

Análise de dados de universidades via PCA

Atividade para as disciplinas SCC0252 e SCC5836

Esta atividade consiste em analisar dados das 800 universidades ranqueadas na "Times World University Ranking" utilizando PCA (Principal Component Analysis).

Instruções:

Para obter um resultado satisfatório para esta atividade, sugerimos:

1. Acessar os dados da *Times World University Ranking* em [1].
2. Utilizar o roteiro para realização do PCA do script em [2].
3. Prestar atenção para as etapas de análise descritas no script anterior, de forma a obter informações relevantes a partir dos dados.
4. Expressar suas conclusões na forma de um relatório em que conste observações sobre cada etapa da análise e visualizações dos dados na forma de tabelas e de scatter plots.
5. Entregar junto ao relatório os scripts finais utilizados. Tente manter o código organizado e bem comentado.

Para abrir os dados em [1] de forma conveniente para utilizar em um script Python, recomendamos o procedimento em [3]. A linguagem de programação utilizada para realizar a análise é de livre escolha do aluno. Decida o que fazer com os dados, considere: os campos faltantes, os campos não numéricos, usar nomes e países?

Extra:

- Visualizações em 3 dimensões.
- Conclusões a respeito de componentes principais além das duas primeiras.
- Análises dos dados com recursos além dos disponibilizados em [2] (e.g. classificação).
- Script com uma transcrição da rotina em [2] para outra linguagem de programação.

Documentação de suporte:

- PDF com descrição do PCA [4].
- Vídeo com apresentação do conteúdo dado em aula [5].

Links:

[1]:

[https://github.com/ttm/dataVisualization/raw/master/data/Times%20World%20University%20Rankings%20\(2016\).xlsx](https://github.com/ttm/dataVisualization/raw/master/data/Times%20World%20University%20Rankings%20(2016).xlsx)

[2]: <https://github.com/ttm/dataVisualization/raw/master/src/pca.py>

[3]: <https://github.com/ttm/dataVisualization/raw/master/src/abreXlsx.py>

[4]: <https://github.com/ttm/dataVisualization/raw/master/latex/pca.pdf>

[5]: <https://www.youtube.com/watch?v=.....>