos algoritmos de *clusterização*.

Técnicas de *Machine Learning*

Aprendizagem Supervisionada

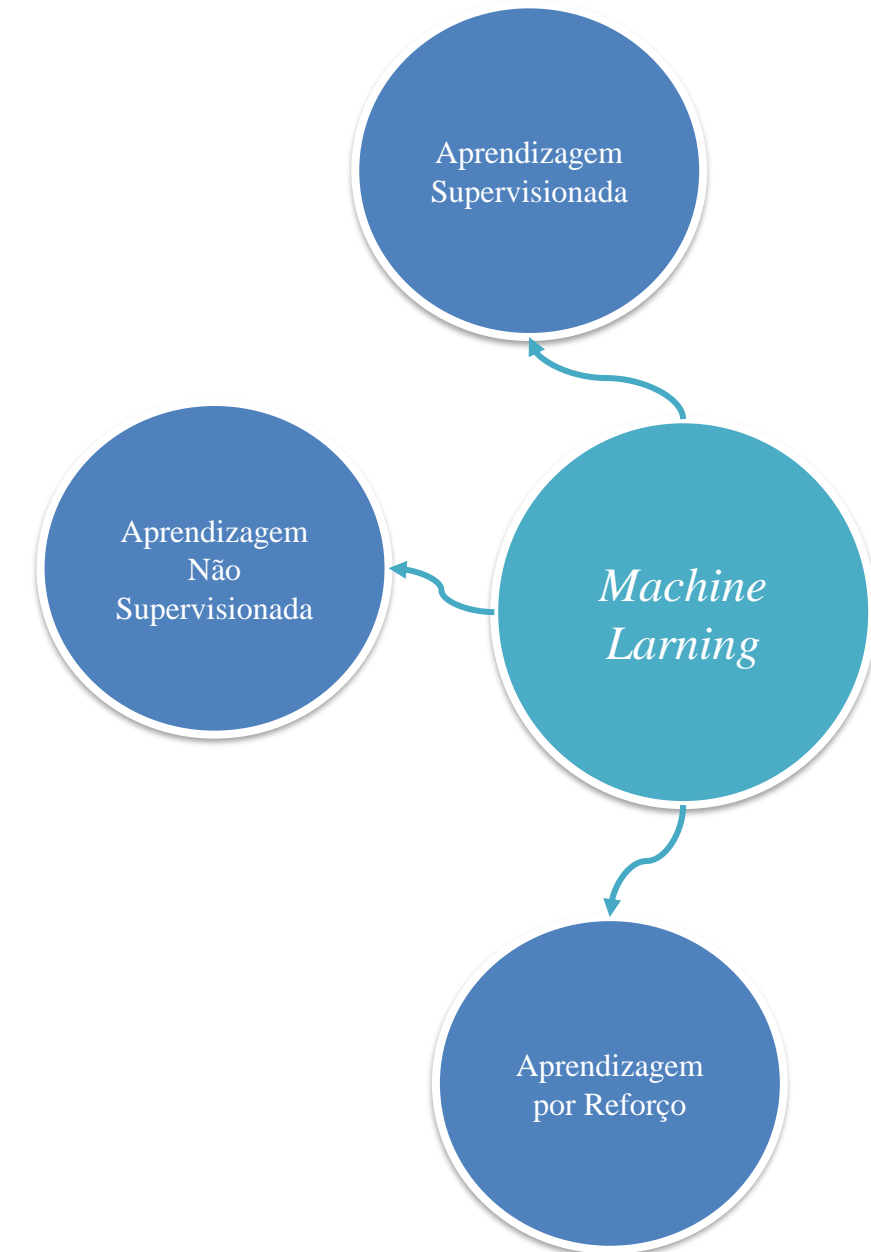
- Ocorre quando o modelo aprende a partir de resultados pré-definidos.
- O modelo possui uma referência daquilo que está certo e daquilo que está errado.

Aprendizagem Não Supervisionada

- Não existem resultados pré-definidos para o modelo utilizar como referência para aprender.

Aprendizagem Por Reforço

- A máquina tenta aprender qual é a melhor ação a ser tomada, dependendo das circunstâncias na qual essa ação será executada.



Agrupamento x Classificação

Agrupamento ou *Clusterização*

NÃO SUPERVISONADO

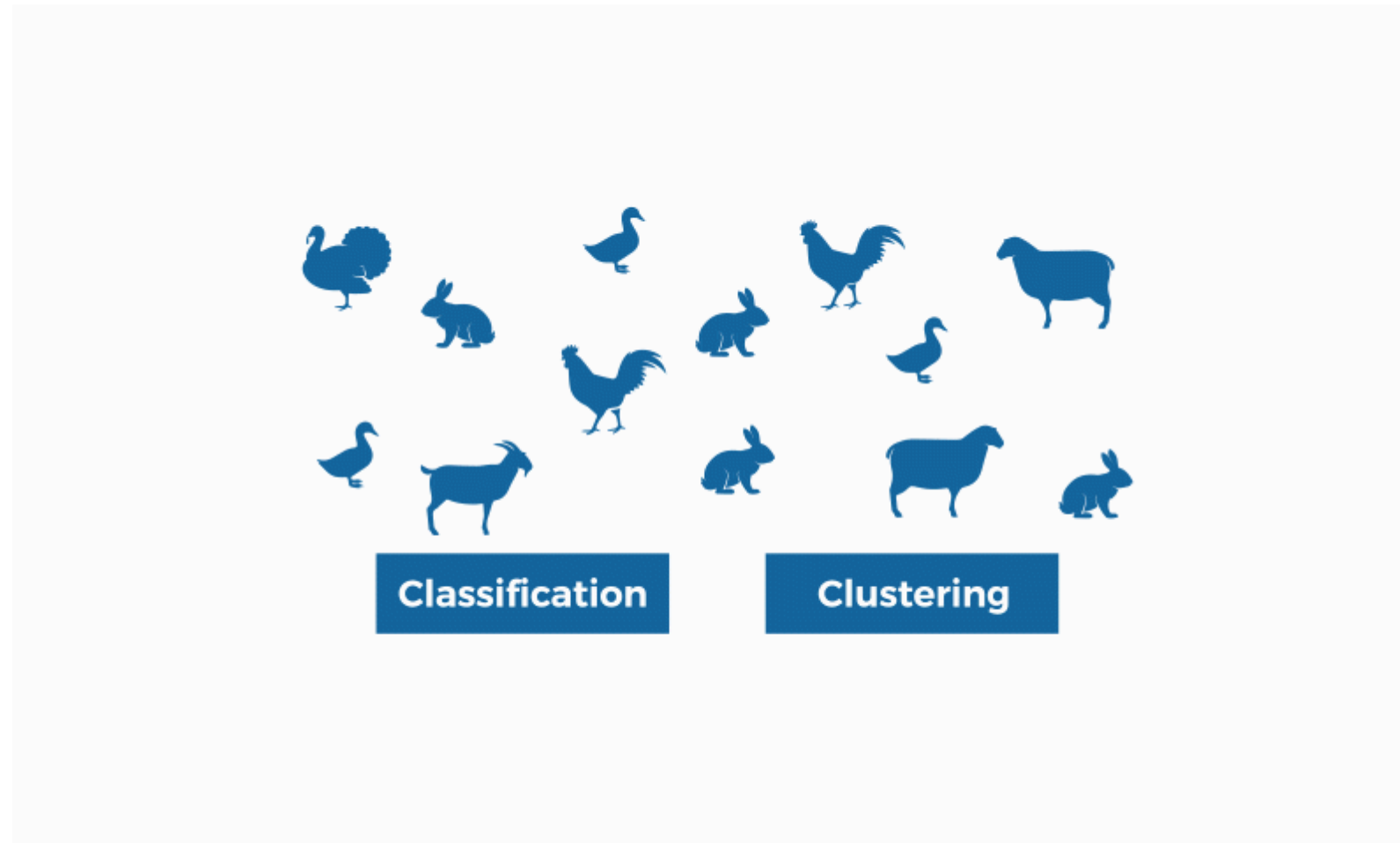
- Encontrar os rótulos das categorias (grupos ou *clusters*) e possivelmente o número de categorias diretamente a partir dos dados.
- É a indução de grupos a partir da base de dados e após agrupados esses grupos serão cuidadosamente estudados

