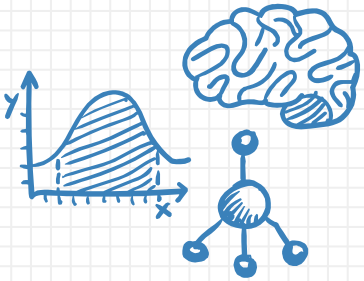
PUCRS



APRENDIZADO NÃO SUPERVISIONADO

AGRUPAMENTO DE DADOS (CLUSTERING)

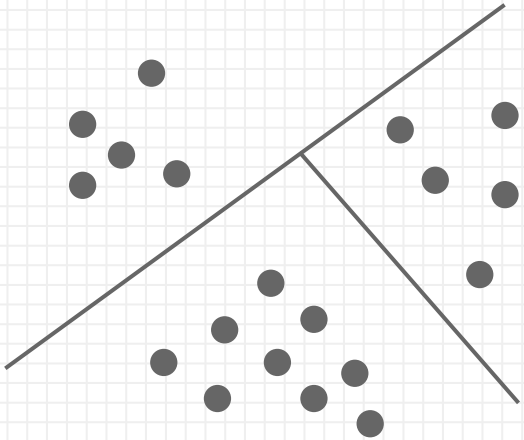
Parte II

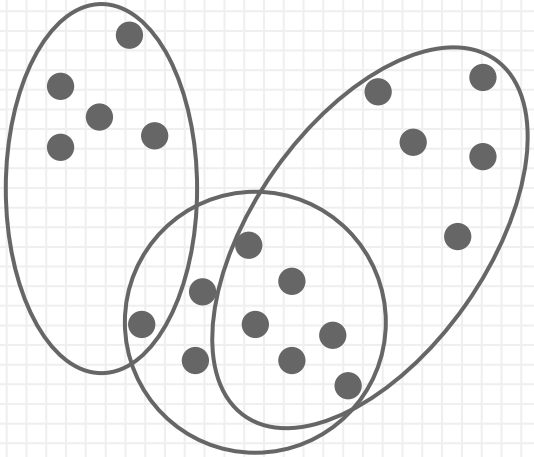
Agenda de Hoje

- ✘ Algoritmos de agrupamento
  - ✘ Estatística 101
  - ✘ Gaussian Mixture Models
  - ✘ DBSCAN

Recapitulando: tipos de agrupamento

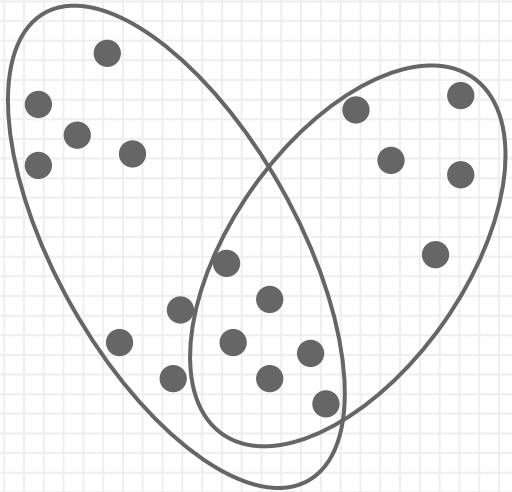
Particional sem sobreposição e **particional com sobreposição**





O que há de errado com o K-means?

- ✘ Cada exemplo é atribuído a apenas um cluster
- ✘ Se os clusters se sobrepõe, qual cluster está correto? Como podemos atribuir uma confiança ao nosso agrupamento?



O que há de errado com o K-means?

K-means

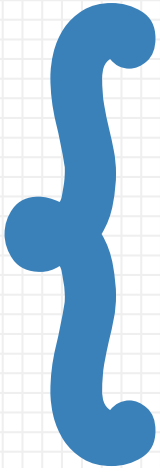
✘ Consegue encontrar apenas clusters esféricos

Encontrar uma forma de transformar os limites esféricos em elipses

✘ Possui uma atribuição não probabilística

Medir incerteza num cluster comparando a distância de cada instância para todos os centróides

Ai que entram GMMs :)



Estatística 101

Recapitulando alguns conceitos importantes

Média, Desvio Padrão, Variância...

- Dados valores [10, 3, 8, 5, 4], qual a média?

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \longrightarrow \mu = \frac{10+3+8+5+4}{5} = 6$$

Média, Desvio Padrão, Variância...

- Dados valores [10, 3, 8, 5, 4], qual a variância?

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad \mu = 6$$

$$\sigma^2 = \frac{1}{4}[(10-6)^2 + (3-6)^2 + (8-6)^2 + (5-6)^2 + (4-6)^2]$$

$$\sigma^2 = 8.5$$

Média, Desvio Padrão, Variância...

- Dados valores [10, 3, 8, 5, 4], qual o desvio padrão?

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2} \qquad \sigma^2 = 8.5$$

$$\sigma \approx 2.92$$

Mais um parênteses...

✘ Calculamos média 6, variância 8.5 e desvio padrão 2.92.

```
In [1]: import numpy as np

        a = [10, 3, 8, 5, 4]

        print('Média:', np.mean(a))
        print('Variância:', np.var(a))
        print('Desvio padrão:', round(np.std(a), 2))
```

```
Média: 6.0
Variância: 6.8
Desvio padrão: 2.61
```

Por quê obtivemos resultados diferentes?

Correção de Bessel!

Mais um parênteses...

Correção de Bessel

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2}$$

- ✘ Corrige o *bias* na estimativa da variância/desvio padrão de uma população
- ✘ Necessário quando a média da **população** não é conhecida e se está estimando tanto a média quando variância de uma **amostra**.

Fechando o parênteses...

- Para usar a correção de Bessel, adicionamos um parâmetro no cálculo

```
In [2]: import numpy as np

        a = [10, 3, 8, 5, 4]

        print('Média:', np.mean(a))
        print('Variância:', np.var(a, ddof=1))
        print('Desvio padrão:', round(np.std(a, ddof=1), 2))
```

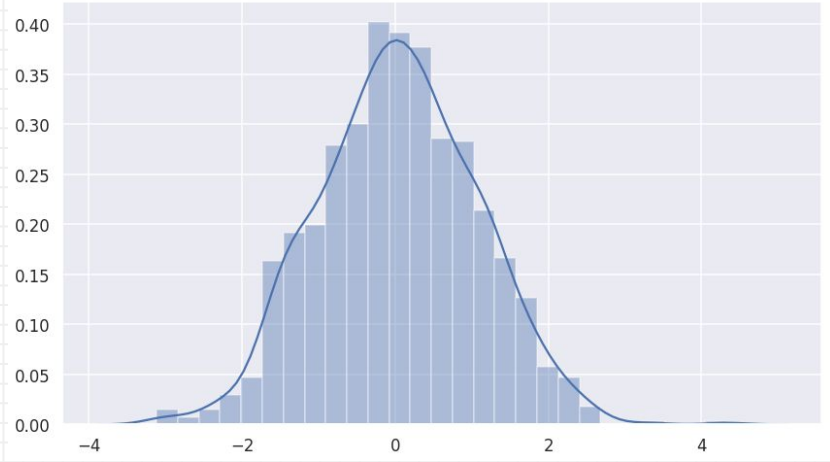
```
Média: 6.0
Variância: 8.5
Desvio padrão: 2.92
```

Devemos sempre usar a correção?

- Depende. Com a correção, o Erro Quadrático Médio aumenta e pode influenciar no resultado de alguns algoritmos (como o PCA, que veremos mais adiante).

O que é uma distribuição?

Histograma

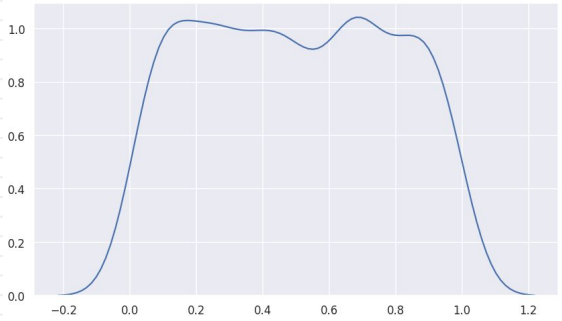


Histograma "com curva"

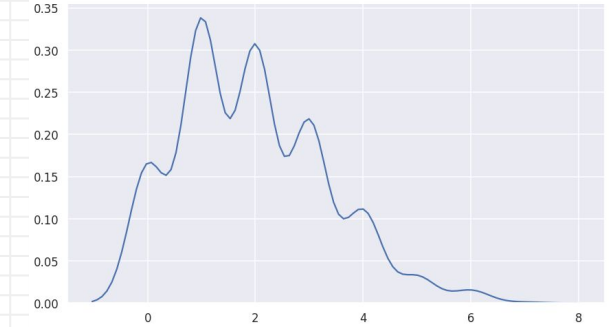
✘ Tanto o histograma quanto a curva são distribuições. Eles nos mostram como as probabilidades de observações são distribuídas.

O que é uma distribuição?

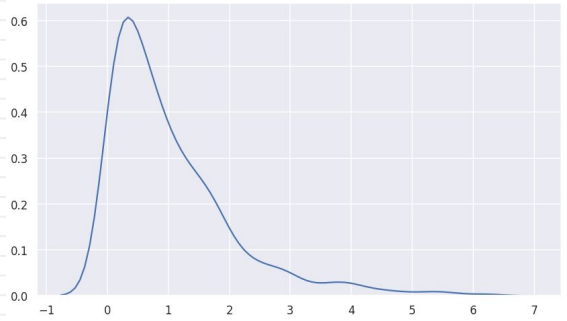
- Existem vários tipos de distribuições. Alguns exemplos.



Uniforme



Poisson

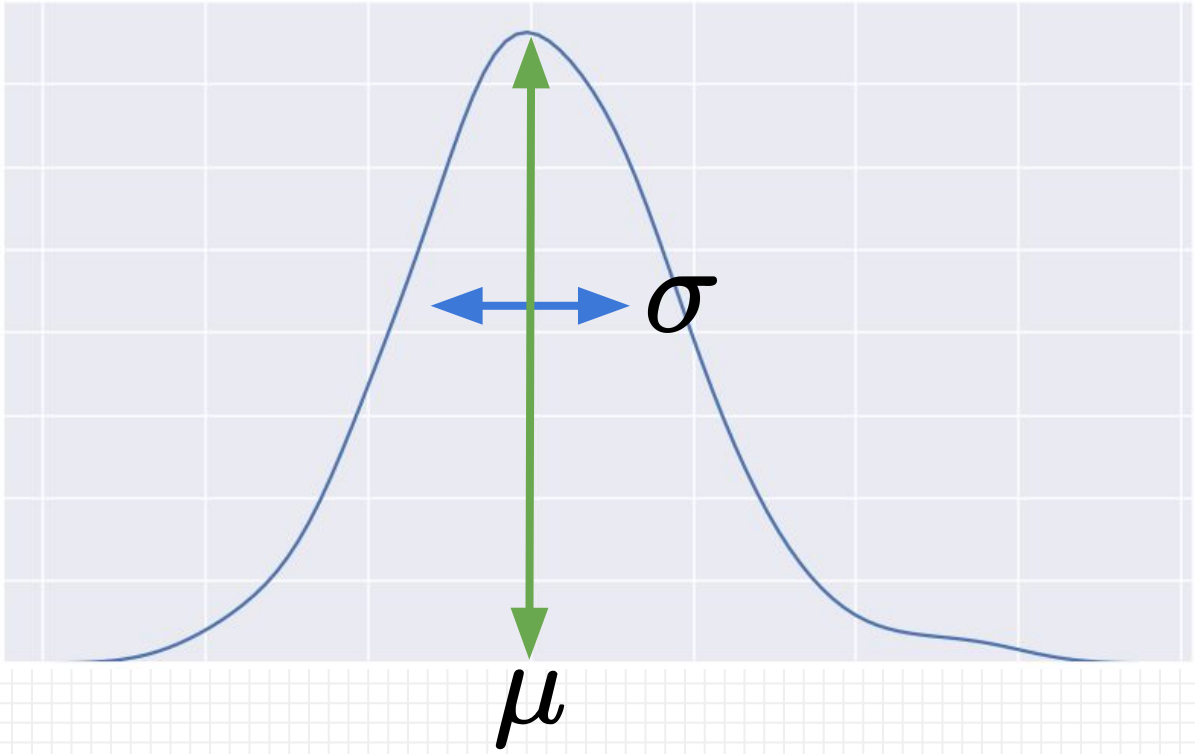


Exponencial

Carl Friedrich Gauss

- 1777 - 1855
- Matemático alemão, conhecido como “o maior matemático desde a antiguidade”
- Contribuições em campos como álgebra, astronomia, eletricidade, magnetismo e estatística
  - E também a distribuição normal

