

PROGRAMA

1. Métodos inferenciais computacionalmente intensivos. Testes baseados em simulação e aleatorização. Bootstrap e jackknife. Algoritmo Expectation-Maximization (EM). Métodos Monte Carlo via Cadeias de Markov (MCMC).

2. Técnicas de redução de dimensionalidade. Análise de componentes principais. Análise de correspondência múltipla. Análise fatorial. Escalonamento Multidimensional.

3. Técnicas de agrupamento. Medidas de distância e similaridade. Agrupamento Hierárquico. K-médias. Agrupamento baseado em densidade e agrupamentos baseados em modelos. Métricas de qualidade de agrupamentos.

4. Modelos de regressão. Regressão linear simples e múltipla. Modelos Lineares Generalizados. Seleção de variáveis. Diagnóstico.

5. Fundamentos de aprendizado supervisionado. Funções de perda e estratégias de ajuste para classificação e regressão. Vício e Variância. Regularização. Métodos de Validação. Medidas de capacidade preditiva.

6. Métodos estatísticos supervisionados. Análise discriminante. Classificadores bayesianos. Métodos baseados em árvores de decisão. Máquina de vetores de suporte.

7. Modelos de Inteligência Artificial. Redes neurais. Perceptron e multilayer perceptron. Aprendizado profundo. Redes neurais recorrentes. Generative adversarial networks (GANs).

8. Big Data. Métodos de subsampling e reamostragem para grandes volumes de dados. Divide-and-conquer. mapReduce e paralelização. Métodos estatísticos para alta dimensionalidade. Plataformas de dados para computação em Nuvem.

BIBLIOGRAFIA

EFRON, B.; HASTIE, T. *Computer age statistical inference: Algorithms, evidence, and data science*. Cambridge: Cambridge University Press, 2016.

FARAWAY, J. J. *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. Boca Raton: CRC Press, Taylor & Francis, 2005.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge: MIT Press, 2016.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: Data mining, inference, and prediction*. 2. ed. New York: Springer, 2013.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An introduction to statistical learning: With applications in R*. 2. ed. New York: Springer, 2021.

JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 6. ed. Upper Saddle River: Pearson, 2007.

LURASCHI, J.; KUO, K.; RUIZ, E. *Mastering Spark with R: The complete guide to large-scale analysis and modeling*. Sebastopol: O'Reilly Media, 2019.

MARSLAND, S. *Machine learning: An algorithmic perspective*. 2. ed. Boca Raton: CRC Press, 2011.

MORETTIN, P. A.; SINGER, J. M. *Estatística e Ciência de Dados*. Rio de Janeiro: LTC, 2022.

PRAJAPATI, V. *Big data analytics with R and Hadoop*. Birmingham: Packt Publishing Ltd, 2013.

RIZZO, M. L. *Statistical computing with R*. Boca Raton: Chapman and Hall/CRC, 2019.

ROSS, S. M. *Simulation*. 4. ed. Amsterdam: Elsevier Academic Press, 2006.

WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J. *Data mining: Practical machine learning tools and techniques*. 4. ed. San Francisco: Morgan Kaufmann, 2016.