Métodos Matriciais e Análise de Clusters

Laura de Oliveira F. Moraes

Olá!

Eu sou Laura Moraes

Graduada em Engenharia Eletrônica

Mestre em Engenharia de Computação

Doutoranda em Inteligência Artificial

Engenheira de Computação no CERN por 4 anos

Co-fundadora da start-up Twist Systems

Imersão de 3 meses no programa Data Science for Social Good

Programa

1. Modelagem dos dados
2. Cálculo de similaridade
3. Redução de dimensionalidade
4. Clusterização
5. Avaliação de clusters

Avaliação

O curso será avaliado através de um **trabalho** e uma **prova** individuais baseada no trabalho.

- O trabalho vale 30% da nota
- A prova vale 70% da nota

Literatura Complementar

1. (O mais técnico) James, Gareth, et al. An Introduction to Statistical Learning: with Applications in R. Springer, 2017.
2. (Misto entre técnico e alto-nível) Provost, Foster, and Tom Fawcett. Data Science para Negócios: O que você precisa saber sobre mineração de dados e pensamento analítico de dados. Alta Books, 2016.
3. (Para colar como faz) Wickham, Hadley, and Garrett Grolemund. R For Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly, 2017.
4. (Pensar com estatística) Huff, Darrell, and Irving Geis. How to Lie with Statistics. Intrínseca, 2016.
5. (Em português) Faceli, K., & Lorena, A., & Gama, J. & de Carvalho, André. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. Editora LTC, 2011
6. (Repositório Online) https://github.com/rsouza/FGV_Intro_DS

Plano de Aula - aula 01 - 18/05/2019

FGV | EDUCAÇÃO EXECUTIVA

- **Dados**
  - Por que tomar decisões baseadas em dados? [Livro 2, Caps. 1 e 2]
  - Como representar as informações? [Livro 5, Caps. 2 e 3]
    - Modelagem
    - Vetor, matriz e transformações lineares
    - Pré-processamento: normalização, limpeza e dados faltantes
    - Representação de dados categóricos
- **Similaridade**
  - Conceito de similaridade: [Livro 2, Cap. 6; Livro 5, Caps. 2 e 11]
    - Matriz de covariância
    - Métricas para cálculo de similaridade
- **Redução de dimensionalidade** [Livro 1, Cap. 10]
  - Similaridade em alta dimensão
  - Decomposição da base vetorial:
    - Teorema espectral
    - SVD e PCA
    - Visualização

Plano de Aula - aula 02 - 01/06/2019

FGV | EDUCAÇÃO EXECUTIVA

- **Clusterização** - [Livro 1, Caps. 2 e 10; Livro 2, Cap. 2]
  - O que é e por que agrupar?
    - Diferença entre aprendizado supervisionado x não-supervisionado
    - Desafios do aprendizado não-supervisionado
  - Técnicas de clusterização - [Livro 1, Cap. 10; Livro 2, Cap. 6; Livro 5, Cap. 12]
    - K-Means
    - Agrupamento hierárquico
    - Vizinhos mais próximos
    - Agrupamento espectral
- **Avaliação** - [Livro 2, Cap. 6; Livro 5, Cap. 14]
  - O que define uma boa clusterização?
    - Distância intra-cluster e inter-cluster
    - Silhueta, índice Calinski-Harabaz e índice Davies-Bouldin
    - Estabilidade
  - Compreendendo os resultados da clusterização