Classificação

SUPERVISONADO

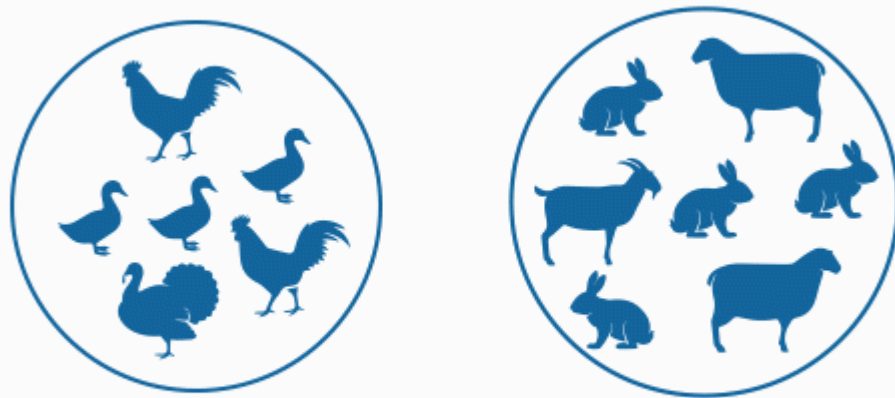
- Aprender um método para predizer as categorias (classes) de padrões não vistos a partir de exemplos pré-rotulados (classificados).

Agrupamento x Classificação (exemplos)



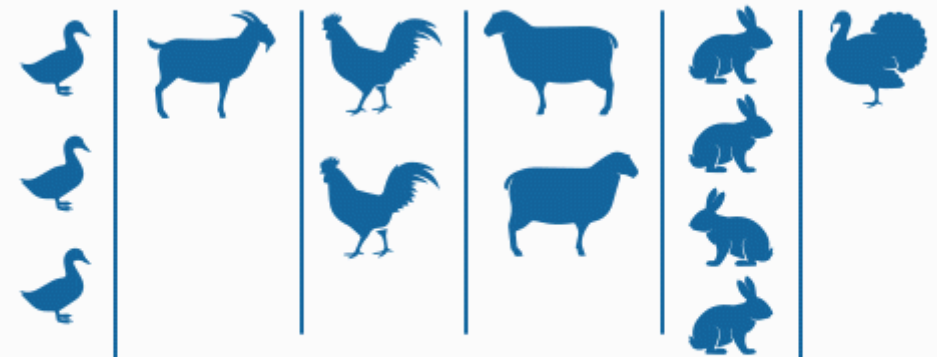
Agrupamento x Classificação (exemplos)

Agrupamento



Clustering

Classificação



Classification



Universidade
Federal de Juiz
de Fora

Modelo de misturas gaussianas

GMM (*Gaussian Mixture Model*)

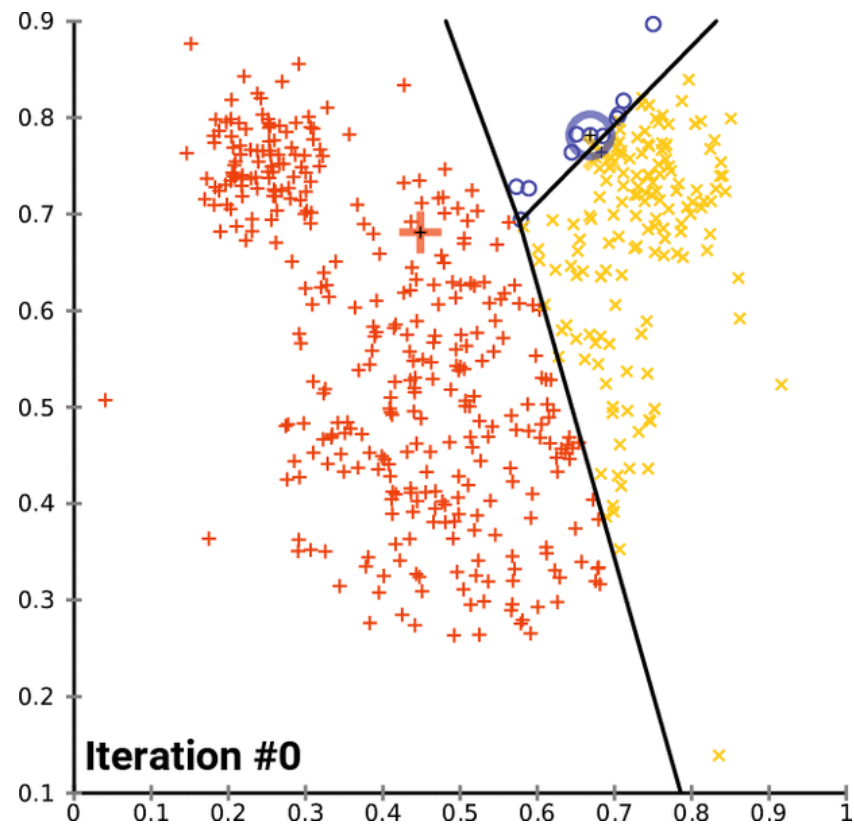
- Método de aprendizagem não supervisionado.
- Ocorre a partir de dados não rotulados e sem conhecimento prévio das categorias presentes no conjunto de dados.
- Busca compreender automaticamente a organização dos padrões existentes nos dados.
- Para finalmente obter conclusões úteis a respeito deles.