Distribuição Normal (ou Gaussiana)



Distribuição Normal (ou Gaussiana)

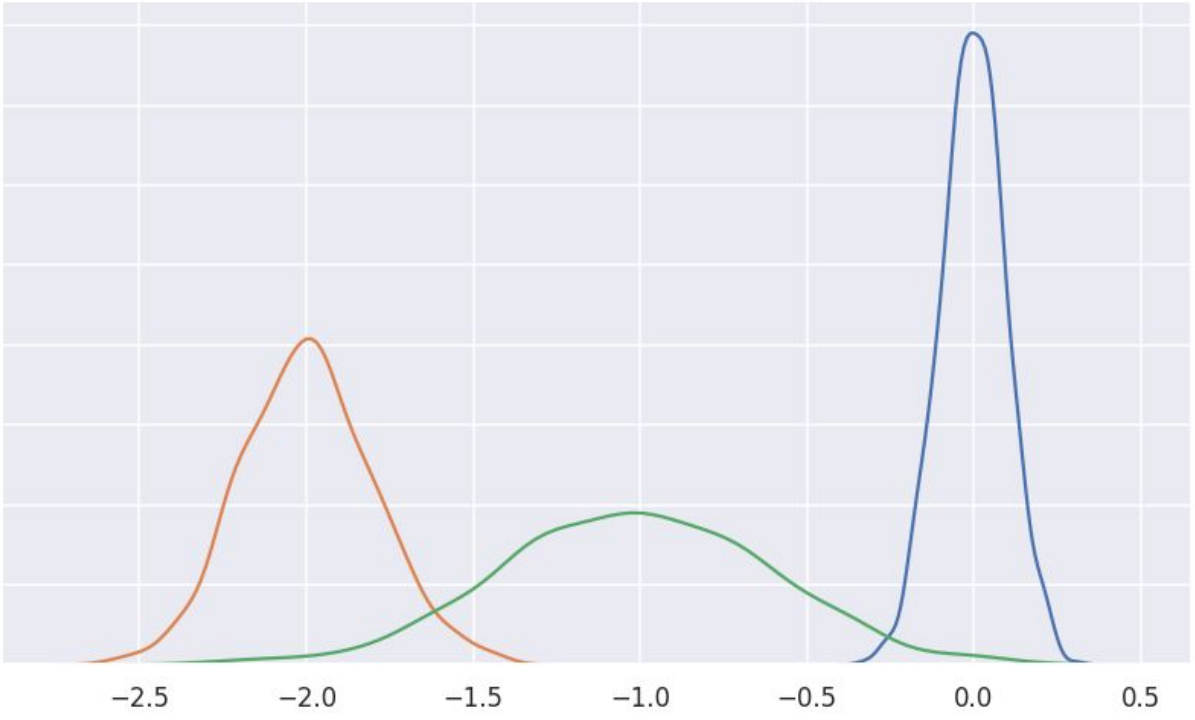
- Definida pela seguinte função

$$\mathcal{N}(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ média

σ desvio padrão

Exercício



Sabendo que:

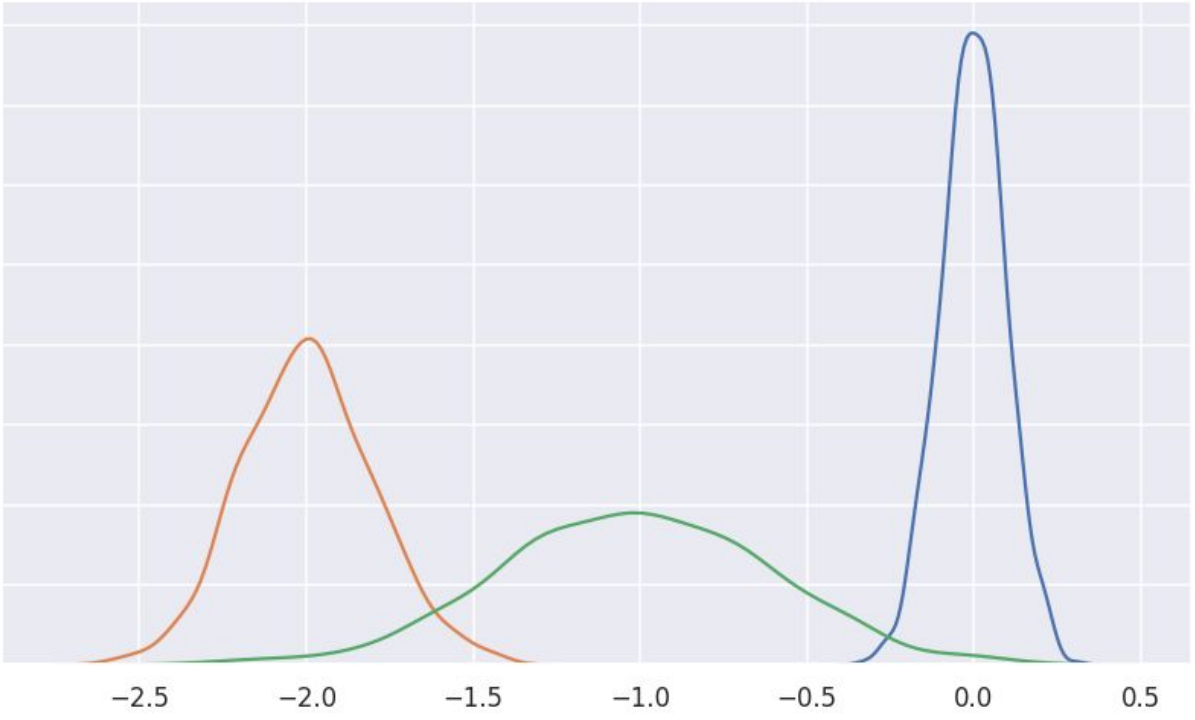
σ = 0.2

μ = -2

Qual é a Gaussiana?

Laranja

Exercício



Sabendo que:

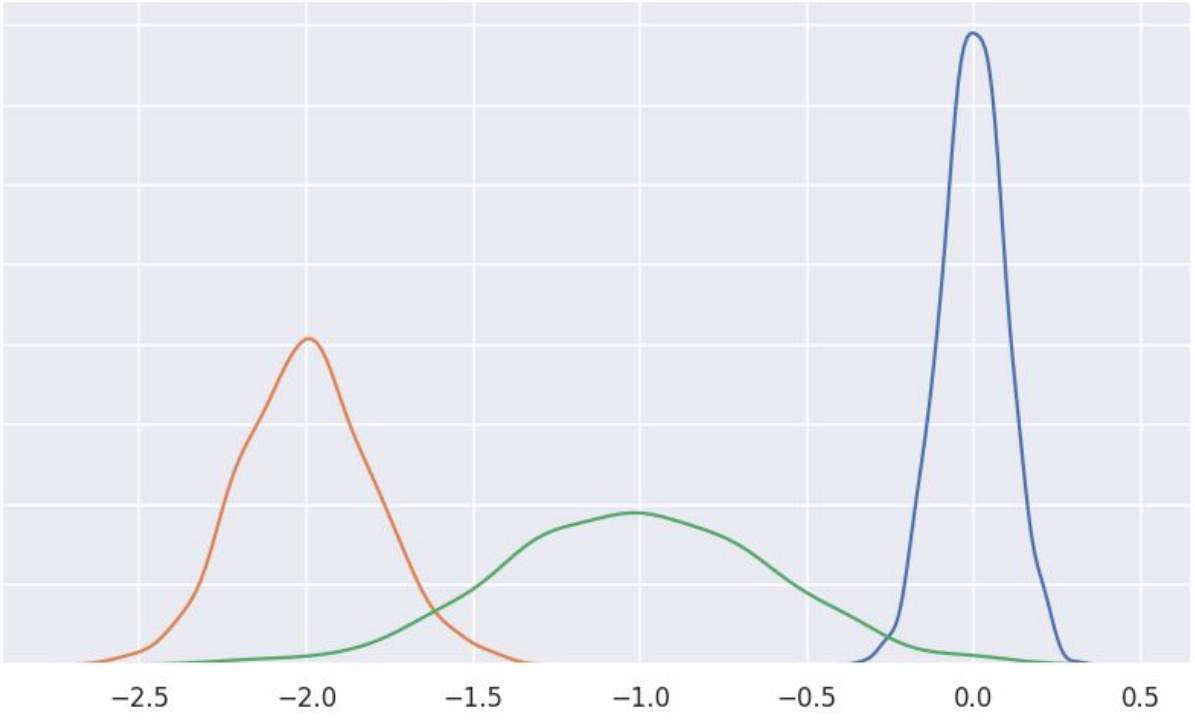
$\sigma = 0.1$

$\mu = 0$

Qual é a Gaussiana?

Azul

Exercício



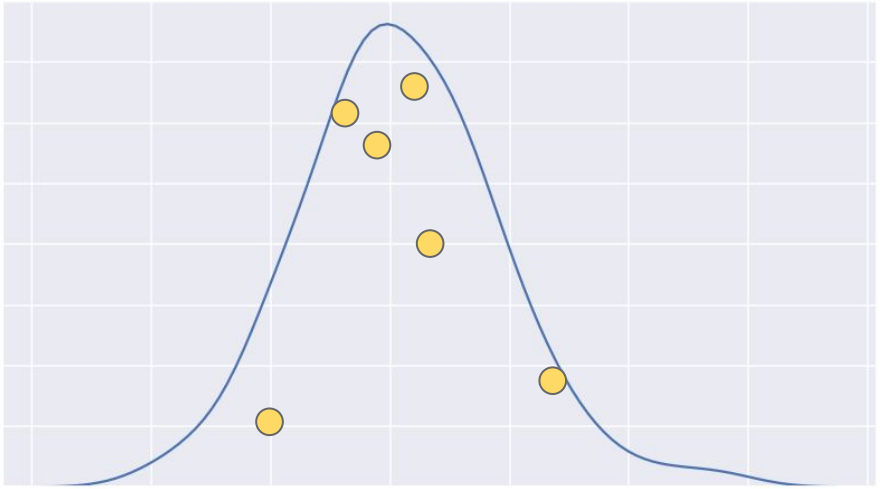
Sabendo que é a Gaussiana **Verde**, quais são os parâmetros?

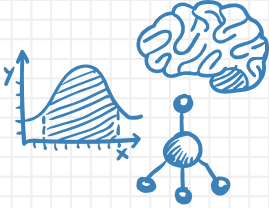
μ = -1

σ = 0.4

“Amostrar” de uma distribuição

- Utilizar o computador para “amostrar” um número aleatório com base nas probabilidades descritas por uma distribuição de probabilidades.





