



## **Aprendizagem de Máquina**

### **Relatório Trabalho 3**

Acadêmicos: Bianca Carvalho de Oliveira e Gabriela Marim de Souza

Data: 27/25/2019

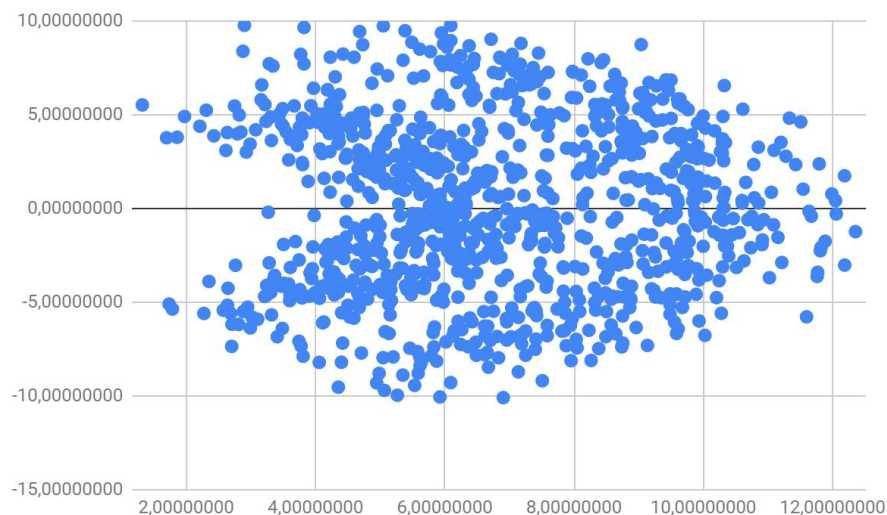
#### **Descrição do que é o trabalho**

Este trabalho tem como objetivo comparar os agrupamentos, segundo cada um dos métodos de avaliação de qualidade (coesão, separabilidade, entropia e silhueta), das estratégias de agrupamento DBSCAN, K-means e Agnes. Para tal foi necessário implementar cada um dos métodos de avaliação.

#### **Descrição do conjunto de dados**

O conjunto de dados utilizado durante os experimentos consiste em uma base, que é composta por 1000 instâncias, descritas por três atributos sendo eles dois pontos (x e y) e o terceiro uma classe usada para entropia.

O comportamento da base é descrito no gráfico abaixo.



## Descrição passo-a-passo do experimento

O experimento foi implementado na linguagem JAVA com o apoio da biblioteca Weka. No primeiro momento carregou-se a base e em seguida implementou-se os métodos para a avaliação da qualidade dos agrupamentos através das abordagens de coesão, separabilidade, entropia e silhueta.

No segundo momento implementou-se as funções para encontrar os melhores parâmetros para cada método de agrupamento, sendo eles :

**K-means:** para a estratégia K-means (baseada em protótipo ou centralidade) os parâmetros foram definidos através da implementação de funções. Essas funções contém três for um dentro do outro, onde no primeiro for variou-se, de 1 em 1 dentro de intervalo [1, 50], a semente de inicialização que determina a posição inicial dos centróides, já no segundo for variou-se, de 10 em 10 dentro de intervalo [10, 500], o número de iterações para a convergência, e por último no terceiro for variou-se, de 1 em 1 dentro de intervalo [2, 20], o número de centróides.

Obtendo-se assim, para cada método de avaliação, os seguintes parâmetros:

Coesão:

- Posição inicial dos centróides: 12
- Número de iterações para convergência: 20
- Número de centróides: 20

Separabilidade:

- Posição inicial dos centróides: 1
- Número de iterações para convergência: 10
- Número de centróides: 2

Entropia:

- Posição inicial dos centróides: 6
- Número de iterações para convergência: 40
- Número de centróides: 17

Silhueta:

- Posição inicial dos centróides: 50
- Número de iterações para convergência: 500
- Número de centróides: 20

**DBSCAN:** para a estratégia DBSCAN (baseada em densidade) os parâmetros foram definidos através da implementação uma função assim como foi feito para o K-means. Essa função contém dois for um dentro do outro, onde no primeiro for variou-se, de 1 em 1 dentro de intervalo [1, 50], o número mínimo de pontos e no segundo for variou-se, de 0.01 em 0.01 dentro de intervalo [0.2, 10], o tamanho do raio adotado.

Obtendo-se assim, para cada método de avaliação, os seguintes parâmetros:

Coesão:

- Número mínimo de pontos: 1

- Tamanho do raio adotado: 0.2
- Separabilidade:
- Número mínimo de pontos: 1
  - Tamanho do raio adotado: 0.2

- Entropia:
- Número mínimo de pontos: 1
  - Tamanho do raio adotado: 0.2

- Silhueta:
- Número mínimo de pontos:1
  - Tamanho do raio adotado: 0.2

**Agnes** : para a estratégia Agnes(baseada em hierarquia) os parâmetros adotados foram os default da biblioteca weka.

## Avaliação dos Experimentos

Depois de implementadas a leitura, escolha dos melhores parâmetros para cada estratégia de agrupamento, foi executado as três estratégias de agrupamento para cada um dos método de avaliação, obtendo-se os seguintes resultados:

	K-means	DBSCAN	Agnes
<b>Coesão</b>	122,08418	50168,49083419665	50168,49083
<b>Separabilidade</b>	19774987,28611	0,0	82803,31762
<b>Entropia</b>	0,08600	1,0	0,5
<b>Silhueta</b>	0,00594	-0,001	0,4995

*Tabela 1: Resultados*

## Análise Resultados

Analisando os resultados obtidos através das Tabela 1, foi possível levantar os seguintes dados:

**Coesão:** para este método de avaliação a melhor estratégia de agrupamento adotada, foi a kmean que obteve uma coesão de 122,08418. Porém, durante os testes de

melhores parâmetros, notou-se, que quanto maior o número de centróides menor a coesão, em vista disso, conclui se que o método de coesão não é um bom método de avaliação para a base de dados usada.

**Separabilidade:** para este método de avaliação a melhor estratégia de agrupamento adotada, foi a kmean que obteve uma separabilidade de 119774987,28611.

**Entropia:** para este método de avaliação a melhor estratégia de agrupamento adotada, foi a kmean que obteve uma entropia de 0,08600.

**Silhueta:** para este método de avaliação a melhor estratégia de agrupamento adotada, foi a kmean que obteve uma silhueta de 0,00594.

Como é possível ver acima o método de agrupamento mais adequado foi o K-mean, pois obteve o melhor resultado , em todos os métodos de avaliação analisado.