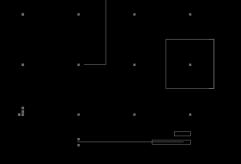
THE SCIENCE SCIENCE

MACHINE LEARNING E DATA MINING



BIG DATA SCIENCE

REFERÊNCIAS

DIÓGENES JUSTO

SLIDER I



Stanford

Minning Massive Datasets

Machine Learning

Introdução a Ciência de Dados

Especialização Machine Learning:

(1) Fundamentos (2) Regressão (3) Classificação (4)Cluster (5) Sistemas de recomendação e redução de dimensão (6) Projeto (Capstone)



+



Especialização Data Mining:

(1) Padrões (2) Texto e Buscadores (3) Cluster (4) Text Mining e Visualização de dados

Especialização Ciência de Dados:

(1) Ferramentas (2) R (3) Extração e Limpeza (4)
 Análise Exploratória (5) Pesquisa Reproduzível (6)
 Inferência Estatística (7) Modelos Regressivos (8)
 Aprendizagem Automática (Machine Learning) (9)
 Desenvolvendo Aplicação

Especialização Executiva em Data Science

(1) Introdução (2) Equipe (3) Gestão (4) Problemas Reais (5) Projeto (Capstone)



StanfordMinning MassiveDatasets

- MapReduce (modelo programação Hadoop)
- Uso de PageRank, grafos, medidas de similaridade
- Clustering, sistemas de recomendação
- Técnicas para grandes volumes de dados: redução de dimensão, data stream, LSH, SVM
- Boa parte do conteúdo em mmds.org (inclusive um livro)

Stanford Machine Learning

- Forte embasamento matemático
- Uso de equações diferenciais eálgebra matricial (matemática aplicada) para método de mínimos quadrados
- Todos exercícios com uso de MatLab/Octave
- Transcrição das aplicações matemáticas em algoritmos

W

UNIVERSITY of WASHINGTON Introdução a Ciência de Dados

- Foco em programação
- Fornece uma máquina virtual
- Sentiment Analysis Python (twitter's crawler)
- MapReduce: implementação em Python
- Competição Kaggle
- NoSQL: Pig (sobre Hadoop), utilizando AWS



Especialização Mining Data

- Embasamento matemático e estatístico
- Ferramenta específica (Meta C++) para Text Mining
- Aplicações:
 - 1. Ex.: Fralda + Cerveja
 - 2. Ex.: Google rank (progr.)
 - 3. Ex.: Vários tipos de cluster

DESCOBRINDO PADRÕES APLICÁVEIS NA MINERAÇÃO DE DADOS

RECUPERAÇÃO DE TEXTO E MOTORE DE BUSCA ANÁLISE DE CLUSTER EM DATA MINING

MINERAÇÃO DE TEXTO E ANÁLISE ESTATÍSTICA VISUALIZAÇÃO DE DADOS







Especialização Data Science

- Todos os módulos com uso de R
 - 1. Overview R, técnicas
 - 2. Uso do R
 - 3. Extração com R
 - 4. Gráficos, descrição dados
 - 5. R. Markdown
 - Estatística clássica
 - 7. Regressão linear e variantes
 - 8. Básico de Machine Learning
 - 9. Aplicações com Rerver/Slices

Todo material em: github.com/DataScienceSpecialization/



AS FERRAMENTAS DO CIENTISTA DE DADOS

LINGUAGEM R

OBTENÇÃO E LIMPEZA DE DADOS

ANÁLISE EXPLORATÓRIA DE DADOS

PESQUISA REPRODUTÍVEL INFERÊNCIA ESTATÍSTICA

MODELOS REGRESSIVOS APRENDIZAGEM AUTOMÁTICA NA PRÁTICA

DESENVOLVIMEN TO DE DADOS

CURSOS NA PLATAFORMA

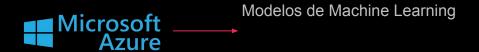


- Aplicações
- Cloud, Databases
- Plataformas distribuídas
- NoSQL, Fast Algorithms
- Data Compression
- Overview BigData
- Spark, pySpark
- Spark MLLib

Introdução a Big Data com Apache Spark



PLATAFORMAS CLOUD





R, Python, Scala, Spark integrados



Python, Octave (Matlab)

Estatístia Aplicada:
 Levine, Stepah, Krehbiel, Berenson, ESTATÍSTICA TEORIA
 E APLICAÇÕES (com uso do Excel). 5.a edição, LTC.

BIBLIOGRAFIA

- Aplicações em Business Intelligence e Big Data:
 Dean, Jared, BIG DATA, DATA MINING AND MACHINE LEARNING. Wiley.
- Data Mining:
 Han, Kamber, DATA MINING– Concepts and Techniques, 2.a edição, Morgan Kaufmann.
- Data Mining e Big Data:
 Leskovec, Rajaraman, Ullman, MINING OF MASSIVE DATABASES. 2014, http://mmds.org.

FÁVERO, L. P. L, et al. *Análise de Dados – Modelagem Multivariada para Tomada de Decisão*, São Paulo, Campus/Elsevier, 2009.

BIBLIOGRAFIA

Free books

- COGHLAN, Avril. A Little Book of R for Multivariate Analysis (free book)
 Disponivel em: http://bit.ly/1PTsN48, 2014.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. The Elements of Statistical Learning, 2nd Ed. (free book), Disponível em: http://stanford.io/20SfKqL, 2008
- JAMES, Gareth, et al. *An Introduction to Statistical Learning with applications in R (free book)*, Disponível em: http://bit.ly/Grdeen

Obrigado



ProfDiogenes.Justo@fiap.com.br



Copyright © 2018 | Diógenes Justo

Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente proibido sem consentimento formal, por escrito, do professor/autor.