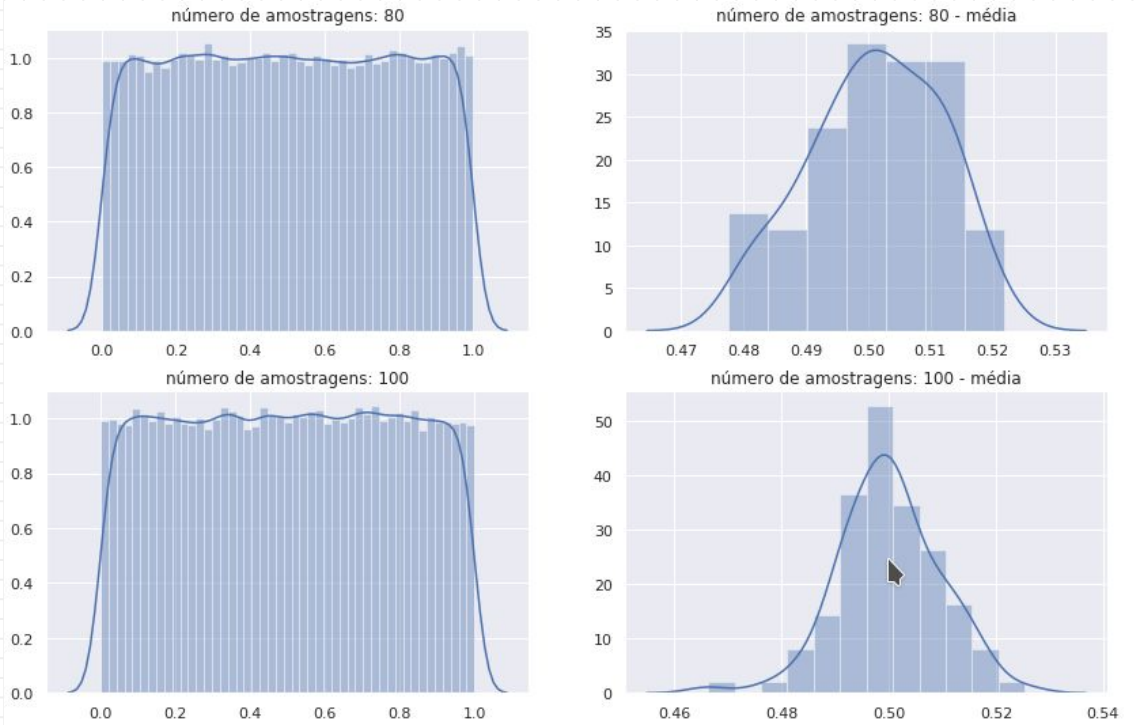
“Even if you’re not normal, the average is normal”

Joshua Starmer, Assistant Professor of University of North Carolina

Teorema do Limite Central

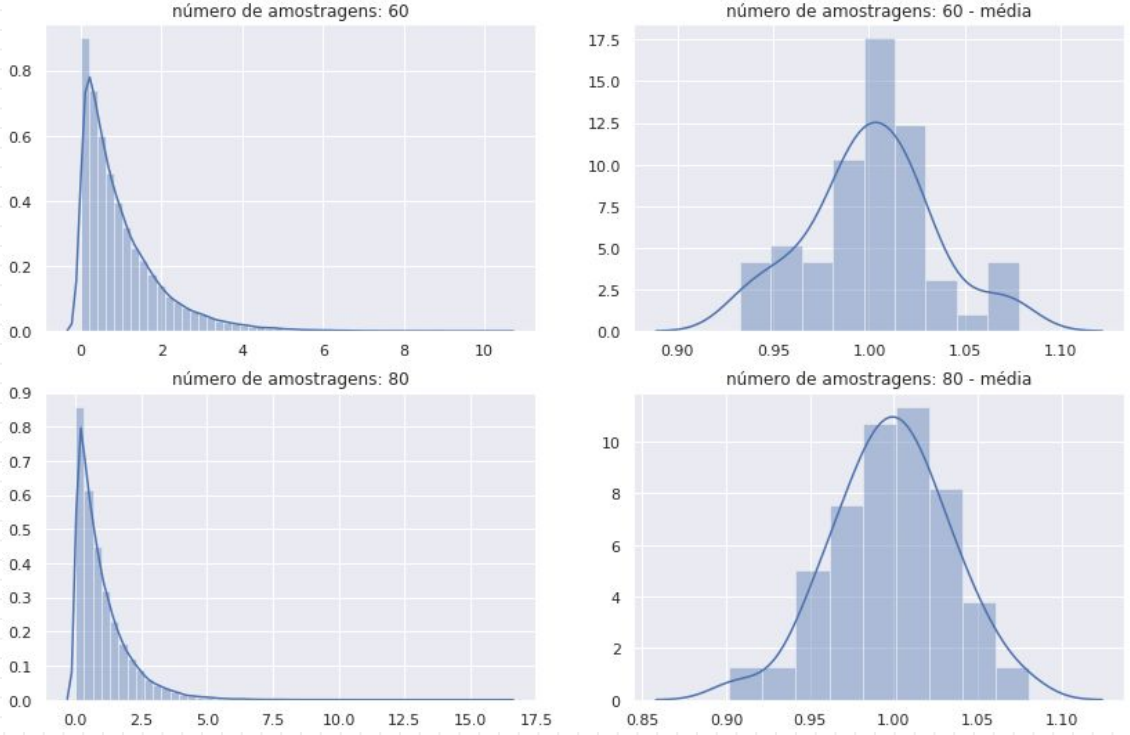
- Se amostrarmos aleatoriamente exemplos de qualquer distribuição, a distribuição das médias destes exemplos se aproximará de uma distribuição normal
- Pode-se aplicar técnicas estatísticas existentes sem precisar ter qualquer conhecimento da distribuição original dos dados
  - Por exemplo, é improvável que a média dos exemplos esteja a mais de dois desvios padrão da média

Teorema do Limite Central



Exemplo com distribuição uniforme.

Teorema do Limite Central

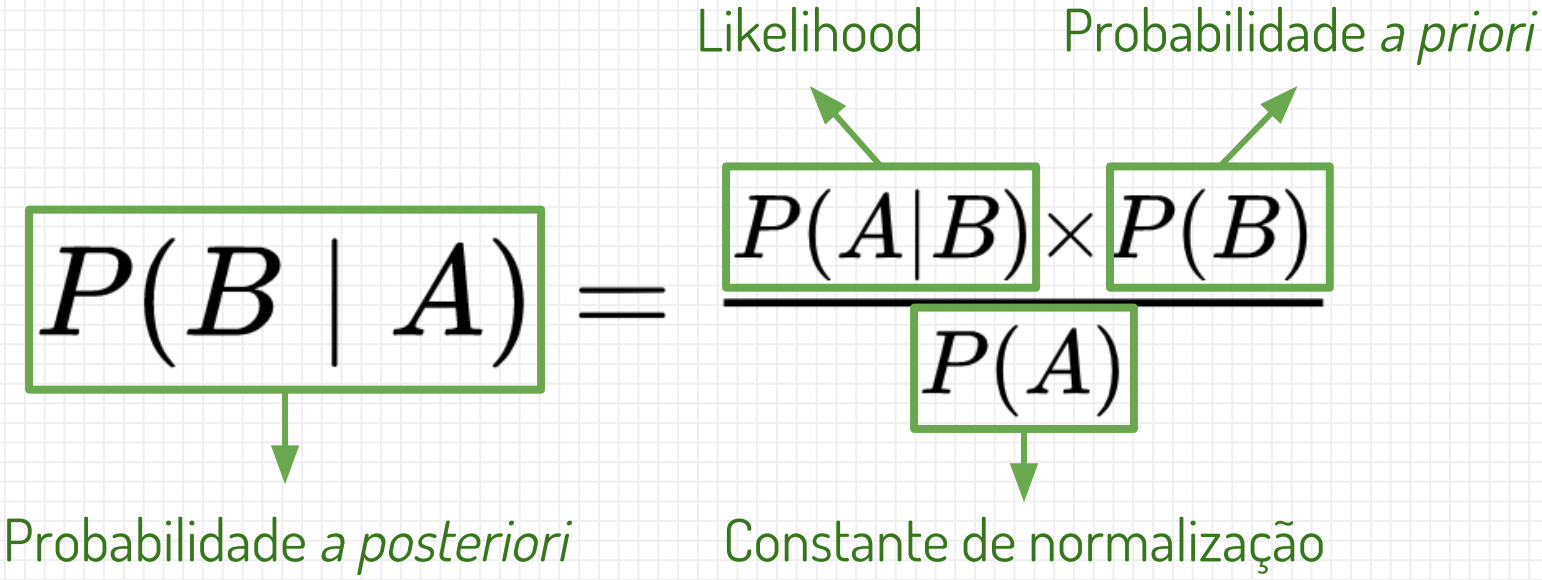


Exemplo com distribuição exponencial.

Thomas Bayes

- 1702 - 1761
- Matemático inglês
- Principal contribuição é o **Teorema de Bayes**
  - Inferência Bayesiana
  - Estatística Bayesiana