GMM x K-means

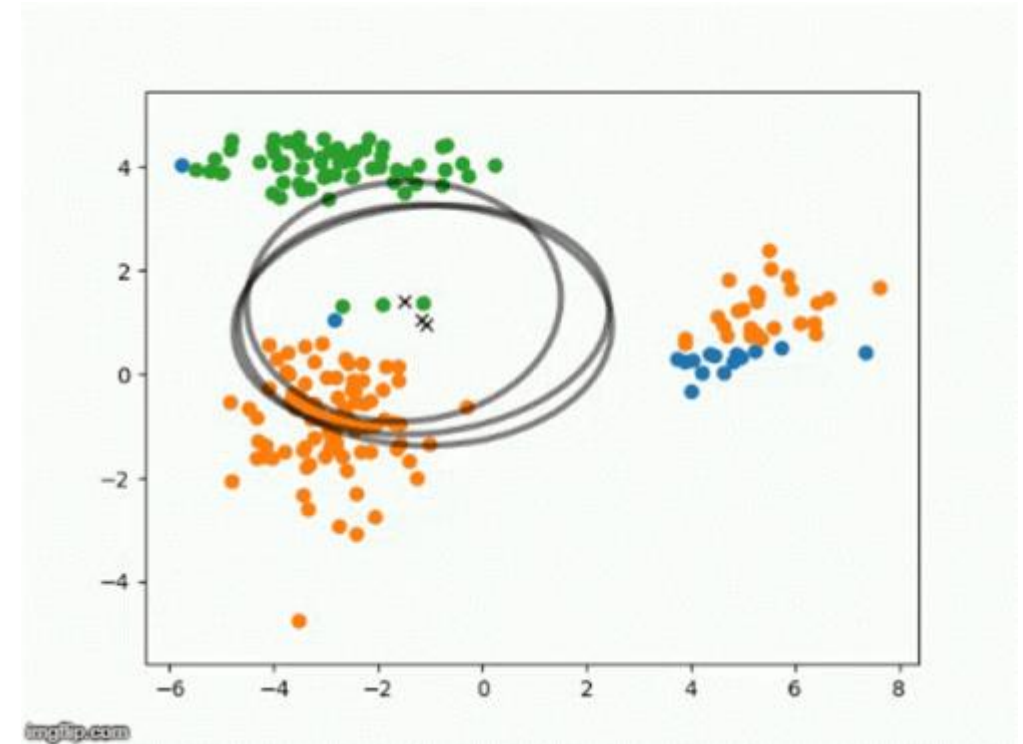
- Ambos são modelos de agrupamento.
- No entanto, muitos cientistas de dados, tendem a escolher um algoritmo K-Means.
- Porém, o GMM pode se provar superior em certos problemas de agrupamento
- Os dois modelos oferecem um desempenho diferente em termos de velocidade e robustez.
- Por último, é possível usar K-Means como um inicializador para o GMM, o que tende a aumentar o desempenho do modelo de agrupamento.

Como o K-means e o GMM trabalham?

K-means



GMM



GMM (Gaussian Mixture Model)

Um GMM é representado pela *p.d.f*:

Diagram illustrating the components of a Gaussian Mixture Model (GMM) and its probability density function (p.d.f):

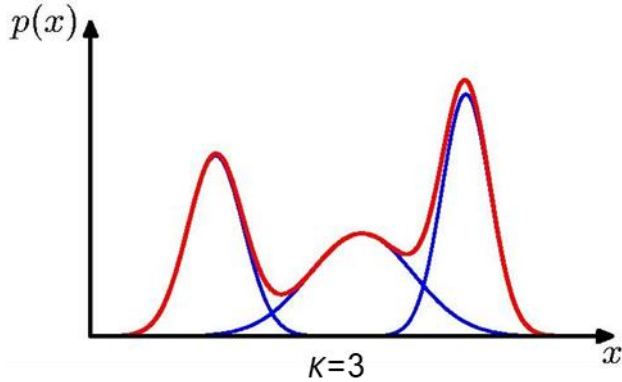
The p.d.f is defined as:

$$p(x/\theta) = \sum_{k=1}^K \pi_k p(x/\mu_k, \Sigma_k)$$

where $\sum_{k=1}^K \pi_k = 1$

The components are:

- Número de componentes (K)
- Coefficientes da mistura (π_k)
- Componentes da mistura ($p(x/\mu_k, \Sigma_k)$)
- Centro da i-ésima Gaussiana (vetor da mesma dimensão de \mathbf{x}) (μ_k)
- Matriz de covariância da i-ésima Gaussiana (Σ_k)

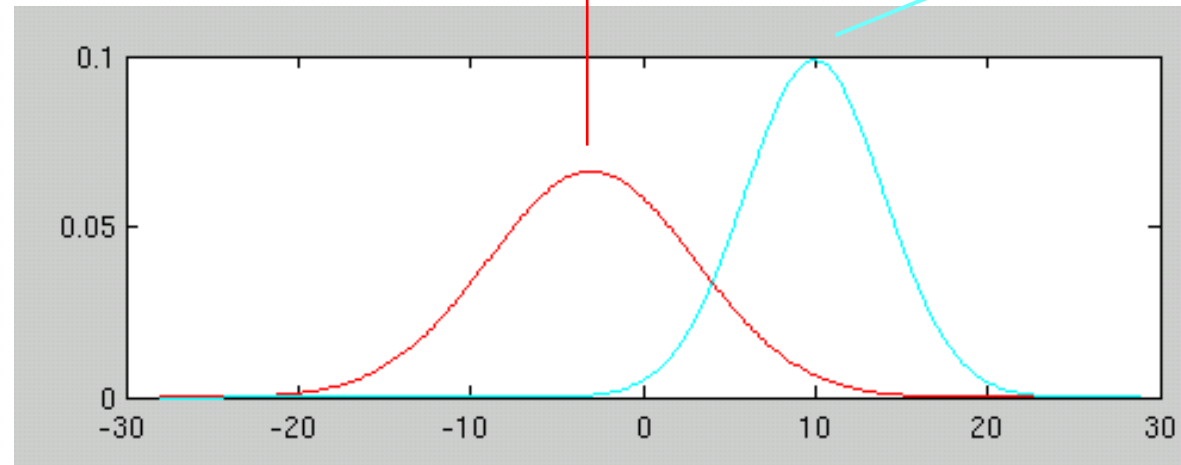


$$p(x/\mu_k, \Sigma_k) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\}$$

GMM (*Gaussian Mixture Model*)

Expandindo para um exemplo com 2 gaussianas

$$p(x/\theta) = \underbrace{\pi_1 p(x/\mu_1, \Sigma_1)}_{\text{red}} + \underbrace{\pi_2 p(x/\mu_2, \Sigma_2)}_{\text{cyan}}$$



$$\theta = \pi_1, \pi_2, \mu_1, \mu_2, \Sigma_1, \Sigma_2$$

EM para Mistura de Gaussianas

- O algoritmo EM (*Expectation Maximization*) parte do princípio do método da máxima verossimilhança.
- O EM é utilizado para encontrar estimadores de máxima verossimilhança (EMV's) de parâmetros de modelos estatísticos nos casos em que as equações não podem ser resolvidas analiticamente.
- Tipicamente, isso ocorre porque tais modelos envolvem variáveis latentes.
- Além disso, os parâmetros dos dados observados são desconhecidos
- Modelo mais utilizado: **Mistura de Gaussianas**

Formulação Matemática para o algoritmo EM

- Dado um modelo estatístico que gera um conjunto X de observações, um conjunto de variáveis latentes Z e um vetor de parâmetros Θ , temos que a função verossimilhança é dada:

$$L(\theta; X, Z) = p(X, Z | \theta)$$

- O Estimador de Máxima Verossimilhança do vetor de parâmetros Θ é determinado pela maximização da verossimilhança marginal dos dados observados.
- A marginalização é feita através da integração da variável latente Z .

