Universidade Federal de São Carlos - Campus Sorocaba
Disciplina: Inteligência Artificial
Professor responsável: Katti Faceli

Trabalho prático

- Data de entrega: 22/01/2019
- Data de avaliação: a partir de 23/01/2019
- Deverá ser feito em grupos de 3 ou 4 alunos
- Os resultados deverão ser descritos e discutidos na forma de um relatório resumido de 2 a 3 páginas (não precisa de capa, só um cabeçalho com as informações relevantes)
- Será feita uma avaliação por meio de entrevista com a presença de todos os membros do grupo
- A data da entrevista será marcada com cada grupo individualmente

Entrega:

- via classrom
- Entregar arquivo .zip nomeado com o nome dos integrantes do grupo contendo:
    - todas as implementações feitas
    - partições produzidas, organizadas por algoritmo
    - planilha com a avaliação da qualidade das partições, gráficos comparativos, etc -
    - relatório resumido no formato pdf

Tarefas:

- Implementar o algoritmo k-medias
    - Entrada: um arquivo texto com o conjunto de dados, k - número de clusters desejado e o número de iterações que o algoritmo deverá executar
    - Saída um arquivo com uma partição do conjunto de dados
- Implementar os algoritmos single-link e average-link
    - Entrada: um arquivo texto com o conjunto de dados, kMin e kMax - intervalo de valores para k (número de clusters) em que serão produzidas partições a partir de cortes no dendrograma
    - Saída um ou mais arquivos, cada um com uma partição do conjunto de dados (dependendo do intervalo fornecido)
- Nos dois casos, tanto para entrada quanto para saída, usar formato conforme conjuntos de dados e partições reais fornecidos
- Aplicar os 3 algoritmos nos 3 conjuntos de dados fornecidos, rodando os algoritmos para produzir partições com os números de clusters indicados em cada caso:
    - c2ds1-2sp.txt --- k entre 2 e 5
    - c2ds3-2g.txt --- k entre 2 e 5
    - monkey.txt --- k entre 5 e 12
- Implementar ou usar uma versão pronta do índice Rand ajustado (AR), conforme a definição dada em https://en.wikipedia.org/wiki/Rand_index

    - atenção para trabalhar com a versão ajustada
- Avaliar a qualidade das partições com o índice Rand ajustado (AR)
    - Para cada conjunto de dados, calcular o índice para todas as partições comparando-as com a partição real correspondente:
        - c2ds1-2spReal.clu
        - c2ds3-2gReal.clu
        - monkeyReal1.clu
    - Anotar resultados em uma planilha identica à resultados disponibilizada
    - Produzir gráfico(s) para ajudar a comparar os resultados
    - Fazer discussão no resumo (incluir dados/gráficos que julgar apropriado)
- Para cada conjunto de dados, escolher melhor partição de cada algoritmo e visualizar os dados mostrando os clusters dessas partições, como ilustrado a seguir
- Discutir os resultados obtidos comparando o desempenho dos algoritmos para os três conjuntos de dados, conforme o(s) tipo(s) de cluster que eles apresentam. Usar o AR e os gráficos