Teorema de Bayes



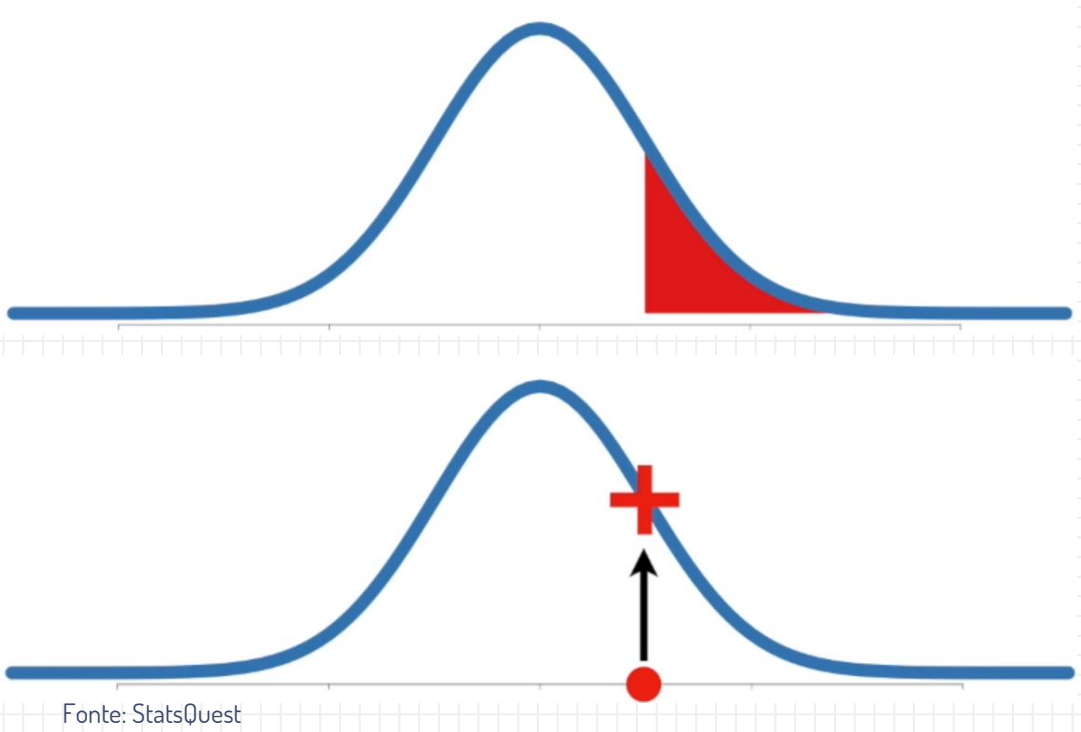
Probabilidade vs Likelihood

Probabilidade

$$P(x_i|\mu,\sigma)$$

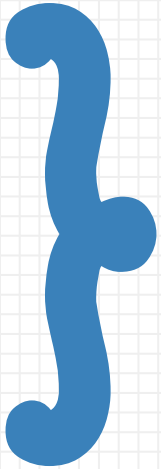
Likelihood

$$\mathcal{L}(\mu,\sigma|x_i)$$



Fonte: StatsQuest

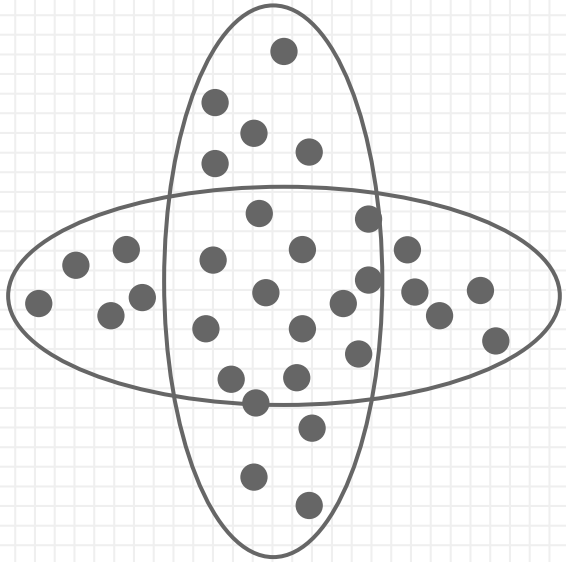
DEMO TIME



Gaussian Mixture Models

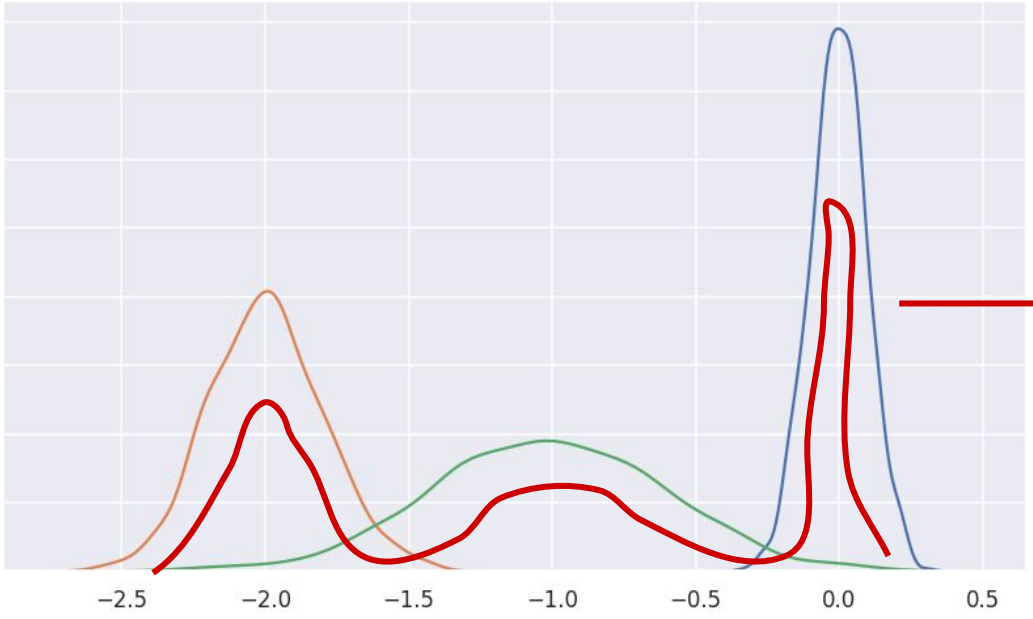
GMM

- Gaussian Mixture Model (GMM) é um modelo probabilístico para representar subpopulações distribuídas normalmente dentro de uma população maior
- Pode ser utilizado para realizar agrupamento de dados com partição *soft*
  - Possui atribuição probabilística de instâncias aos clusters
  - Consegue encontrar clusters não elípticos/esféricos

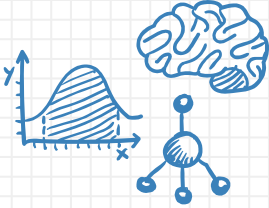


Intuição

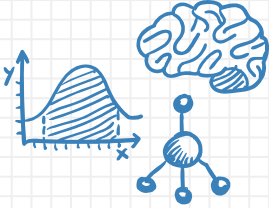
- Imagine que nossos dados podem ser representados por Gaussianas



GMM: combinação linear das diferentes Gaussianas



O que precisamos saber para definir uma distribuição normal qualquer?



O que precisamos saber para definir uma distribuição normal qualquer?

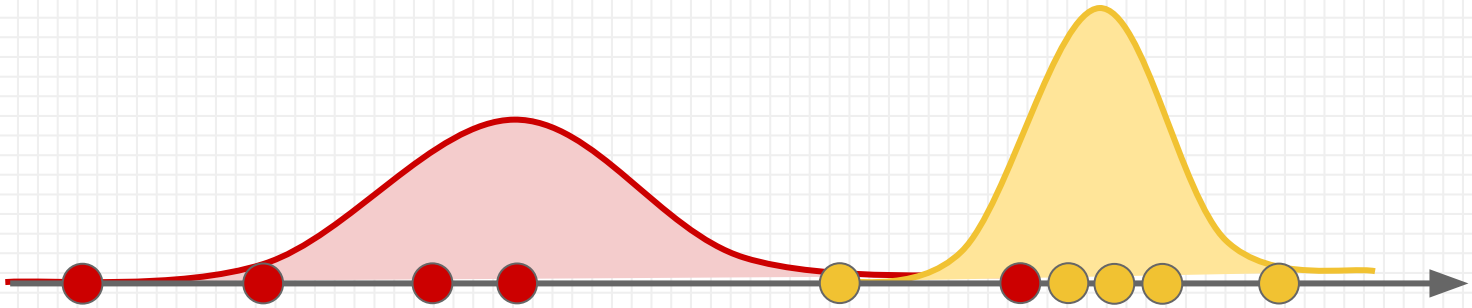
$$\mathcal{N}(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

GMM em 1D

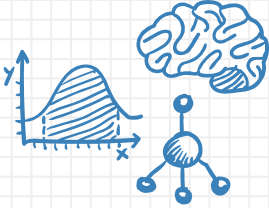
- ✘ Cada instância tem apenas uma dimensão;
- ✘ Vamos assumir que sabemos as fontes dos dados (**vermelho** e **amarelo**);
- ✘ μ e σ não são conhecidos.

$$\mu_v = \frac{1}{n_v} \sum_{i=1}^{n_v} x_i$$

$$\sigma_v = \sqrt{\frac{1}{n_v - 1} \sum_{i=1}^{n_v} (x_i - \mu_v)^2}$$



Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.



O que acontece se não soubermos quais são as fontes de dados (e.g. removermos as cores)?

GMM em 1D

- ✘ Cada instância tem apenas uma dimensão;
- ✘ Não sabemos as fontes de dados (cores);
- ✘ μ e σ não são conhecidos.

Assumindo que sabemos μ e σ, o que podemos fazer ?

Dizer se uma instância possui maior probabilidade de ser **vermelha** ou **amarela**.

$$P(v \mid x_i) = \frac{P(x_i|v)P(v)}{P(x_i|v)P(v)+P(x_i|a)P(a)}$$

$$P(x_i \mid v) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{(x_i-\mu_v)^2}{2\sigma_v^2}}$$

Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

GMM em 1D

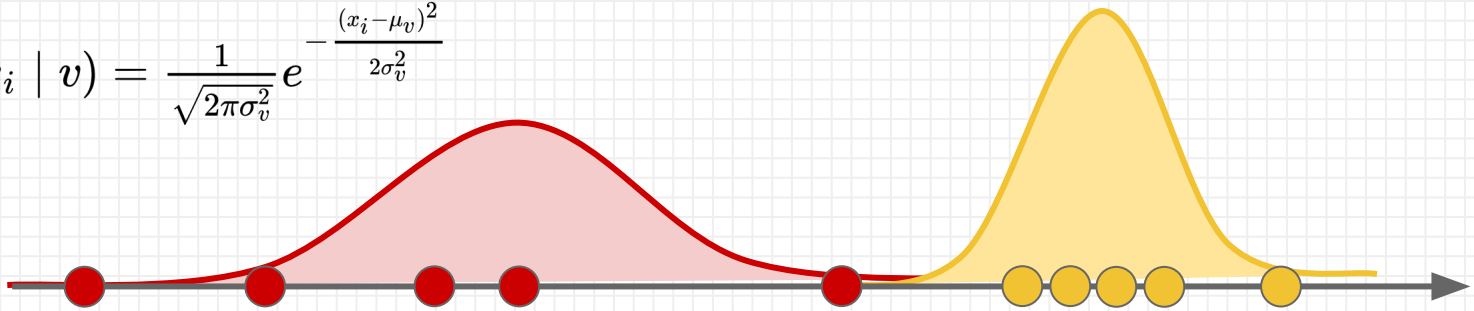
- ✘ Cada instância tem apenas uma dimensão;
- ✘ Não sabemos as fontes de dados (cores);
- ✘ μ e σ não são conhecidos.

Assumindo que sabemos μ e σ, o que podemos fazer ?

Dizer se uma instância possui maior probabilidade de ser **vermelha** ou **amarela**.

$$P(v \mid x_i) = \frac{P(x_i|v)P(v)}{P(x_i|v)P(v)+P(x_i|a)P(a)}$$

$$P(x_i \mid v) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{(x_i-\mu_v)^2}{2\sigma_v^2}}$$

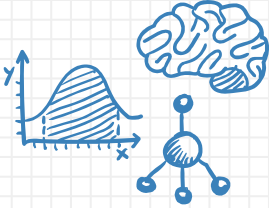


Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Expectation Maximization (EM)

- ✘ *Chicken and egg problem*, pois precisamos
  - ✗ da média e desvio padrão de cada cor para atribuir instâncias
  - ✗ saber as cores das instâncias para calcular média e desvio padrão

Ai que entra o algoritmo Expectation Maximization!



"In statistics, an expectation–maximization (EM) algorithm is an iterative method to find maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables."

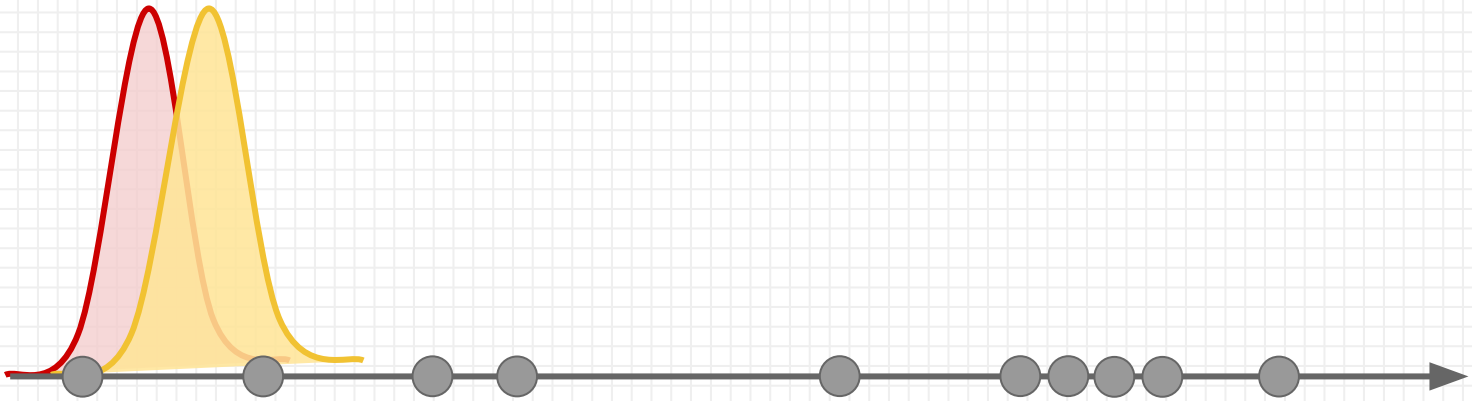
Wikipedia

Expectation Maximization (EM)

1. Iniciar com **K** Gaussianas escolhidas aleatoriamente
2. Para cada instância, calcular probabilidade de pertencer a cada uma das **K** Gaussianas **(Expectation)**
3. Recalcular médias e desvios padrão das **K** Gaussianas de acordo com os dados **(Maximization)**
4. Iterar até convergir

Expectation Maximization (EM)

