

Introdução a Ciências de Dados

Programação do Curso

Francisco A. Rodrigues
ICMC/USP
francisco@icmc.usp.br



Aula 1

Introdução

e

Ciências de Dados e suas Etapas

Tópicos

- O que é Ciência de Dados
- Problemas e Soluções em Ciência de Dados
- Tipos de dados.
- Estatística descritiva.
- Visualização

Leitura Complementar

- Tan, Steinbach, Karpatne, Kumar, **Introduction to Data Mining**, Pearson, 2013 (capítulo 1)
- Chen, Härdle, Unwin, **Handbook of Data Visualization**, Springer.

Aula 2

Selecionando Métodos e Ajustando Modelos

Tópicos

- Modelos preditivos
- Overfitting
- Bias-variance tradeoff
- Escolhendo modelos

Leitura Complementar

- Bishop, **Pattern Recognition and Machine Learning**, Springer (capítulo 1).
- Online:
<https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>
- Lindholm et al., **Supervised Machine Learning**, 2019.
http://www.it.uu.se/edu/course/homepage/sml/literature/lecture_notes.pdf

Aula 3

Tratamento e transformação de dados

Tópicos

- Técnicas para Tratamento e Transformação de Dados.
- Normalização
- Análise dos componentes principais.

Leitura Complementar

- **Inteligência Artificial: Uma abordagem por aprendizado de máquina**, Facelli, Lorena, Gamma e Carvalho, LTC. (Capítulos 2 e 3)
- **PCA: A Tutorial on Principal Component Analysis**, Jonathon Shlens, 2014
<https://arxiv.org/abs/1404.1100>

Aula 4

Técnicas de Agrupamento de dados

Tópicos

- K-means,
- Agrupamento Hierárquico,
- Avaliando Agrupamentos

Leitura Complementar

- **Inteligência Artificial: Uma abordagem por aprendizado de máquina**, Facelli, Lorena, Gamma e Carvalho, LTC. (Capítulos 11 e 12)
- Tan, Steinbach, Karpatne, Kumar, **Introduction to Data Mining**, Pearson, 2013 (capítulo 8)

Aula 5

Modelos de regressão

Tópicos

- Regressão Linear,
- Simplificando Modelos via Regularização,
- Avaliando e Interpretando modelos.

Leitura Complementar

- Lindholm et al., **Supervised Machine Learning**, 2019.
http://www.it.uu.se/edu/course/homepage/sml/literature/lecture_notes.pdf
- James et al., **Introduction to statistical learning with applications in R**, 2014.
<https://www.ime.unicamp.br/~dias/Intoduction%20to%20Statistical%20Learning.pdf>

Aula 6

Classificação

Tópicos

- k-vizinhos mais próximos,
- Regressão Logística,
- Naive Bayes.

Leitura Complementar

- Lindholm et al., **Supervised Machine Learning**, 2019.
http://www.it.uu.se/edu/course/homepage/sml/literature/lecture_notes.pdf
- James et al., **Introduction to statistical learning with applications in R**, 2014.
<https://www.ime.unicamp.br/~dias/Intoduction%20to%20Statistical%20Learning.pdf>

Aula 7

Classificação: árvores, ensembles

Tópicos

- Árvores de Decisão
- Ensemble methods

Leitura Complementar

- Lindholm et al., **Supervised Machine Learning**, 2019.
http://www.it.uu.se/edu/course/homepage/sml/literature/lecture_note_s.pdf
- Tan, Steinbach, Karpatne, Kumar, **Introduction to Data Mining**, Pearson, 2013

Aula 8

Classificação: SVM, Avaliação de modelos e Projetos em Ciência de Dados

Tópicos

- Support Vector Machines
- Avaliando Modelos de Classificação
- Projeto de Ciência de Dados.
- Comunicação de Resultados.
- Questões Éticas em Ciência de dados
- Exemplo de um projeto.

Leitura Complementar

- Tan, Steinbach, Karpatne, Kumar, **Introduction to Data Mining**, Pearson, 2013