$$L(\theta; X) = p(X | \theta) = \int p(X, Z | \theta) dZ$$

Algoritmo EM

- O algoritmo EM busca encontrar o Estimador da Máxima Verossimilhança iterativamente aplicando os dois passos:

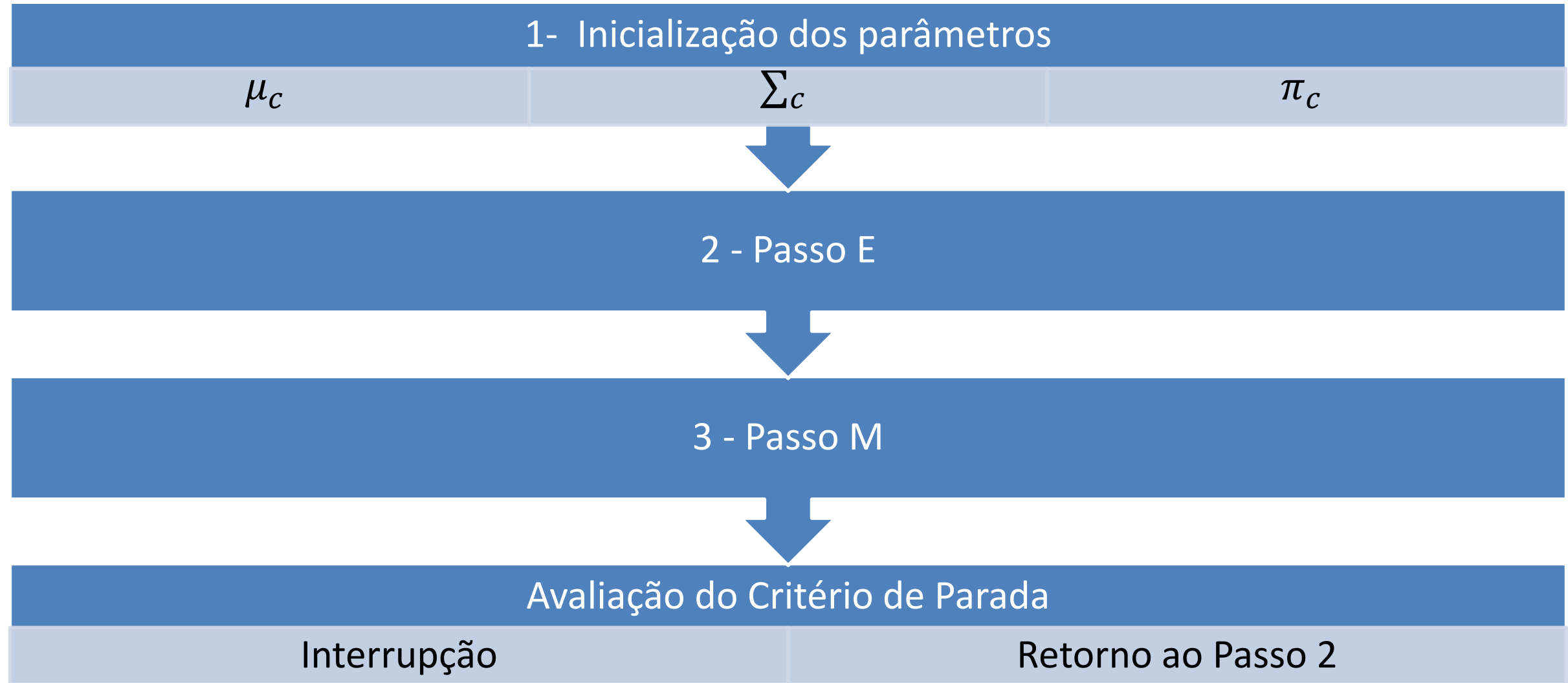
1. **Expectation** (Etapa E): calcula o valor esperado da log- verossimilhança com relação a distribuição condicional de Z dado X, utilizando a estimativa atual dos parâmetros Θ da iteração 't'.

$$Q(\theta|\theta_t) = E_{Z|X,\theta_t}[\log L(\theta; X, Z)]$$

2. **Maximization** (Etapa M): Encontrar os parâmetros Θ que maximizam essa quantidade (Q).

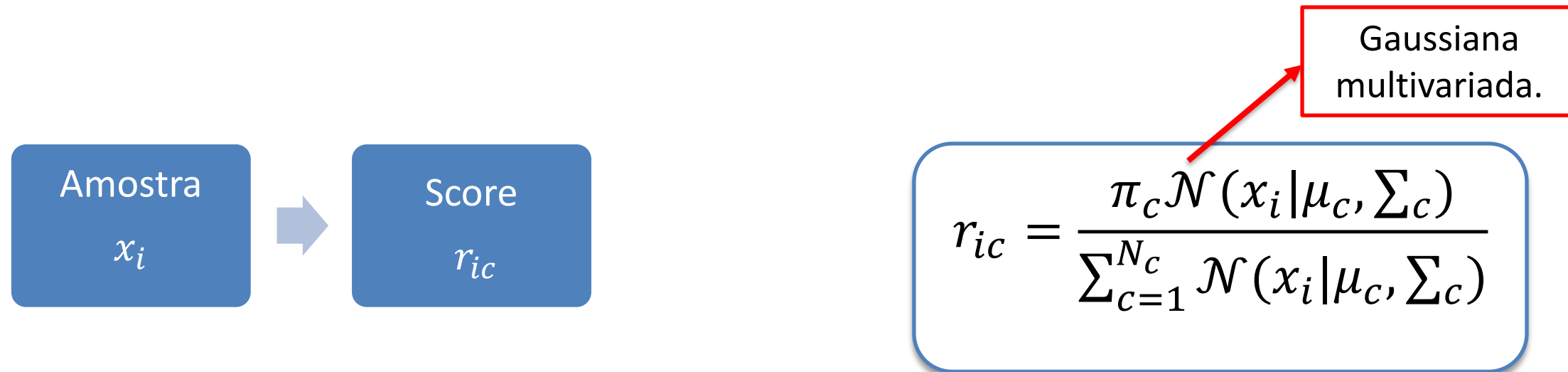
$$\theta_{t+1} = \arg \max_{\theta} Q(\theta, \theta_t)$$

Algoritmo EM



Algoritmo EM

Passo da expectativa



Algoritmo EM

Passo Maximização

Dados

r_{ic}



Atualizar
parâmetros

$$m_c = \sum_{c=1}^{N_c} r_{ic}$$

$$\mu_c = \frac{1}{m_c} \sum_{i=1}^{N_i} r_{ic} x_i$$

$$\Sigma_c = \frac{1}{m_c} \sum_{i=1}^{N_i} r_{ic} (x_i - \mu_c)^T (x_i - \mu_c)$$

$$\pi_c = \frac{m_c}{N_i}$$

Algoritmo EM

Avaliação do Critério de Parada

- Função log - verossimilhança

$$\ln p(X|\mu, \Sigma, \pi) = \sum_{i=1}^{N_i} \ln \left\{ \sum_{c=1}^{N_c} \pi_c \mathcal{N}(x_i | \mu_c, \Sigma_c) \right\}$$

Medida de confiança de que os dados são gerados pelos parâmetros estimados.