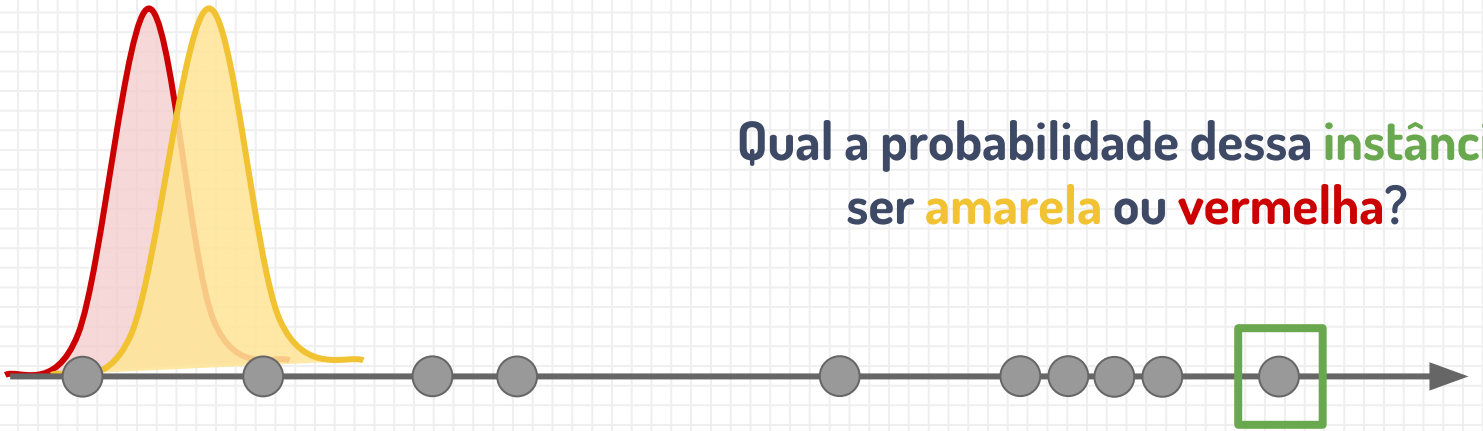
1. Iniciar com **K** Gaussianas escolhidas aleatoriamente



Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Expectation Maximization (EM)

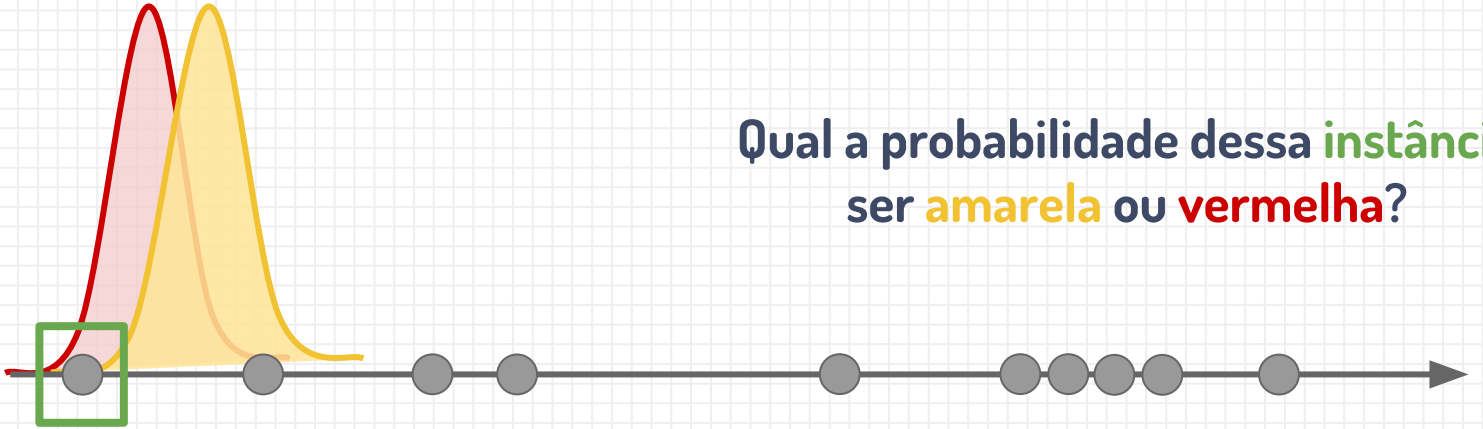
2. Para cada instância, calcular probabilidade de pertencer a cada uma das **K** Gaussianas **(Expectation)**



Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Expectation Maximization (EM)

2. Para cada instância, calcular probabilidade de pertencer a cada uma das **K** Gaussianas **(Expectation)**



Qual a probabilidade dessa instância ser amarela ou vermelha?

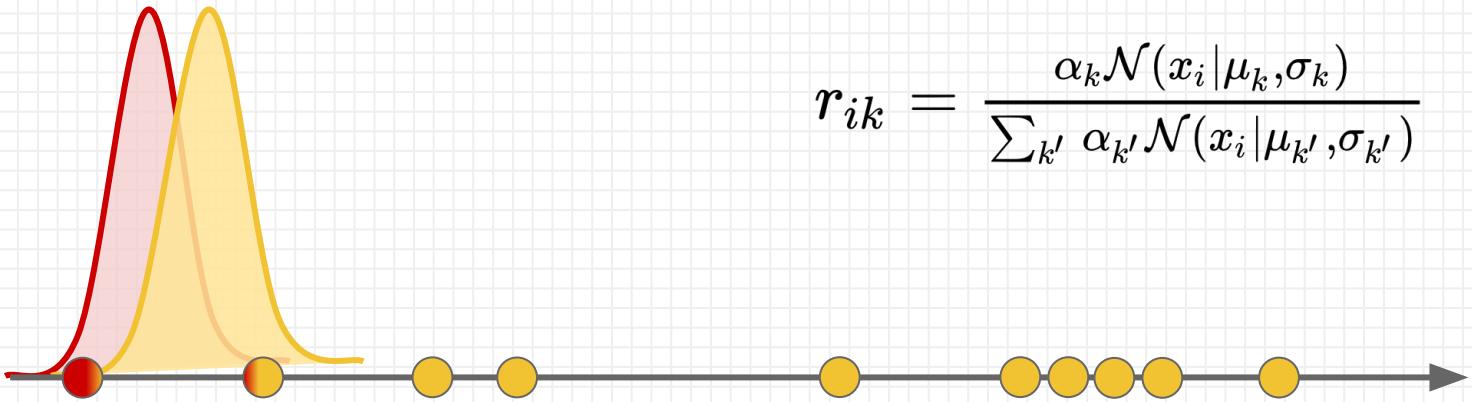
Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Expectation Maximization (EM)

2. Para cada instância, calcular probabilidade de pertencer a cada uma das **K** Gaussianas **(Expectation)**

A probabilidade de uma instância **i** pertencer a Gaussiana **k** é dada pela "responsabilidade":

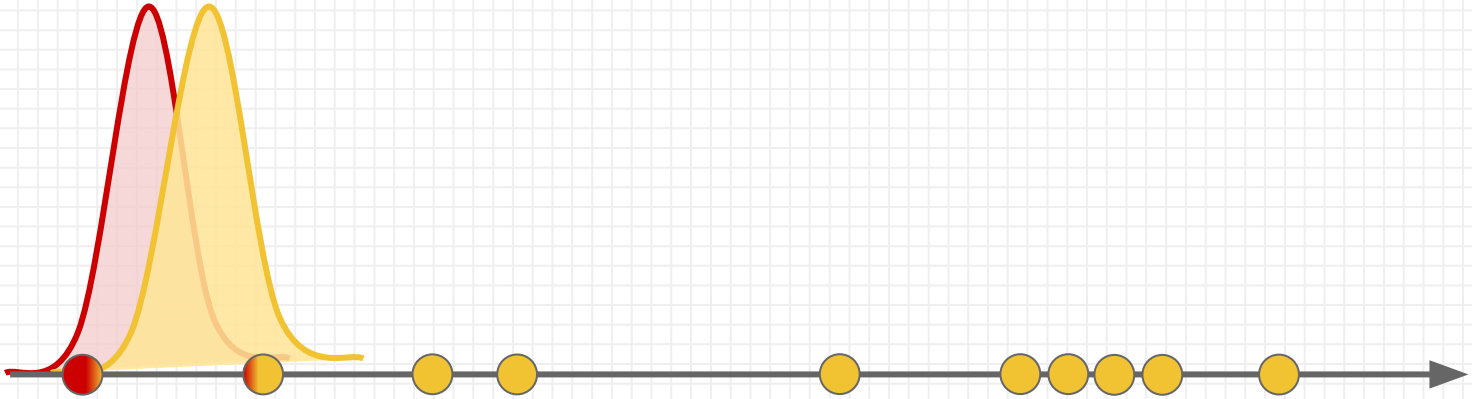
$$r_{ik} = \frac{\alpha_k \mathcal{N}(x_i|\mu_k,\sigma_k)}{\sum_{k'} \alpha_{k'} \mathcal{N}(x_i|\mu_{k'},\sigma_{k'})}$$



Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Expectation Maximization (EM)

3. Recalcular médias e desvios padrão das **K** Gaussianas de acordo com os dados **(Maximization)**



Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Expectation Maximization (EM)

3. Recalcular médias e desvios padrão das **K** Gaussianas de acordo com os dados **(Maximization)**

$$n_k = \sum_{i=1} r_{ik}$$

Responsabilidade total da Gaussiana **k**

$$\sigma_k^{new} = \sqrt{\frac{1}{n_k} \sum_{i=1} r_{ik} (x_i - \mu_k^{new})^2}$$

Desvio padrão das instâncias da Gaussiana **k**

$$\mu_k^{new} = \frac{1}{n_k} \sum_{i=1} r_{ik} x_i$$

Média ponderada das instâncias da Gaussiana **k**

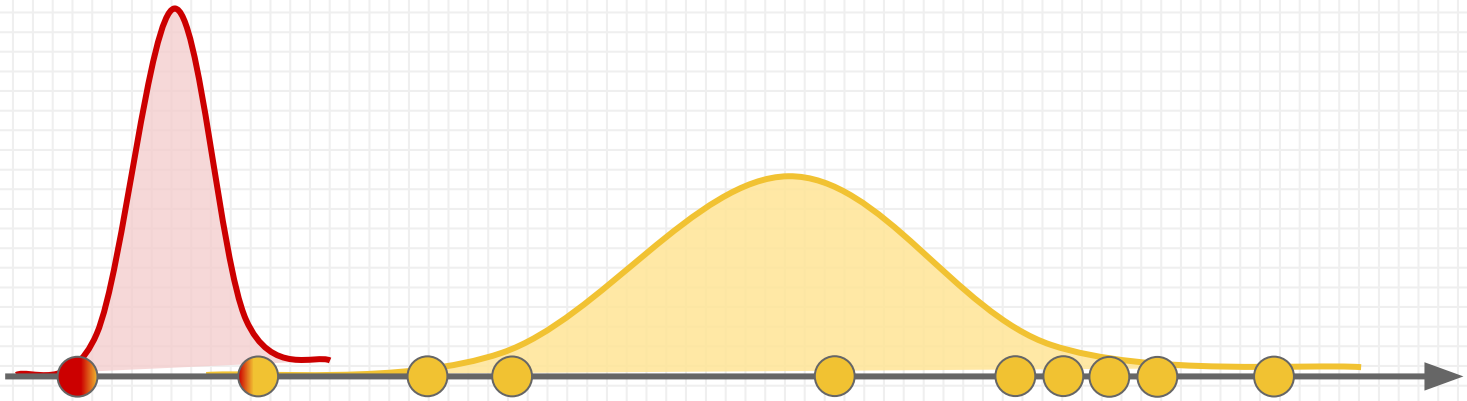
$$\alpha_k^{new} = \frac{n_c}{n}$$

Fração do total atribuída à Gaussiana **k**

Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Expectation Maximization (EM)

