



Aprendizagem de Máquina Trabalho 3

O objetivo do terceiro trabalho da disciplina consiste em comparar o comportamento, em termos de qualidade de separação, de métodos de agrupamentos baseados em diferentes estratégias sobre um mesmo conjunto de dados.

Cada equipe ficará encarregada de uma base de dados distinta. No entanto, todas as bases adotadas terão 1000 instâncias e serão compostas por dois atributos do tipo real. O número de classes entre as bases, no entanto, pode variar.

As bases estão disponíveis no link abaixo. Cada grupo terá um conjunto já definido de acordo com os nomes dos integrantes.

<https://drive.google.com/open?id=1lrJsINC4dHXOu2iOox29Qvtai0qOZhSQ>

O QUE DEVE SER IMPLEMENTADO

Deverão ser implementadas três estratégias de agrupamento (para as abordagens de agrupamento podem ser adotadas implementações disponíveis nas linguagens):

- Baseada em densidade: **DBSCAN**;
- Baseada em protótipo ou centralidade: **K-means**;
- Baseada em hierarquia: **AGNES** (neste caso, será adotada a abordagem aglomerativa com o **Complete Linkage (Max)**).

Para cada uma das estratégias, deve-se encontrar quais são os melhores parâmetros:

DBSCAN

- Tamanho do raio adotado;
- Número mínimo de pontos.

K-means

- Número de centróides;
- Posição inicial dos centróides;
- Número de iterações para convergência.

AGNES

- Não serão adotados parâmetros.

Métodos de Avaliação

Devem ser implementadas (**não deverão ser usadas soluções prontas**) estas abordagens para avaliação da qualidade dos agrupamentos obtidos:

- Coesão;
- Entropia;
- Separabilidade;
- Silhueta.

O QUE DEVE SER ENTREGUE

Relatório detalhado apontando como foram obtidos os melhores parâmetros para cada método.

Discussão de qual método alcançou o melhor agrupamento segundo cada um dos critérios de avaliação de qualidade. Esta análise será feita de forma separada, ou seja, será encontrada a melhor abordagem segundo a coesão, segundo a entropia, segundo a separabilidade e a também para a silhueta.

O relatório deve discutir a razão que pode ter feito com que o método mais adequado foi melhor superior aos demais.

Deve-se discutir também se para ambas as métricas de qualidade o mesmo método foi melhor e, caso não tenha sido, qual seria a razão para tal.

Além do relatório detalhado deve ser entregue o código fonte em que se implementou os métodos de agrupamento e de avaliação da qualidade dos mesmos.

QUANDO

A data de entrega e apresentação dos trabalhos é dia **27/05/2019 às 9:00**.

Nesta aula todas as equipes deverão apresentar os resultados alcançados.

Equipes que não estirem presentes serão penalizadas em 30% da nota do trabalho.