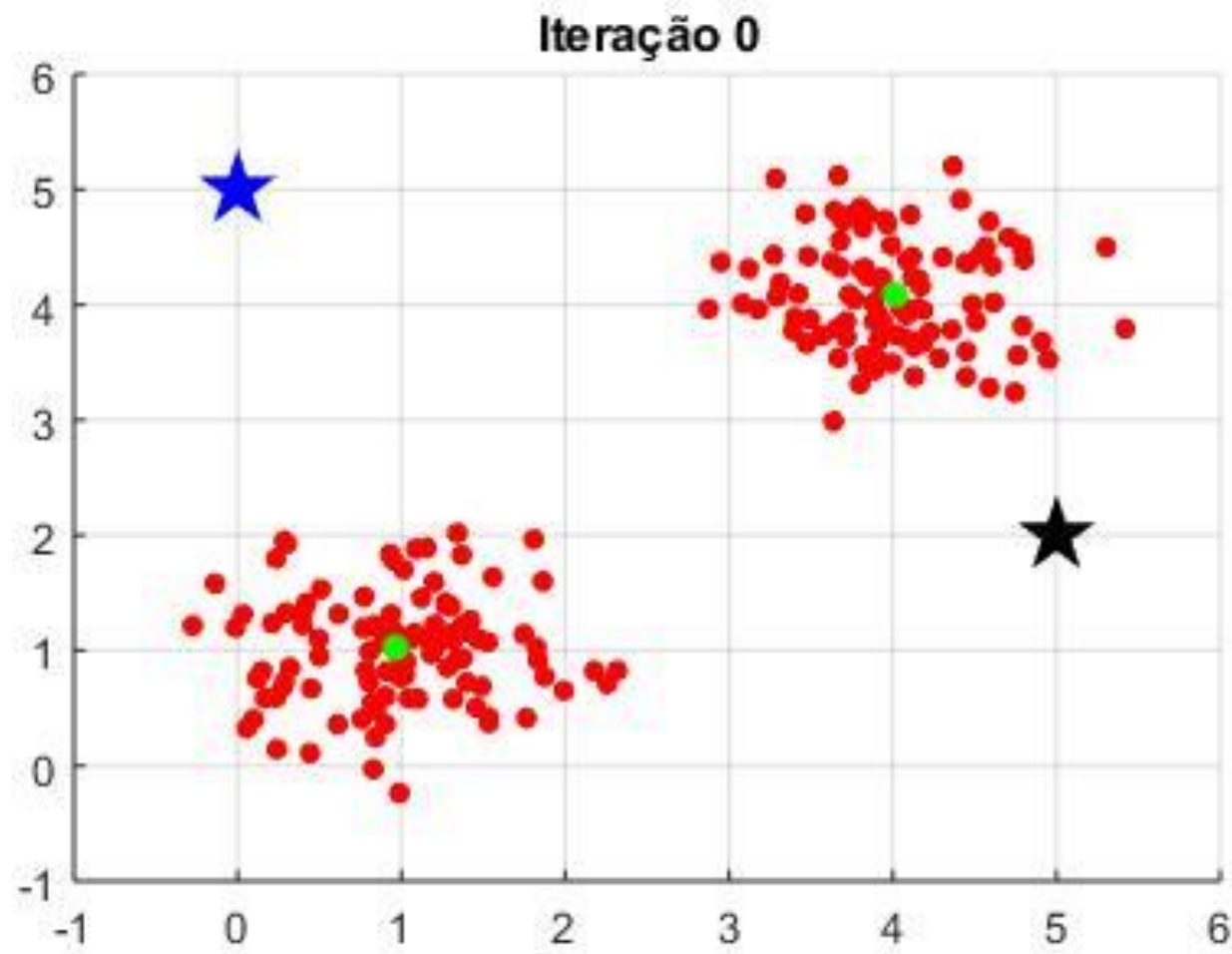
Deve ser computado ao final de cada iteração.

Algoritmo EM

- Número de pontos por cluster $\rightarrow 10$
- Número de clusters $\rightarrow 2$
- Iterações $\rightarrow 6$

Dados originais					
μ_1	μ_2	Σ_1		Σ_2	
1	4	1/4	0	1/4	0
1	4	0	1/4	0	1/4

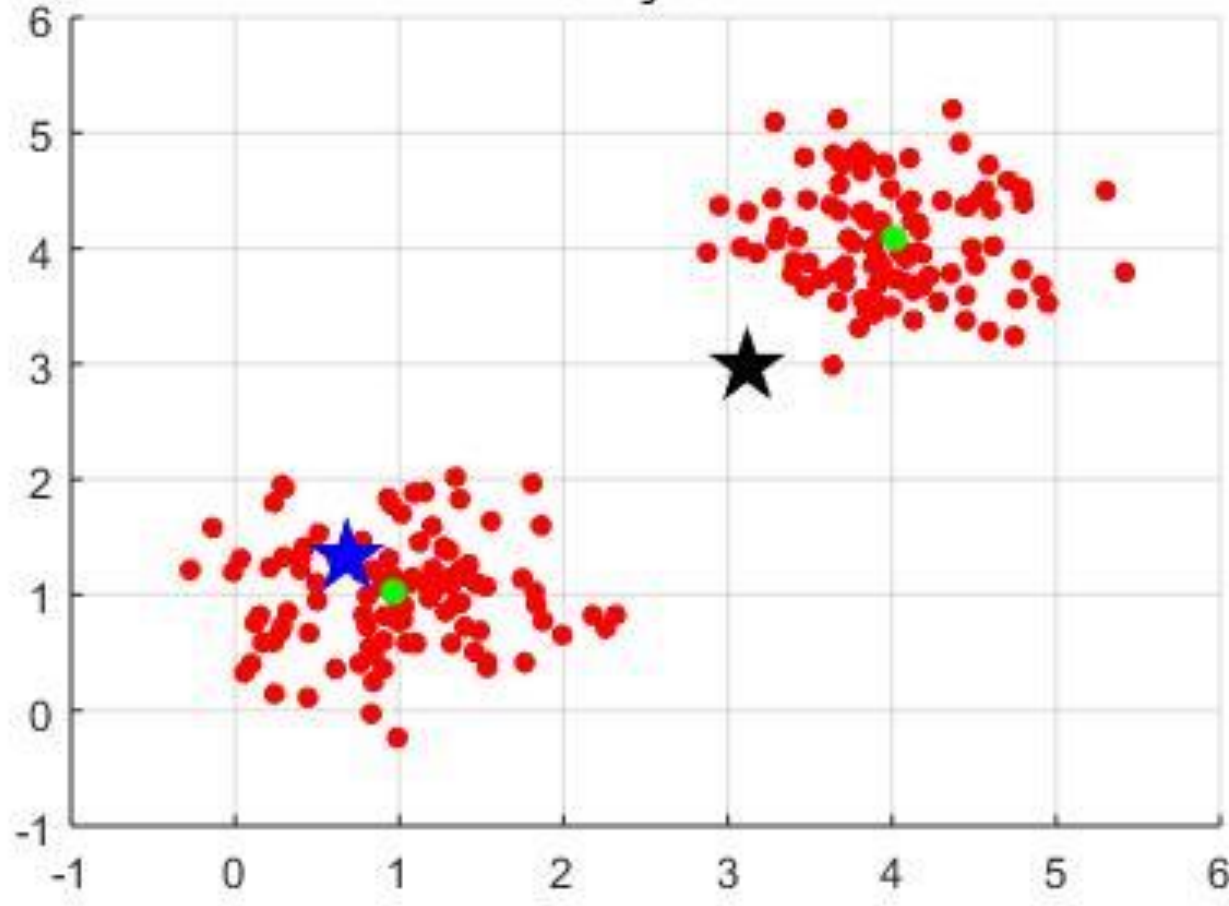
Algoritmo EM



Dados iniciais					
Σ_1		Σ_2		π_1	π_2
1/4	0	1/4	0	0,5	0,5
0	1/4	0	1/4		
μ_1		μ_2			
0		5			
5		2			