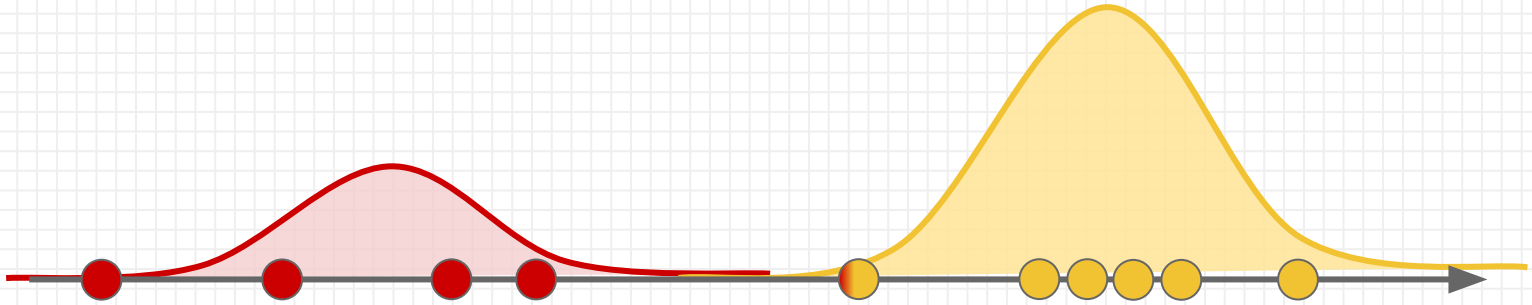
3. Recalcular médias e desvios padrão das **K** Gaussianas de acordo com os dados **(Maximization)**



Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Expectation Maximization (EM)

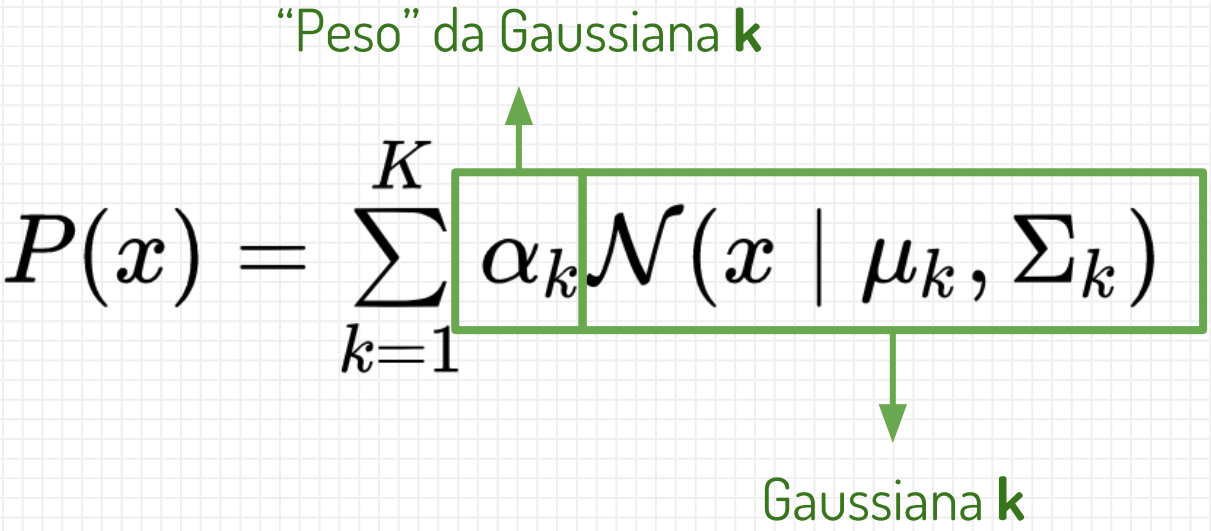
4. Iterar até convergir
   - ✘ Cada passo do EM aumenta o *log-likelihood*
   - ✘ Comparamos o *log-likelihood* da iteração anterior com o *log-likelihood* da iteração atual
   - ✘ Caso a diferença seja menor que um limiar pré-definido, terminamos



Fonte: Baseado no material do professor Dr. Victor Lavrenko, Universidade de Edinburgh.

Formalizando GMMs

- Um modelo de mistura de gaussianas é dado pela seguinte função



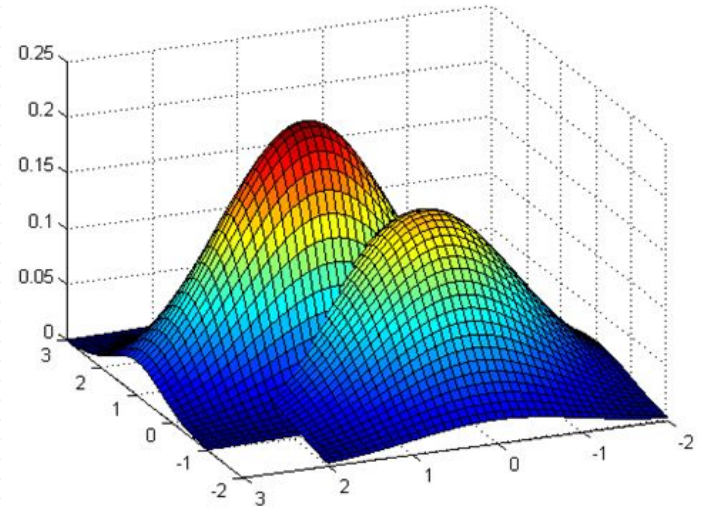
Log-likelihood

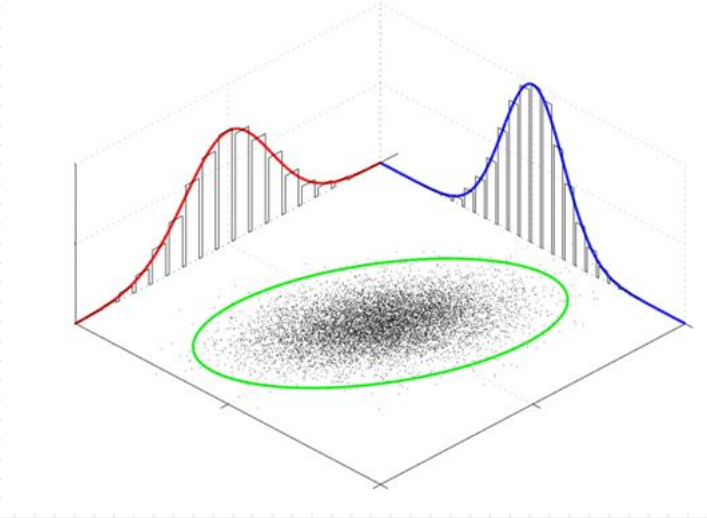
$$\sum_{i=1}^{N} \left(\log \sum_{k=1}^{K} \alpha_k \mathcal{N}(x \mid \mu_k, \Sigma_k) \right)$$

- No caso de GMMs, função é utilizada para avaliar convergência;
- Objetivo é maximizar esse valor (Maximum Likelihood Estimation);
- Podemos entender como maximizar a compatibilidade das Gaussianas com os dados.

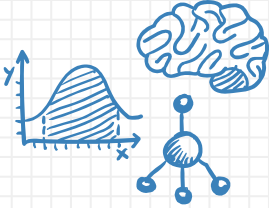
GMM multivariada

- ✘ Vimos até agora exemplos com uma dimensão;
- ✘ No entanto, GMM pode ser aplicada para dados com mais dimensões.





DEMO TIME



O que muda quando temos mais dimensões?

GMM multivariada

- Estruturas de dados ficam mais complexas

$$x_i = (x_{id_1}, x_{id_2}, \cdots, x_{id_n})$$

$$\mu_k = (\mu_{kd_1}, \mu_{kd_2}, \cdots, \mu_{kd_n})$$

O que acontece com o desvio padrão/variância?
Vira uma matriz de covariância! $\sum_k$

GMM multivariada

- Matriz de covariância, considerando duas variáveis

$$\Sigma = \begin{bmatrix} \sigma(x) & cov(x,y) \\ cov(y,x) & \sigma(y) \end{bmatrix}$$

- No passo de atualização do EM, a atualização da matriz ficaria

$$\Sigma_k^{new} = \frac{1}{n_k} \sum_{i=1} r_{ik}(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T$$

GMM multivariada

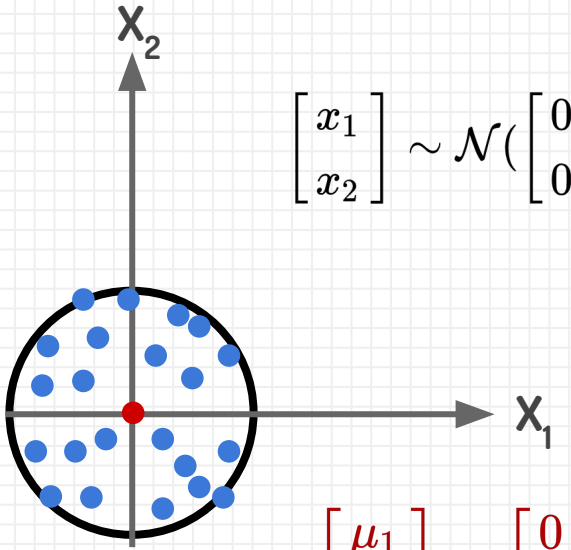
- Gaussiana multivariada

$$\mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

μ médias

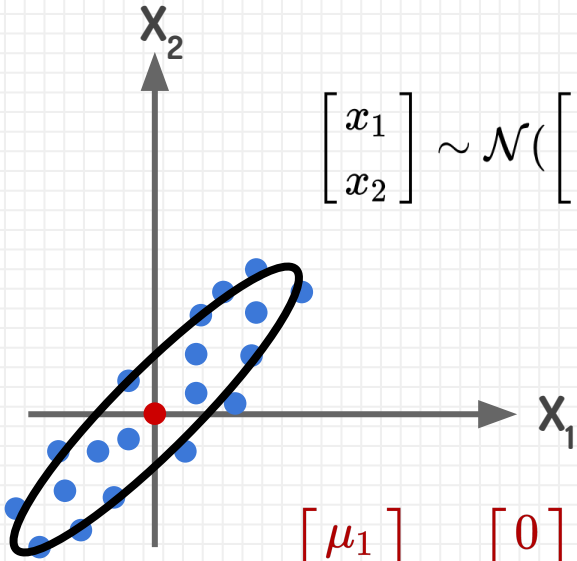
Σ matriz de covariância

Intuição da matriz de covariância



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix})$$

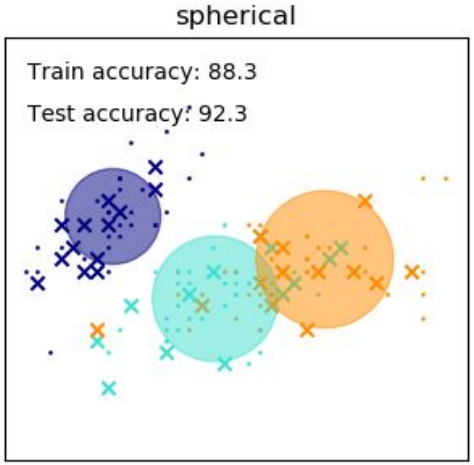
$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

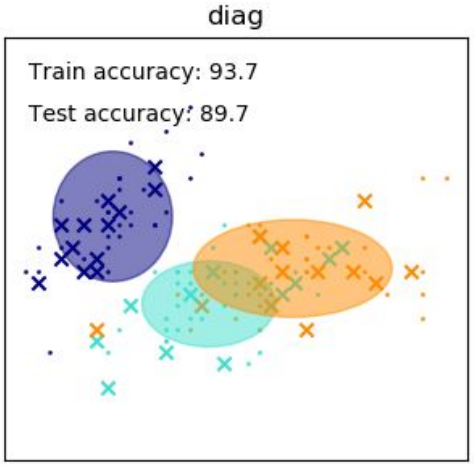


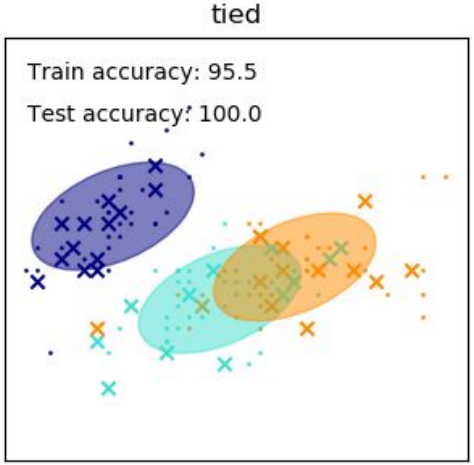
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix})$$

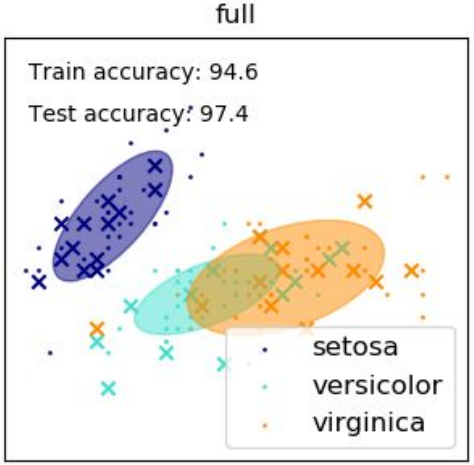
$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Fonte: Roberto Silveira, Uncertainty in Deep Learning (https://www.slideshare.net/rsilveira79/uncertainty-in-deep-learning).





