Título: Agrupamento Fuzzy baseado em Análise Discriminante Linear para Dados com Variáveis Correlacionadas

Orientador: Bruno Almeida Pimentel

Resumo: No contexto da Aprendizagem de Máquina, a análise de agrupamento constitui uma importante abordagem de aprendizado não supervisionado, permitindo identificar padrões e estruturas em conjuntos de dados sem a necessidade de rótulos previamente conhecidos. Entre os métodos existentes, o Fuzzy C-Means (FCM) destaca-se por representar a associação dos objetos aos grupos por meio de graus de pertinência, possibilitando modelar situações em que as fronteiras entre grupos não são claramente definidas [1]. Entretanto, sua formulação clássica utiliza a distância euclidiana, que pode apresentar limitações em conjuntos de dados com variáveis correlacionadas, redundância de informações e grupos com estruturas geométricas complexas. Embora extensões baseadas em matrizes de covariância tenham sido desenvolvidas para lidar com essas características [2][3], desafios relacionados à estabilidade numérica e à alta dimensionalidade ainda motivam novas investigações. Nesse cenário, a Análise Discriminante Linear (Linear Discriminant Analysis – LDA) apresenta potencial para identificar subespaços que favoreçam a separação entre grupos, sendo sua integração com algoritmos de agrupamento uma direção relevante de pesquisa [4][5][6]. Este projeto propõe investigar e desenvolver um método de agrupamento fuzzy baseado em LDA, denominado LDA-FCM, no qual os graus de pertinência estimados pelo FCM sejam utilizados para construir um subespaço discriminante sem recorrer a rótulos externos durante o treinamento. A proposta busca incorporar o tratamento das relações de correlação entre variáveis ao próprio processo de formação dos grupos, considerando aspectos de regularização, estabilidade e eficiência computacional. A avaliação será conduzida por meio de experimentos com dados sintéticos e reais, comparando o método proposto com algoritmos tradicionais de agrupamento fuzzy e técnicas de redução de dimensionalidade. Espera-se contribuir para o desenvolvimento de métodos de agrupamento mais robustos, capazes de explorar estruturas de dependência entre variáveis e ampliar a qualidade e a interpretabilidade dos resultados em aplicações de Ciência de Dados.

Referências Bibliográficas

[1] BEZDEK, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, 1981. https://doi.org/10.1007/978-1-4757-0450-1

[2] GUSTAFSON, D. E.; KESSEL, W. C. Fuzzy clustering with a fuzzy covariance matrix. *IEEE Conference on Decision and Control*, p. 761–766, 1979. https://doi.org/10.1109/CDC.1978.268028

[3] GATH, I.; GEVA, A. B. Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 11, n. 7, p. 773–780, 1989. https://doi.org/10.1109/34.192473

[4] DING, C.; LI, T. Adaptive dimension reduction using discriminant analysis and K-means clustering. *Proceedings of the 24th International Conference on Machine Learning*, p. 521–528, 2007. https://doi.org/10.1145/1273496.1273562

[5] ZHI, X.; FAN, J.; ZHAO, F. Fuzzy Linear Discriminant Analysis-guided maximum entropy fuzzy clustering algorithm. *Pattern Recognition*, v. 46, n. 6, p. 1604–1615, 2013. https://doi.org/10.1016/j.patcog.2012.12.007

[6] WANG, F.; WANG, Q.; NIE, F.; LI, Z.; YU, W.; WANG, R. Unsupervised Linear Discriminant Analysis for Jointly Clustering and Subspace Learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 33, n. 3, p. 1276–1290, 2021. https://doi.org/10.1109/TKDE.2019.2939524