Algoritmo EM

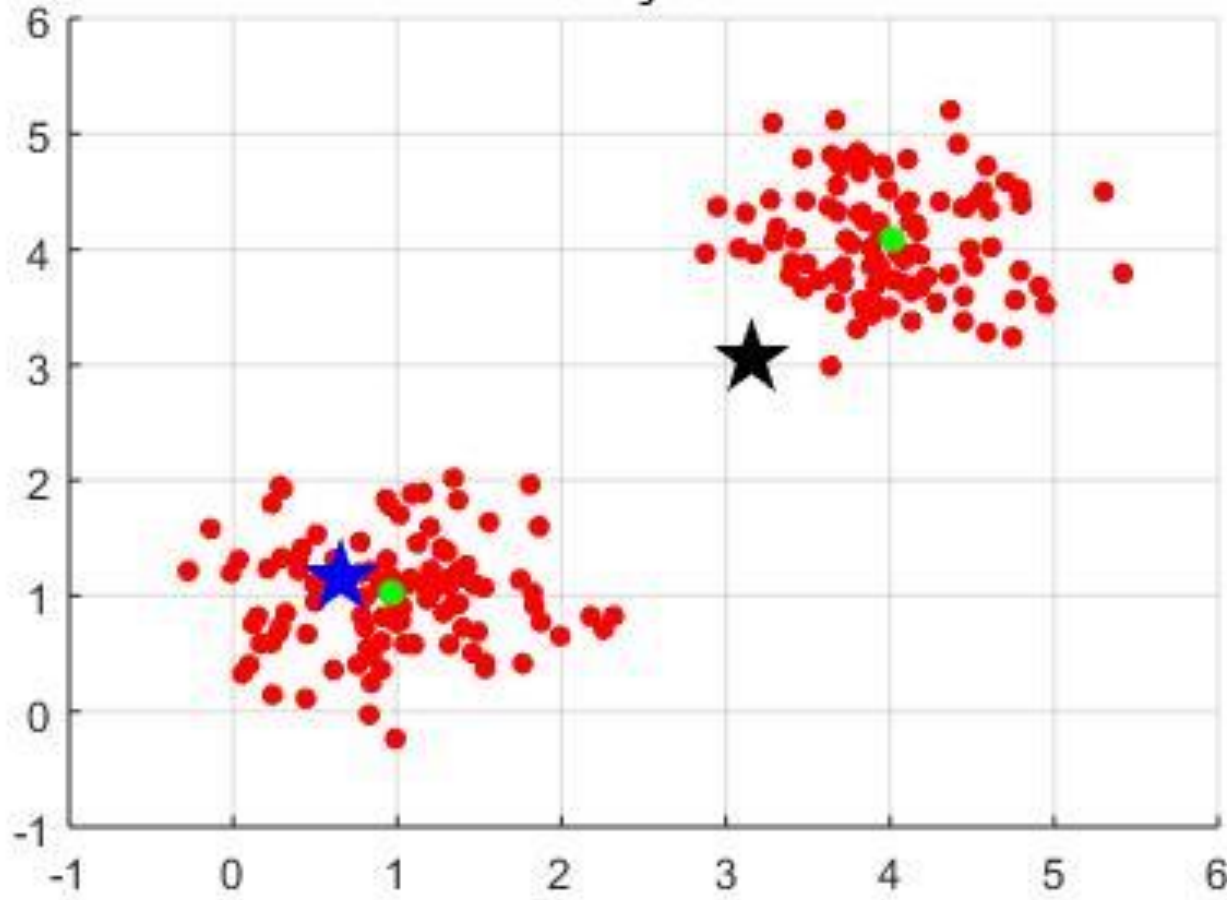
Iteração 1



Iteração 1					
Σ_1		Σ_2		π_1	π_2
0,4155	0,4209	1,8269	1,9600	0,2571	0,7429
0,4209	0,6952	1,9600	2,5388		
μ_1		μ_2			
0,6799		3,1183			
1,328		2,975			