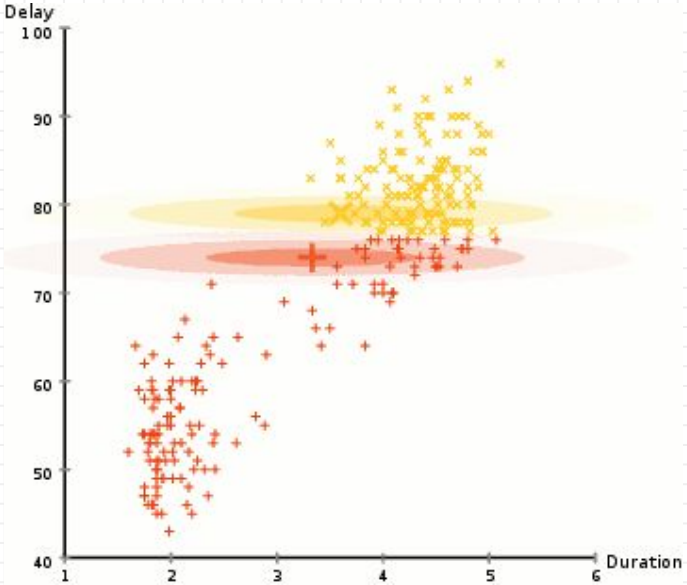


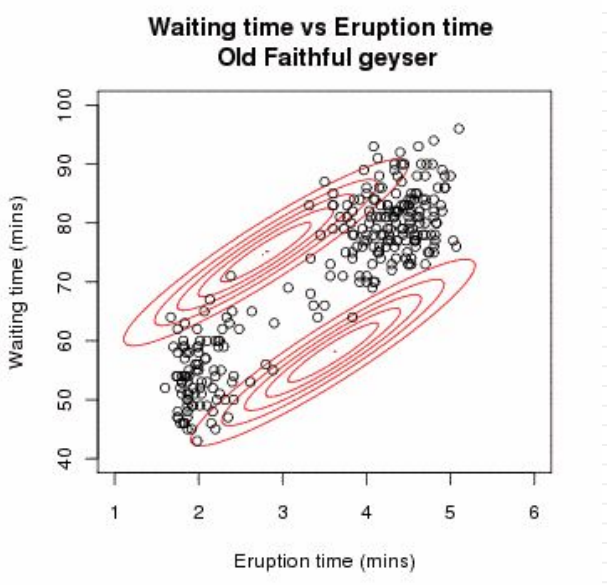


Fonte: Scikit Learn

DEMO TIME

EM para GMM multivariada – Exemplo





Old Faithful Geyser Data

Fonte: Wikipedia.

GMM – Alguns Prós e Contras

- **Prós**
  - Produz resultados ricos (probabilidade de instâncias pertencerem a determinados grupos)
  - Pode produzir uma partição rígida (maiores probabilidades)
  - Capaz de encontrar grupos em formato de elipse
  - Pode ser utilizado como um modelo generativo
- **Contras**
  - Necessita do número de clusters *a priori*
  - Inversão da matriz de covariância tem custo computacional alto
  - Inicialização
    - Uma forma de resolver é utilizar K-means como inicialização

DEMO TIME

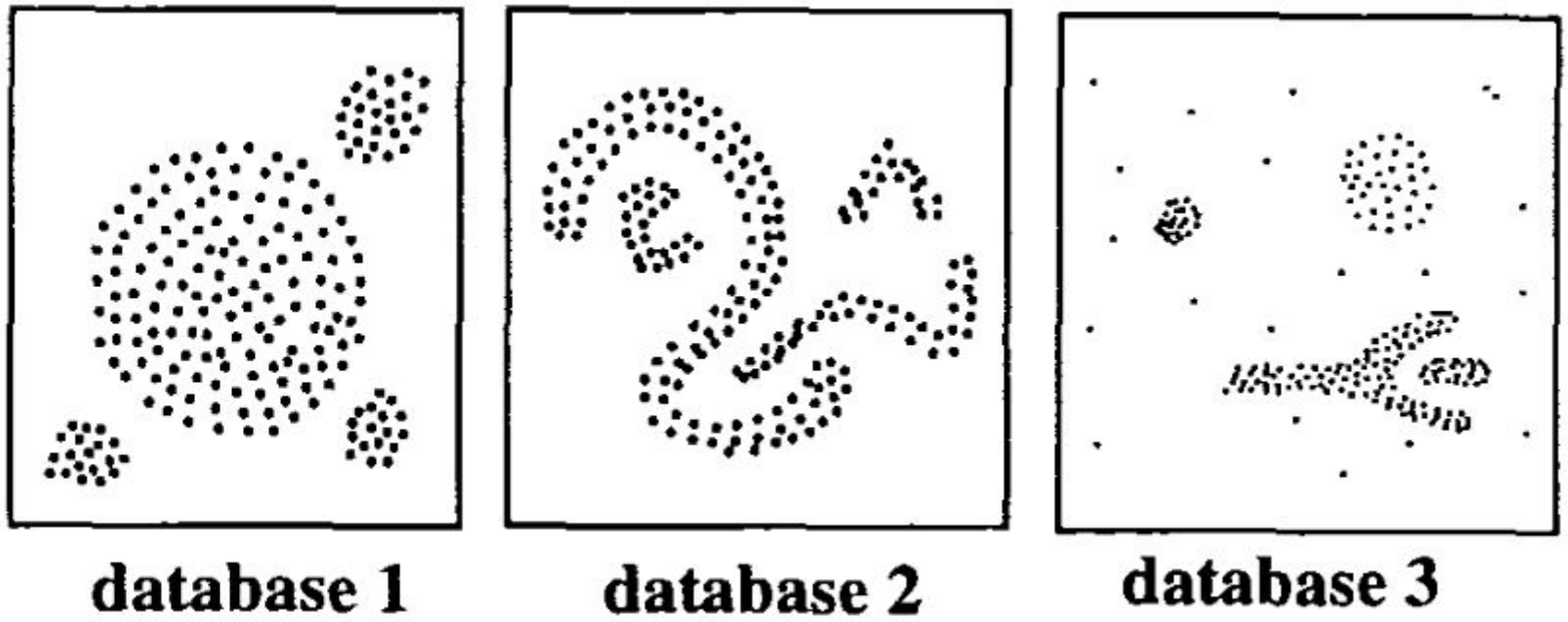
DBSCAN

Agrupamento baseado em densidade

- **Clusters são regiões densas** no espaço de dados, as quais são separadas por regiões de baixa densidade;
- Sua ideia é baseada na forma com que seres humanos percebem grupos de dados.

Vamos ver um exemplo.

Agrupamento baseado em densidade



Fonte: Ester, Martin, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." Kdd. Vol. 96. No. 34. 1996.

DBSCAN

- Density-Based Spatial Clustering of Applications with Noise (1996)
- A ideia chave é que para cada instância de um cluster, a vizinhança em um dado raio deve conter um mínimo número de instâncias
- Diferentemente do K-means, o DBSCAN
  - Não necessita do número de clusters *a priori*
  - Consegue encontrar clusters de diferentes formatos
  - Consegue identificar outliers

DBSCAN


KDD
SIGKDD AWARDS
Home / Awards / 2014 SIKDD Test of Time Award Winners
2014 SIKDD TEST OF TIME AWARD WINNERS
2014 KDD Test of Time Award Award
The KDD Test of Time Award recognizes outstanding papers from past KDD Conferences beyond the last decade that have had an important impact on the data mining research community.
The 2014 Test of Time award recognizes the following influential contributions to SIGKDD that have withstood the test of time:
A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise [KDD 1996]
Martin Ester, Hans-Peter Kriegel, Joerg Sander, Xiaowei Xu

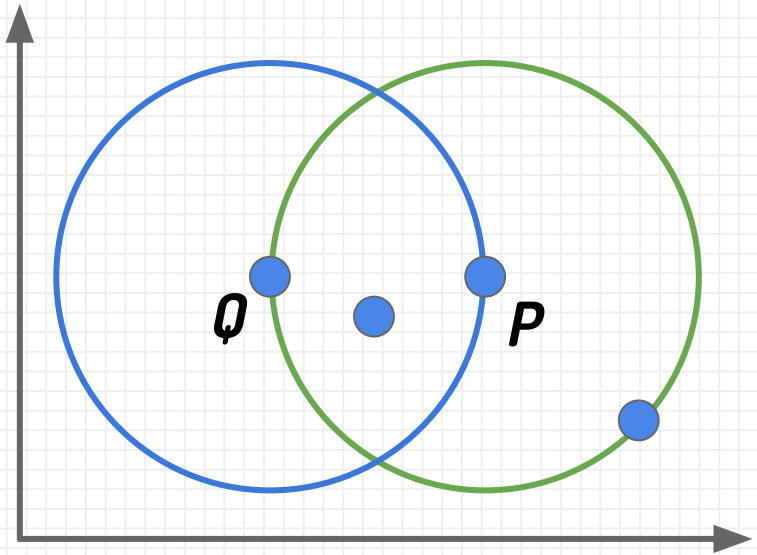
DEMO TIME

DBSCAN

- Como funciona (em alto nível)
  - Três tipos de pontos: *core*, *border*, *noise*
  - Conecta pontos *core* nos *clusters*
  - Atribui pontos *border* aos clusters
- Densidade de um ponto é definida através de dois parâmetros
  - ε (epsilon) – raio da vizinhança do ponto
  - ***MinPoints*** – número mínimo de pontos da vizinhança

O que é alta densidade?

- ε-vizinhança de uma instância contém no mínimo ***MinPoints***



ε-vizinhança de ***P***

ε-vizinhança de ***Q***

Qual é a densidade quando MinPoints = 4?

