SCC5836/SCC0252
Visualização Computacional
M. Cristina 2023

LISTA DE EXERCÍCIOS 5

Visualização por redução de dimensionalidade

1. Explique qual é a relação entre uma técnica de redução de dimensionalidade e uma técnica de projeção multidimensional.

2. Qual é a entrada e qual é a saída de uma técnica de projeção multidimensional? Explique.

3. Explique porque a definição de uma função de dissimilaridade no espaço multidimensional é um requisito necessário para aplicar uma técnica de projeção multidimensional a um conjunto de dados. Dê exemplos de funções de dissimilaridade que podem ser aplicadas nesse contexto.

4. Descreva como funciona o PCA (textualmente, e depois matematicamente).

5. Que propriedades um conjunto de dados multidimensional deve apresentar para que o PCA seja uma alternativa interessante para redução de dimensionalidade?

6. Em que situação a transformação pelo PCA não se mostra adequada como estratégia para a visualização dos dados por meio de gráficos de dispersão 2D ou 3D das componentes principais (CPs)? Existem estratégias alternativas que poderiam ser utilizadas para visualizar os dados por meio das suas CPs? Discuta.

7. Dada a matriz de dados [2, 0, 3, -1; 0, -2, -3, 1], calcule a variância após projetar os dados no seu primeiro CP.

8. Se os autovetores e autovalores da matriz de covariância de um conjunto dados 2D são como informado abaixo, pergunta-se: (i) em termos percentuais, quanto da variância desses dados é explicada pelo PC1 e pelo PC2, respectivamente? (ii) Descreva como obter a projeção dos dados originais na PC1.

$$v1 = \begin{bmatrix} 0.6778736 \\ 0.7351785 \end{bmatrix} \qquad \lambda_1 = 1.284028$$

$$v2 = \begin{bmatrix} -0.7351785 \\ 0.6778736 \end{bmatrix} \qquad \lambda_2 = 0.04908323$$

9. Explique que informação é apresentada em um *scree plot* associado a uma decomposição por PCA.

10. Explique no que consistem os *loading factors* no PCA, e qual a sua interpretação, i.e., o que informam sobre os dados no contexto do PCA.

11. Explique como funciona a técnica de projeção multidimensional MDS. Qual é

função de stress a ser minimizada, e o que ela mede?

12. Explique (i) porque o MDS é uma técnica de custo computacional alto; (ii) porque é uma técnica de preservação global.

13. Explique a diferença entre uma técnica de projeção *paramétrica* vs. *não paramétrica*.

14. Explique a diferença entre uma técnica de projeção *global* vs. *local*.

15. Explique a diferença entre uma técnica de projeção *supervisionada* vs. *não supervisionada*.

16. Explique a diferença entre uma técnica de projeção *estocástica* vs. *não estocástica*.

17. Descreva qual a função de erro otimizada na t-SNE, e o que ela mede.

18. Descreva quais os parâmetros da t-SNE e qual a interpretação de cada um.

19. Diz-se que a t-SNE é uma técnica que prioriza a preservação de vizinhanças. Explique o que isso significa. Discuta quais as implicações dessa propriedade ao interpretar uma visualização que exibe um gráfico de pontos 2D ou 3D com o resultado da t-SNE.

20. Descreva qual a função de erro otimizada na UMAP, e o que ela mede.

21. Descreva quais os parâmetros da UMAP e qual a interpretação de cada um.

22. Quais as vantagens da UMAP sobre a t-SNE?

23. Discuta possíveis maneiras de avaliar a qualidade do resultado de uma projeção multidimensional aplicada um conjunto de dados, considerando um conjunto de dados rotulado (i.e., instâncias têm classes).

24. Discuta possíveis maneiras de avaliar a qualidade do resultado de uma projeção multidimensional aplicada um conjunto de dados, considerando um conjunto de dados não rotulado (i.e., instâncias não têm classes).