

Inteligência Artificial

Atividade - Explorando Agrupamento

Prof.

Fernando H. Carvalho Silva

Atividade

1 Atividade

Enunciado

Entrega : De acordo com o Canvas, apresentação em aula (arguição)

Enunciado: Aplicar as técnicas de agrupamento vistas em aula ((bi-secting)
K-Means, Linkage) em duas bases de dados, sendo uma delas a base Iris, a outra base
deve ser escolhida pelo aluno (algumas sugestões de repositórios nos slides adiante)

Enunciado Específico

Especificamente, usando os critérios de avaliação e técnicas de parametrização discutidas em aula:

- 1 Aplicar um algoritmo hierárquico e um particional nas duas bases
- 2 Selecionar melhores parâmetros utilizando técnica de joelho/elbow ou alguma das métricas de avaliação (interna ou externa¹)
- 3 Avaliar grupos gerados
- 4 Descrever e discutir informações obtidas ao aplicar os algoritmos.

¹Não se esqueçam de omitir os rótulos quando aplicarem os algoritmos. Além disso, se usarem os rótulos da base procurem usá-los de forma mais indireta, procurem focar no aspecto não-supervisionado, assumindo que esses rótulos não estariam disponíveis.

Avaliação

Apresentação do código e relatório que contém:

- 1 Algoritmos aplicado às bases de dados selecionadas (4 PONTOS)
- 2 Gráficos gerados pertinentes a cada algoritmo e base de dados (1 PONTO)
- 3 Escolha e justificativa dos parâmetros usados (3 PONTOS)
- 4 Indicar e justificar qual trouxe melhores resultados (2 PONTOS)

Fontes de Bases de Dados

- https://scikit-learn.org/stable/datasets/toy_dataset.html²
- https://scikit-learn.org/stable/datasets/real_world.html³
- <https://www.openml.org/>
- <https://archive.ics.uci.edu/datasets?skip=0&take=10&sort=desc&orderBy=NumHits&search=>
- <https://www.kaggle.com/datasets>

²só tem classificação/regressão, mas podem omitir as classes ou tentar encontrar outros padrões

³só tem classificação/regressão, mas podem omitir as classes ou tentar encontrar outros padrões

Discussões e Reflexões

- A base Iris possui 3 grupos/classes no mundo real, mas a base tem um formato que indica apenas 2 grupos; tentem encontrar a quantidade de grupos em função da base em si, evitando ver os rótulos reais
- Se os rótulos estiverem disponíveis podem fazer parte da análise, mas pensem como uma informação que foi obtida depois do estudo preliminar
- Como o K-means se comporta em relação a um corte no bisecting k-means
- A técnica de joelho permite aplicação de qualquer critério de validação interna, como escolher? O joelho é sensível a escolha do critério de validação, ou não faz muita diferença?

Links e Métodos Úteis

Referência rápida:

- https://scikit-learn.org/stable/datasets/loading_other_datasets.html
- <https://scikit-learn.org/stable/modules/preprocessing.html#normalization>
- <https://scikit-learn.org/stable/modules/decomposition.html#pca>
- <https://plotly.com/python/pca-visualization/>
- <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html#sklearn.cluster.KMeans>

Links e Métodos Úteis

Referência rápida:

- <https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html>
- <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.BisectingKMeans.html>
- <https://scikit-learn.org/stable/modules/clustering.html#overview-of-clustering-methods>