

Disciplina: **Agrupamento de Dados e Aplicações (PCS5782) – 3º Período de 2020.**

Curso: Pós-Graduação em Engenharia Elétrica.

Dia-Horário: Segunda-Feira, 8:30-11:30 (online).

Professor: Eduardo Raul Hruschka.

Planejamento das Atividades:

Aula	Data	Assunto / Atividade
1	14/09	Apresentação da Disciplina (1a) / Introdução (1b), Representação de Dados (2)
2	21/09	Medidas de Proximidade e Conversão de Escalas (2), Métodos Hierárquicos (3)
3	28/09	Métodos Hierárquicos (3)
4	05/10	Métodos Particionais (4), Exercícios, Dúvidas para Prova
	12/10	Feriado
5	19/10	Prova 1
6	26/10	Métodos Particionais (4), Correção da Prova, Vista Prova
	02/11	Feriado
7	09/11	Métodos Particionais DBSCAN + LDA (5)
8	16/11	Validação (6)
9	23/11	Validação (6), Métodos Avançados (7) --- Ensembles, Seleção Atributos, F-EAC --- Dúvidas
10	30/11	Aplicações
11	07/12	Prova 2
12	14/12	Prova Substitutiva (opcional)

Critérios de Avaliação do Aproveitamento:

- A avaliação será composta por duas provas obrigatórias ($P_1 \in [0,10]$ e $P_2 \in [0,10]$);
- As notas obtidas serão consolidadas em uma única nota final, $N_F = (P_1 + P_2)/2$;
- Haverá uma prova opcional, $P_{SUB} \in [0,10]$, envolvendo todo o conteúdo da disciplina, cujo objetivo é substituir a menor entre $\{P_1, P_2\}$. Todos os alunos podem realizá-la, mesmo aqueles que não tenham tido justificativa formal (e.g., atestado médico) por não terem realizado uma das provas.
- Aprovação em função da N_F : **C**[5,7] **B**(7,8.5] **A**(8.5,10] (Frequência $\geq 75\%$).

Objetivos:

Apresentar métodos e algoritmos clássicos para agrupamento de dados amplamente utilizados na prática. Abordar a validação de grupos de dados. Ilustrar aplicações práticas de ferramentas para agrupamento de dados.

Justificativa:

Métodos e algoritmos para agrupamento de dados são tradicionalmente aplicados para abordar diversos problemas práticos, tais como: sumarização de dados, segmentação de mercado, processamento de imagens, categorização automática de textos e documentos etc. Além disso, algoritmos para agrupamento de dados podem servir de base para projetar sistemas inteligentes. Algoritmos para agrupamento de dados têm sido estudados por décadas, mas continuam se constituindo numa área de pesquisa efervescente nos dias atuais, especialmente em áreas do conhecimento que necessitam processar grandes quantidades de dados (e.g., aplicações de big data, predictive analytics e data science). Neste contexto, pode-se afirmar que ferramentas para agrupamento de dados são essenciais para muitas aplicações práticas. Esta disciplina tem por objetivo principal apresentar métodos e algoritmos para agrupamento de dados, bem como várias de suas aplicações práticas importantes. O conhecimento de tais algoritmos agrega ao profissional um diferencial relevante de mercado, em particular para áreas da indústria que necessitam processar grandes quantidades de dados.

Conteúdo:

1. Introdução.
 - 1.1 Motivação.
 - 1.2 Conceitos Básicos.
 - 1.3 Visão Geral.
2. Representação de Dados.
 - 2.1 Tipos de Variáveis.

- 2.2 Índices de Dis(similaridade).
- 2.3 Normalização de Dados.
- 3. Métodos e Algoritmos para Agrupamento de Dados.
 - 3.1 Métodos Hierárquicos.
 - 3.1.1 Métodos Aglomerativos.
 - 3.1.1.1 Algoritmos baseados em vinculação simples, completa e média.
 - 3.1.1.2 Dendrogramas.
 - 3.1.2 Métodos Divisivos.
 - 3.2 Métodos Particionais.
 - 3.2.1 Partições Rígidas.
 - 3.2.1.1 Algoritmo k-means.
 - 3.2.1.2 Algoritmo k-medoids.
 - 3.2.2 Partições com Sobreposição e Modelos de Tópicos (e.g., LDA)
 - 3.2.2.1 Algoritmo EM para Modelos de Misturas de Gaussianas.
 - 3.2.2.2 Algoritmo fuzzy c-means.
- 4. Validação de Grupos de Dados.
 - 4.1 Índices de validade internos, relativos e externos.
 - 4.2 Validade de Estruturas Hierárquicas.
 - 4.3 Validade de Estruturas Particionais.
 - 4.4 Validade de Grupos Individuais.
 - 4.5 Tendência de Agrupamentos.
 - 4.6 Validade de Agrupamentos com Sobreposição.
 - 4.7 Determinação Automática do Número de Grupos.
- 5. Tópicos Avançados.
 - 5.1 Algoritmos Híbridos.
 - 5.2 Algoritmos Baseados em Conjuntos de Partições (e.g., Ensembles, Consensus).
 - 5.3 Seleção de Atributos para Agrupamento de Dados.
- 6. Aplicações.

Bibliografia:

1. XU, R., WUNSCH, D., Clustering, IEEE Press, 2009.
2. GAN, G., Ma, C., WU, J., Data Clustering: Theory, Algorithms, and Applications, SIAM, 2007.
3. KOGAN, J., Introduction to Clustering Large and High-Dimensional Data, Cambridge University Press, 2006.
4. EVERITT, B. S., LANDAU, S., LEESE, M., Cluster Analysis, Hodder Arnold Publication, 2001.
5. ARABIE, P., HUBERT, L. J., DE SOETE, G., Clustering and Classification, World Scientific, 1999.
6. KAUFMAN, ROUSSEAU, **Finding Groups in Data** – An Introduction to Cluster Analysis, Wiley, 1990.
7. JAIN, A. K., DUBES, R. C., **Algorithms for Clustering Data**, Prentice Hall, 1988.
8. Artigos recentes