Densidade de ***P*** é alta

Densidade de ***Q*** é baixa

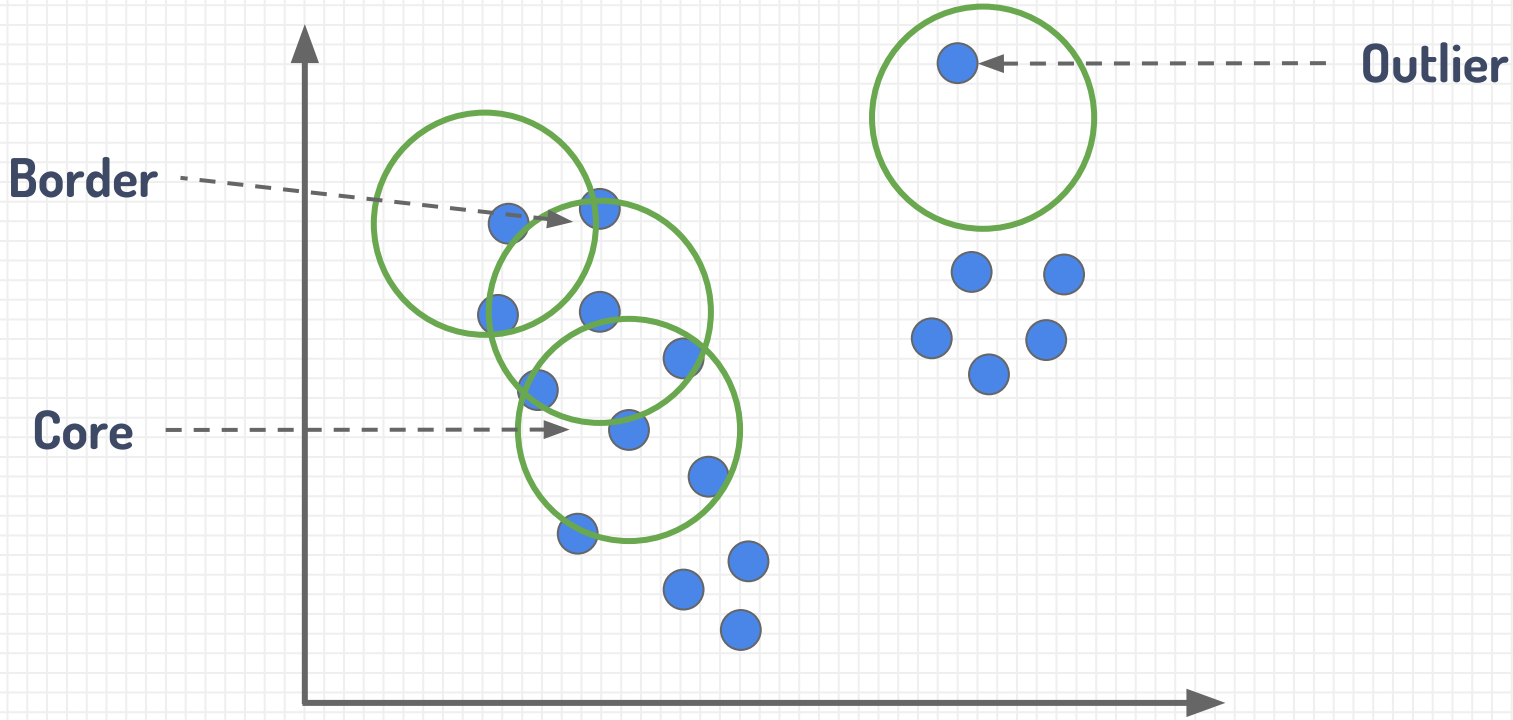
Fonte: Moise, Izabela et al., "Density-based Clustering", ETH Zürich, 2015

Core, Border e Outlier

- **Core (Central):** instância cuja densidade é alta
  - Ou seja, possui mais do que ***MinPoints*** no raio Epsilon – pontos que estão no interior de um cluster
- **Border (Borda):** instância cuja densidade é baixa (mas está na vizinhança de alguma instância Core)
  - Ou seja, possui menos do que ***MinPoints*** no raio Epsilon
- **Noise (Outlier):** Qualquer instância que não é Core nem Border

Fonte: Moise, Izabela et al., "Density-based Clustering", ETH Zürich, 2015

Core, Border e Outlier



Fonte: Moise, Izabela et al., "Density-based Clustering", ETH Zürich, 2015

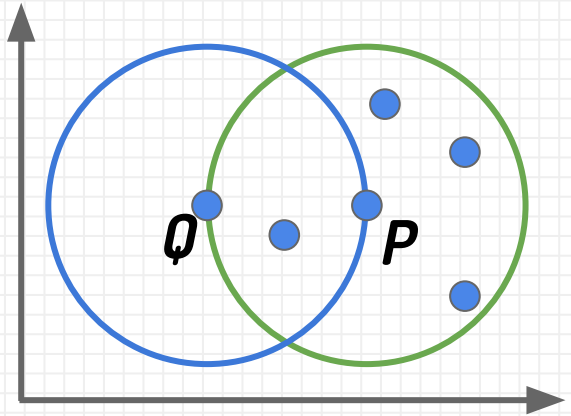
DEMO TIME

Como os clusters são formados?

- Para entendermos isso, precisamos compreender alguns conceitos
  - Alcance por Densidade, onde instâncias são:
    - Diretamente Alcançáveis por Densidade
    - Alcançáveis por Densidade
  - Conexão por Densidade, onde instâncias são
    - Conectadas por Densidade

Alcance por Densidade

- Uma instância ***Q*** é diretamente alcançável por densidade de uma instância ***P*** se:
  - ***P*** é uma instância Core
  - ***Q*** está na Epsilon-vizinhança de ***P***



Fonte: Lee, Wondong, http://www.cs.fsu.edu/~ackerman/CIS5930/notes/DBSCAN.pdf

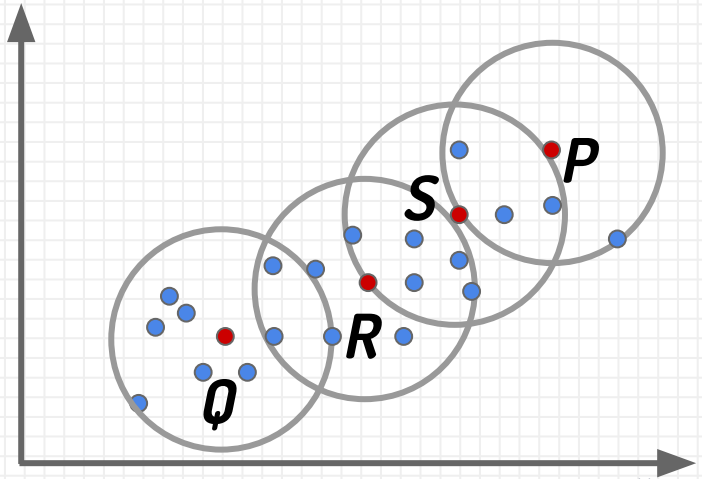
Considerando MinPoints=4

Q **é diretamente alcançável por densidade de** ***P*****?** Sim!

P **é diretamente alcançável por densidade de** ***Q*****?** Não! Por quê?

Alcance por Densidade

- Uma instância ***P*** é alcançável por densidade de uma instância ***Q*** se:
  - Existe uma cadeia de instâncias de ***Q*** até ***P***, onde todas as instâncias são diretamente alcançáveis por densidade



Considerando MinPoints=7

P é diretamente alcançável por densidade de ***S***

S é diretamente alcançável por densidade de ***R***

R é diretamente alcançável por densidade de ***Q***

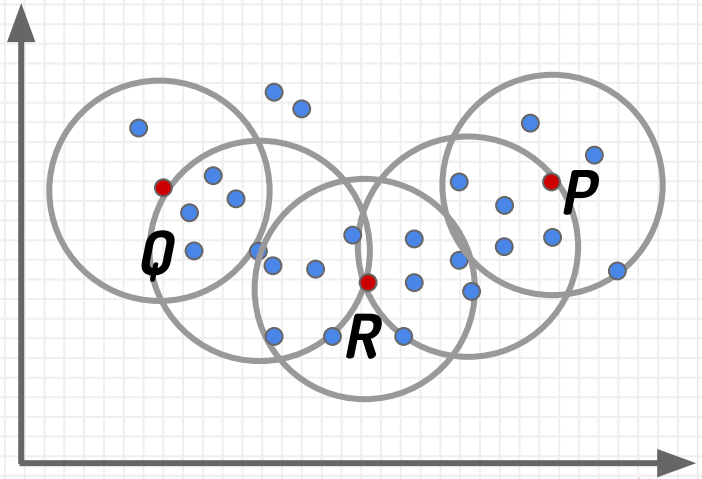
P é (inderamente) alcançável por densidade de ***Q***

***Q* é alcançável por densidade de *P*?** Não! Por quê?

Fonte: Lee, Wondong, http://www.cs.fsu.edu/~ackerman/CIS5930/notes/DBSCAN.pdf

Conexão por Densidade

- Um par de instâncias ***Q*** e ***P*** é conectado por densidade se ambos são alcançáveis por densidade de uma instância ***R***



Considerando MinPoints=7

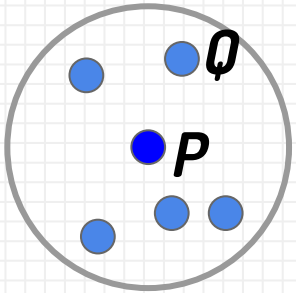
P é alcançável por densidade de ***R***

Q é alcançável por densidade de ***R***

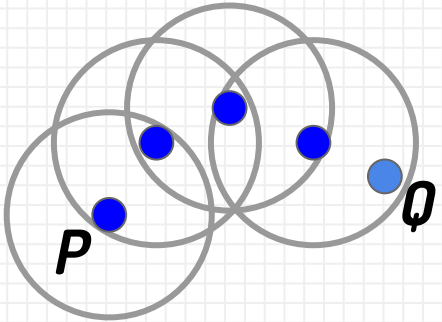
Portanto, ***Q*** e ***P*** são conectados por densidade através de ***R***

Fonte: Lee, Wondong, http://www.cs.fsu.edu/~ackerman/CIS5930/notes/DBSCAN.pdf

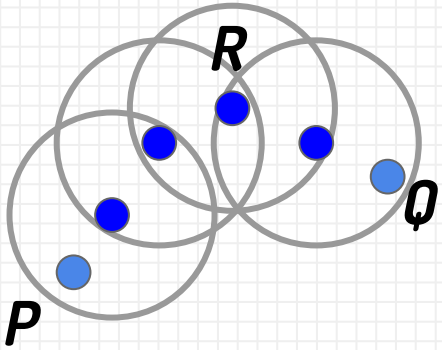
Em resumo



Q é diretamente alcançável por densidade de **P**



Q é alcançável por densidade de **P**



Q e **P** são conectados por densidade através **R**

Um cluster é um conjunto de instâncias que são conectadas por densidade.

Fonte: Schlitter, Nico, and Tanja Falkowski. "Dengraph-ho: Density-based hierarchical community detection for explorative visual network analysis." Research and Development in Intelligent Systems XXVIII. Springer, London, 2011.

DBSCAN – Alguns Prós e Contras

- **Prós**
  - Consegue encontrar clusters de formatos diferentes
  - É capaz de identificar outliers
  - Não é necessário informar número de grupos
- **Contras**
  - Não consegue lidar com densidades variadas
  - É sensível à escolha de hiperparâmetros
  - Mais difícil de utilizar em dados com altas dimensionalidade

DEMO TIME

DBSCAN vs K-means

- Algumas diferenças

DBSCAN	K-means
Descarta instâncias classificadas como ruído	Em geral, agrupa todas as instâncias
Baseado no conceito de densidade	Baseado no conceito de protótipos (centróides)
Consegue lidar com clusters de diferentes tamanhos e formatos	Tem dificuldade para lidar com clusters não esféricos ou de diferentes tamanhos
Geralmente tem problemas com dados de alta dimensionalidade	Pode ser aplicado neste cenário
Produz mesmos clusters em diferentes execuções com mesmos dados	Produz clusters diferentes em diferentes execuções com os mesmos dados

Considerações finais

- ✘ Vimos na aula de hoje diferentes métodos de agrupamento de dados (GMM e DBSCAN);
- ✘ GMM, além de um algoritmo de agrupamento que gera uma partição não rígida, é também um modelo generativo;
- ✘ DBSCAN é baseado em densidade e não necessita da definição do número de grupos
  - ✗ Porém tem dois parâmetros que devem ser escolhidos