



# Introdução ao Affinity Propagation

|           |   |
|-----------|---|
| ≡ Ciclo   | Ciclo 08: Outros algoritmos Clusterização |
| # Aula    | 63  |
| 🕒 Created | @June 26, 2023 7:11 PM                    |
| ☑ Done    | <input type="checkbox"/>                  |
| ☑ Ready   | <input checked="" type="checkbox"/>       |

## Objetivo da Aula:

- ☐ O que é Affinity Propagation?
- ☐ Os 5 passos do treinamento
- ☐ Resumo
- ☐ Próxima aula

## Conteúdo:

### ▼ 1. O que é Affinity Propagation?

Affinity Propagation é um algoritmo de clusterização que usa uma abordagem baseada em grafos para encontrar automaticamente um número de clusters ou agrupamentos em um conjunto de dados. O algoritmo não requer a especificação prévia do número de clusters desejados, o que pode ser uma vantagem em algumas situações.

O Affinity Propagation usa uma matriz de similaridade para modelar as relações entre os elementos do conjunto de dados. Essa matriz é usada para construir um grafo de similaridade, onde cada nó representa um elemento do conjunto de dados e as arestas representam a similaridade entre os elementos. O algoritmo usa esse grafo para identificar um conjunto de "exemplares", que são pontos que representam os agrupamentos.

#### ▼ 1.1 As 4 matrizes do algoritmo Affinity Propagation

##### ▼ 1.1.1 A matriz similaridade (S)

###### ▼ Matriz similaridade (S):

| Participantes | Alice | Bob | Cary | Doug | Edna |
|---------------|-------|-----|------|------|------|
| Alice         | -16   | -1  | 1    | -6   | -11  |
| Bob           | 10    | -15 | -10  | -10  | -15  |
| Cary          | 11    | -11 | -16  | -12  | -15  |
| Doug          | -9    | -14 | -15  | -19  | 9    |
| Edna          | -14   | -19 | -18  | 14   | -19  |

###### ▼ Descrição:

A matriz de similaridade, também conhecida como matriz de afinidade, é uma matriz que descreve a relação de similaridade entre as amostras em um conjunto de dados.

Cada elemento da matriz de similaridade  $S[i, j]$  representa a medida de similaridade entre as amostras  $i$  e  $j$ . Essa medida pode ser calculada usando diferentes métricas, como distância euclidiana, coeficiente de correlação, distância de Manhattan, entre outras, dependendo do problema e das características dos dados.

A matriz de similaridade é usada como entrada no algoritmo Affinity Propagation para capturar a relação de proximidade entre as amostras. Ela fornece informações sobre quão semelhantes ou diferentes são as amostras em termos de atributos, características ou medidas relevantes.

Em resumo, a matriz de similaridade no algoritmo Affinity Propagation descreve a relação de similaridade entre as amostras do conjunto de dados. Ela é usada para calcular as responsabilidades e disponibilidades entre as amostras durante o processo iterativo do algoritmo, auxiliando na formação dos clusters.

### ▼ 1.1.2 A matriz responsabilidade (R)

#### ▼ Matriz responsabilidade (R):

| Participantes | Alice | Bob | Cary | Doug | Edna |
|---------------|-------|-----|------|------|------|
| Alice         | -16   | -1  | 1    | -6   | -11  |
| Bob           | 10    | -15 | -10  | -10  | -15  |
| Cary          | 11    | -11 | -16  | -12  | -15  |
| Doug          | -9    | -14 | -15  | -19  | 9    |
| Edna          | -14   | -19 | -18  | 14   | -19  |

#### ▼ Descrição:

A matriz de responsabilidade (R) representa a "responsabilidade" que uma amostra atribui a outra amostra em se tornar um "exemplar".

A "responsabilidade", no contexto do algoritmo Affinity Propagation, é uma medida que indica o quanto uma amostra considera outra amostra como um "exemplar". Em outras palavras, é uma forma de "atribuir responsabilidade" a uma amostra em se tornar parte de um cluster.

