

Projeto Aprendizado de Máquina Descritivo
CMC- 15 Inteligência Artificial
Profa. Ana Carolina Lorena
Trabalho em Grupo de Três ou Quatro Alunos

1. Objetivo

Exercitar e fixar conhecimentos adquiridos sobre Aprendizado de Máquina descritivo, por meio do uso e comparação de técnicas do paradigma de aprendizado não supervisionado em um problema prático.

2. Descrição do Trabalho

O trabalho envolverá analisar dados para prever a presença de meteoros em imagens coletadas de um ponto de observação na cidade de São José dos Campos, uma tarefa originalmente supervisionada. Uma descrição do problema pode ser encontrada em: <https://www.facom.ufu.br/~kddbrcompetition/>

Os dados estão em:

<https://www.kaggle.com/competitions/can-i-make-a-wish-detecting-shooting-stars/overview>

O que deve ser feito:

- a) Cada imagem foi descrita por um conjunto de características extraídas por métodos distintos de processamento de imagens. São 21 conjuntos de características. No primeiro projeto, vocês escolheram cinco desses subconjuntos de treinamento para trabalhar. Neste trabalho, escolha **um** desses subconjuntos para trabalhar, justificando sua escolha.
- b) Realize o agrupamento desses dados com o algoritmo k-médias (excluindo a coluna dos labels). Escolha o valor de k com base no método *elbow*. O valor de k bate com o número de classes original? Discuta.
- c) Realize o agrupamento hierárquico desses dados (excluindo a coluna dos labels) com as métricas de integração *single-linkage*, *average-linkage* e *complete-linkage*. Considere que o número de grupos é 2 e avalie como os algoritmos se saem considerando dos índices de validação: interna por silhueta; e externa por índice rand corrigido (ARI - *adjusted rand index*), neste caso comparando com a partição dada pelos *labels* originais. Faça uma tabela como a seguir e discuta os resultados alcançados.

Métrica de integração	Silhueta	ARI
<i>single-linkage</i>	0.8	0.5

<i>average-linkage</i>
<i>complete-linkage</i>		

- d) Acrescente à tabela anterior os resultados do algoritmo k-médias, usando $k=2$. Discuta os resultados, comparando-os com as variações do algoritmo hierárquico.

Fiquem atentos às eventuais necessidades de pré-processamento, justifiquem suas escolhas no material entregue.

3. Material a ser entregue e prazo

Material: Notebook com as implementações, resultados e discussões

Prazo de Entrega: 04/outubro/2024

Estrutura sugerida:

Nomes dos Membros da Equipe

1. Descrição dos dados escolhidos
2. Resultados do k-médias com o método elbow
3. Resultados de variações do algoritmo hierárquico e do k-médias
4. Discussões
5. Conclusões: Comentários e sugestões sobre o trabalho (complexidade/facilidade, sugestões, etc.).

Bom Trabalho!

Profa. Ana Carolina Lorena

aclorena@ita.br