Algoritmo EM

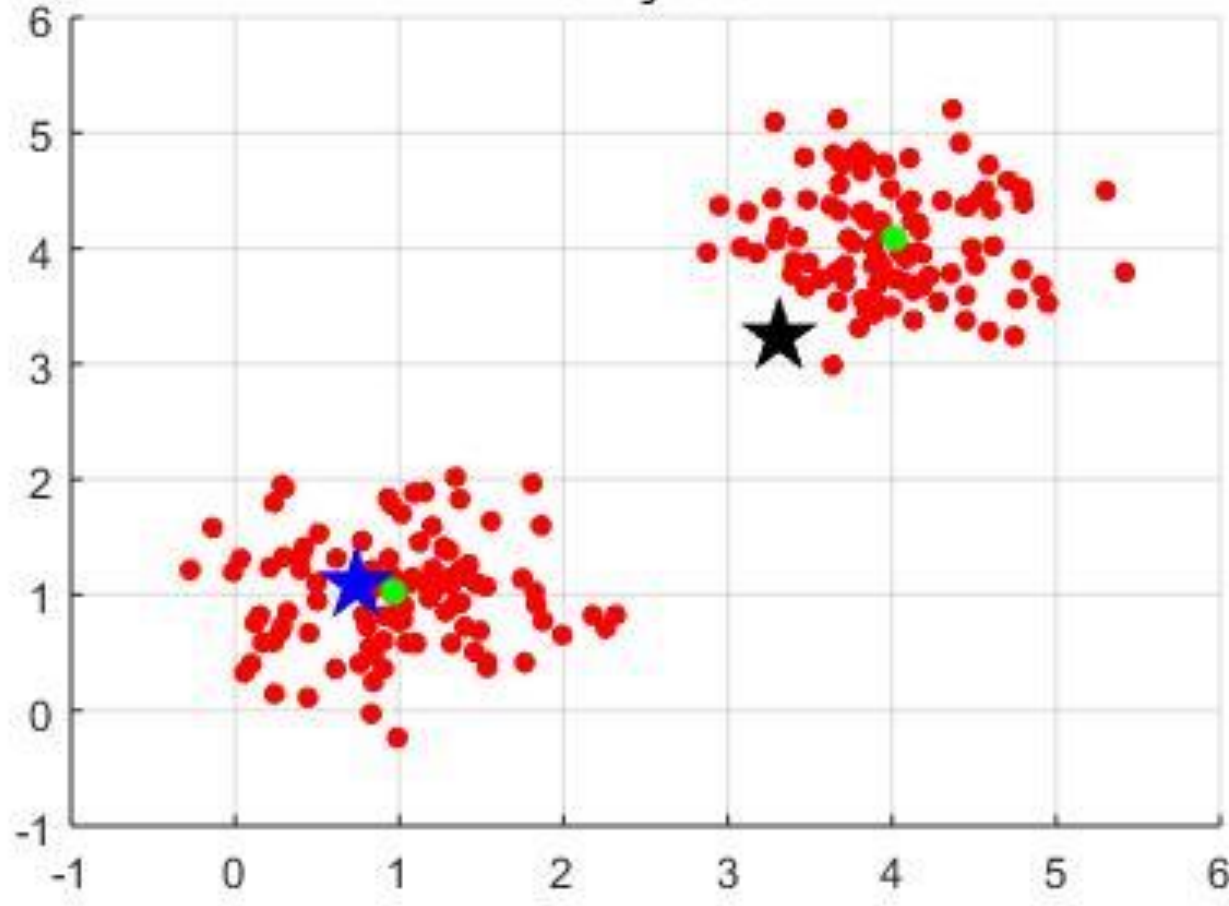
Iteração 2



Iteração 2					
Σ_1		Σ_2		π_1	π_2
0,2070	0,0615	1,8020	1,8898	0,2661	0,7339
0,0615	0,2271	1,8898	2,4767		
μ_1		μ_2			
0,6555		3,1570			
1,1573		3,0571			

Algoritmo EM

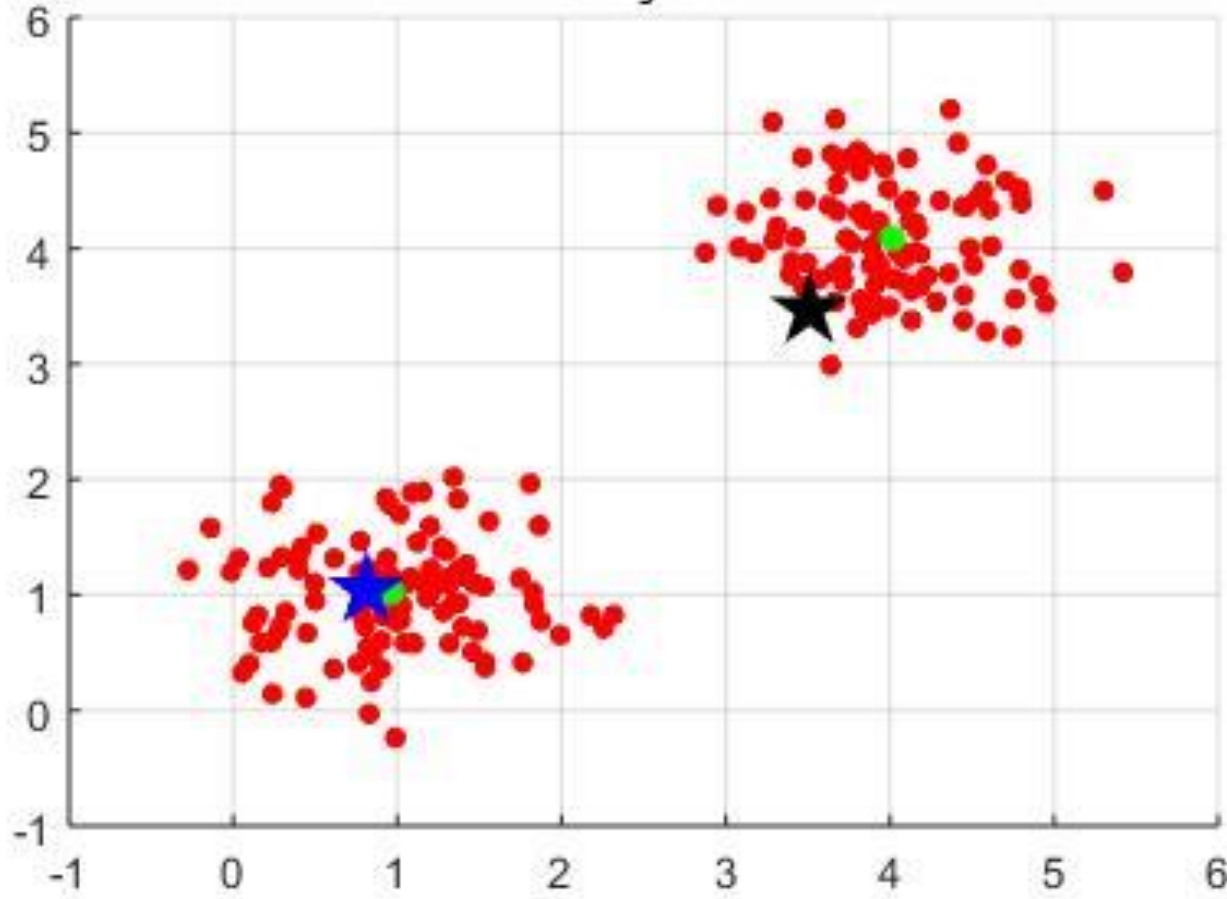
Iteração 3



Iteração 3					
Σ_1		Σ_2		π_1	π_2
0,2128	0,0224	1,6021	1,6508	0,3191	0,6809
0,0224	0,2043	1,6508	2,2294		
		μ_1	μ_2		
		0,7379	3,3131		
		1,0910	3,2361		