Um "exemplar" refere-se a uma amostra que é selecionada para representar um cluster. De outro modo, é um ponto dos dados que é considerado altamente representativo e característico do grupo ao qual pertence e portanto, pertencente ao cluster.

Resumindo, a matriz de responsabilidade (R) representa a medida do quanto uma amostra considera outra amostra como representante de um cluster.

### ▼ 1.1.3 A matriz disponibilidade (D)

#### ▼ Matriz disponibilidade (D):

| Participantes | Alice | Bob | Cary | Doug | Edna |
|---------------|-------|-----|------|------|------|
| Alice         | 21    | -15 | -16  | -5   | -10  |
| Bob           | -5    | 0   | -15  | -5   | -10  |
| Cary          | -6    | -15 | 1    | -5   | -10  |
| Doug          | 0     | -15 | -15  | 14   | -19  |
| Edna          | 0     | -15 | -15  | -19  | 9    |

#### ▼ Descrição:

No contexto do algoritmo Affinity Propagation, a "disponibilidade" refere-se à medida da adequação de uma amostra para ser escolhida como um exemplar. É uma estimativa da capacidade de uma amostra em representar um cluster de forma consistente.

### ▼ 1.1.4 A matriz critério (C)

▼ **Matriz critério (C):**

| Participantes | Alice | Bob | Cary | Doug | Edna |
|---------------|-------|-----|------|------|------|
| Alice         | 5     | -16 | -15  | -11  | -21  |
| Bob           | 5     | -15 | -25  | -15  | -25  |
| Cary          | 5     | -26 | -15  | -17  | -25  |
| Doug          | -9    | -29 | -30  | -5   | -10  |
| Edna          | -14   | -34 | -33  | -5   | -10  |

▼ **Descrição:**

No contexto do algoritmo Affinity Propagation, o valor do critério é a soma da responsabilidade e disponibilidade de cada par de ponto. A coluna com o maior valor de critério para cada linha identifica o “exemplo exemplar” daquele ponto. Os pontos de cada linha que compartilham o mesmo “exemplo exemplar” estão no mesmo cluster.

## ▼ 2. Os 5 passos do treinamento

Os passos para encontrar os grupos (clusters) formados pelos dados, usando o algoritmo de Affinity Propagation são os seguintes:

1. Definição da **métrica de similaridade**
2. Cálculo da similaridade entre todos os pontos do conjunto de dados, formando a **matriz de similaridade (S)**
3. **Até o número n de repetições ser alcançada** ou a variação dos valores das matrizes de responsabilidade e disponibilidade for menor do que um valor  $\epsilon$ , faça:
  - a. Cálculo da **matriz de responsabilidade**
  - b. Cálculo da **matriz de disponibilidade**
4. Para cada ponto, some os valores da matriz de responsabilidade e disponibilidade, formando a **matriz de critério**
5. Atribua o mesmo **cluster** para os pontos que possuem o mesmo **valor de critério**.

## ▼ 3. Resumo

1. Affinity Propagation é um algoritmo de clusterização que usa uma abordagem baseada em grafos para encontrar automaticamente um número de clusters ou agrupamentos em um conjunto de dados.
2. O Affinity Propagation utiliza quatro matrizes: matriz de similaridade, matriz de responsabilidade, matriz de disponibilidade e matriz de critério. Cada uma delas tem uma função específica no algoritmo.
3. Os passos para encontrar os grupos (clusters) formados pelos dados, usando o algoritmo de Affinity Propagation são:
  - a. Definição da métrica de similaridade
  - b. Cálculo da similaridade entre todos os pontos do conjunto de dados, formando a matriz de similaridade (S)
  - c. Processo iterativo para cálculo das matrizes de responsabilidade e disponibilidade
  - d. Cálculo da matriz de critério
  - e. Atribuição do mesmo cluster para os pontos que possuem o mesmo valor de critério.

## ▼ 4. Próxima aula

Definição dos clusters