



Um estudo do método de agrupamento espectral Laplacian Eigenmap

Isabella Tito de O. Silva ¹, Deise Mara B. de Almeida ²

RESUMO

O *Laplacian Eigenmap* é um método de *spectral clustering*, uma técnica de agrupamento que utiliza a teoria espectral dos grafos para mapear grupos dentro de um conjunto de dados de acordo com a similaridade entre eles, onde os vértices representam os dados, e os pesos das arestas, a relação de similaridade. O uso deste método ajuda no entendimento sobre o conjunto de dados e possibilita a descoberta de correlações entre os atributos dos dados que não seriam facilmente visualizados sem seu uso. Ademais, o método trabalha com a redução de dimensionalidade do conjunto de dados, dessa forma, as informações contidas nos dados é melhor interpretada. O objetivo deste trabalho é estudar o *Laplacian Eigenmaps* para um conjunto de dados gerados artificialmente de modo a observar a classificação feita pelo método. Os dados artificiais foram criados através da perturbação de pontos, e para o cálculo da similaridade foi utilizado duas equações relacionadas as distâncias entre os pontos. Assim, foi possível observar o mapeamento de conjunto de pontos, fornecida pelo método, de acordo os grupos de pontos existentes dentro de cada conjunto e da proximidade entre os grupos.

Palavras-chave: agrupamento, redução de dimensionalidade, similaridade.

¹ Aluna do curso de Licenciatura em Matemática, Unidade Acadêmica de Matemática, UFCG, Campina Grande, PB, e-mail: isabella.tito@estudante.ufcg.edu.br

² Doutora, Professora, Unidade Acadêmica de Matemática, UFCG, Campina Grande, PB, e-mail: deise@mat.ufcg.edu.br

Um estudo do método de agrupamento espectral Laplacian Eigenmap

ABSTRACT

Laplacian Eigenmap is a method of spectral clustering, a clustering technique that uses the spectral theory of graphs to map groups within a dataset according to the similarity between them, where the vertices represent the data, and the edge weights represent the similarity relationship. The use of this method helps in the understanding of the dataset, and makes it possible to discover the correlations between the attributes of the data, which wouldn't be easily visualized without its use. Furthermore, this method works with the reduction of the dimensionality of the dataset; in this way, the information in the data can be better interpreted. The objective of this work is to study the *Laplacian Eigenmaps* for an artificially-generated dataset, to observe the classification made by this method. The artificial data was created by a perturbation of points. To calculate the similarity, two equations related to the distances between the points were used. Thus, it was possible to observe the mapping of the set of points provided by the method, according to the groups of points existing within each set, and the proximity between the groups.

Keywords: map, reduction of dimensionality, similarity.