

Algoritmos para Big Data

2021/2022

Exercício (simples)

Tendo como base os notebooks fornecidos nas aulas anteriores, implemente um modelo de agrupamento (*clustering*) de acordo com os seguintes requisitos:

- Utilização do algoritmo k-means que está disponível na biblioteca Apache Spark MLlib.
- Adoção da *pipeline* de aprendizagem automática como princípio orientador.
- Utilização do ficheiro de dados fornecido, de pequena dimensão.

Refira-se que este exercício tem como objetivo principal obter experiência na adaptação de trabalho realizado anteriormente para novos contextos, embora similares. Por isso, é fornecido um ficheiro de dados de pequena dimensão, bastante utilizado em ambiente académico.