Algoritmo EM

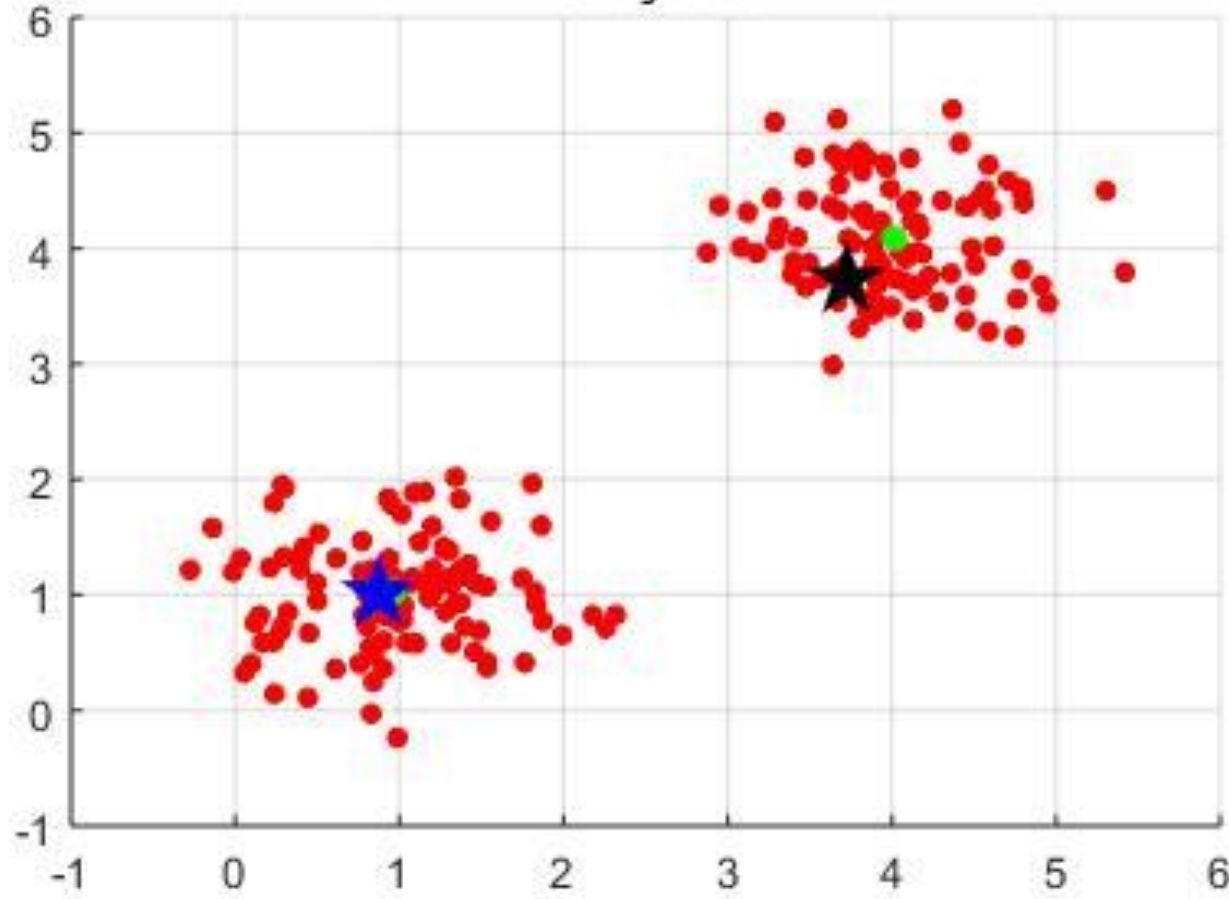
Iteração 4



Iteração 4					
Σ_1		Σ_2		π_1	π_2
0,2332	0,0070	1,2867	1,2743	0,3772	0,6228
0,0070	0,2084	1,2743	1,8088		
		μ_1	μ_2		
		0,8109	3,5091		
		1,0433	3,4651		

Algoritmo EM

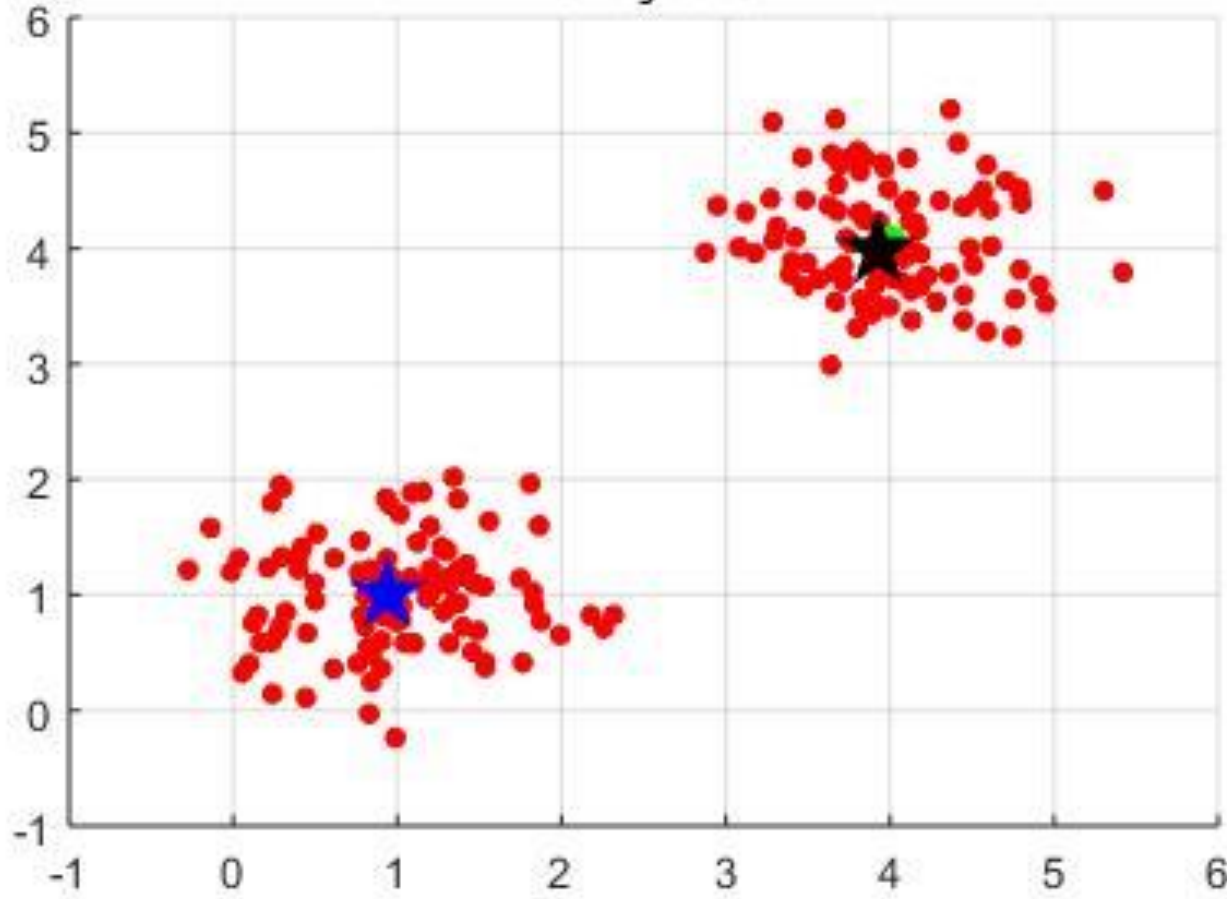
Iteração 5



Iteração 5					
Σ_1		Σ_2		π_1	π_2
0,2602	0	0,8690	0,7636	0,4327	0,5673
0	0,2186	0,7636	1,2013		
		μ_1	μ_2		
		0,8744	3,7246		
		1,0124	3,7255		

Algoritmo EM

Iteração 6



Iteração 6					
Σ_1		Σ_2		π_1	π_2
0,2954	- 0,0017	0,4336	0,2142	0,4802	0,5198
- 0,0017	0,2247	0,2142	0,5111		
		μ_1	μ_2		
		0,9368	3,9274		
		1,0050	3,9801		

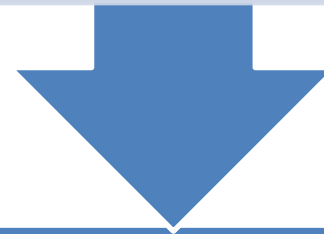
EM × k-means

EM

Informação mais rica

Probabilidades facilmente
convertidas em partição rígida

Custo computacional elevado



K-means

Caso particular do EM