
CURSO: Mestrado em Modelagem Matemática
2º trimestre de 2020/TURMA 2020
DISCIPLINA: **Fundamentos de Data Science**
PROFESSOR: **Jorge Luís Poco Medina**
CARGA HORÁRIA: 45 h
CLASSIFICAÇÃO: Eletiva
HORÁRIO E SALA DE ATENDIMENTO:
SALA:

PLANO DE ENSINO

1. Ementa

Conceitos de Dados, Pré-Processamento de Dados: Limpeza, Avaliação de Outlier, Transformação de Dados, Redução, Análise e Seleção de Variáveis; Representação do Conhecimento, Classificação de Dados a partir de Algoritmos de Árvores de Decisão, Baseados em Instâncias; Clusterização por partição, hierárquica; emprego de Redes Neurais, Algoritmos Genéticos, Lógica Fuzzy e Sistemas Híbridos em MD; Softwares e Ferramentas de MD; Domínios de Aplicação e Estudos de Casos.

2. Objetivos da disciplina

O curso se concentra na análise de dados confusos da vida real para realizar previsões usando métodos estatísticos e de aprendizado de máquina.

3. Procedimentos de ensino (metodologia)

Aulas online

4. Conteúdo programático detalhado

- Semana 1: Introdução
- Semana 2: Exploração de dados, Pandas e Web scrapping
- Semana 3: Introdução à regressão e regressão linear
- Semana 4: Regressão linear múltipla, regressão polinomial e regularização de modelos
- Semana 5: Regularização, EDA, visualização para comunicação
- Semana 6: Regressão logística
- Semana 7: Classificação e PCA
- Semana 8: Árvores de decisão, bagging, e floresta aleatória
- Semana 9: Boosting, Redes Neurais (RN) 1, Perceptron e MLP
- Semana 10: RN 2 e 3, Anatomia da NN e Back Propagation
- Semana 11: RN 4, Métodos de regularização,

- Semana 12: Visualização para interpretação de modelo
- Semana 13: Apresentação do Projeto

5. Procedimentos de avaliação

- **Trabalhos: 50%**, avaliado em sua submissão individual.
- **Projeto: 40%**, avaliado de acordo com os critérios do projeto e sua avaliação por pares.
- **Participação: 10%** avaliados sobre a participação na Piazza e (para estudantes no campus) na frequência da seção.
- **Melhores Projetos:** os três principais projetos que receberão pontos extras.

6. Bibliografia Obrigatória

- James, Witten, Hastie, Tibshirani. An Introduction to Statistical Learning.
- Ballard, D. H. (1999). Introduction to Natural Computation. MIT Press.
- Goldschmidt, R., & Passos, E. (2005). Data Mining. Campus.
- Han, J., Kamber, M., & Pei, J. (2000). Data Mining: Concepts and Techniques. Morgan Kaufmann.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Addison Wesley.
- Witten, I. H., Frank, E., & Hall, M. A. (2000). Data Mining. Morgan Kauffmann.

7. Minicurrículo do Professor

Jorge Luís Poco Medina - I am an associate professor in the School of Applied Mathematics of Fundação Getulio Vargas (FGV-EMAp). Previously I was an assistant professor in the Research and Innovation Center in Computer Science (RICS) at the San Pablo Catholic University (UCSP) and a research associate in the UW Interactive Data Lab (IDL) at the University of Washington working with Jeffrey Heer. I obtained my PhD from the NYU Polytechnic School of Engineering in 2015 under the supervision of Claudio Silva. Before going to NYU I spent a year in the Scientific Computing and Imaging Institute (SCI) at the University of Utah (UoU). I have a M.S. in Computer Science from the Instituto de Ciências Matemáticas e de Computação (ICMC) at the University of São Paulo (USP), Brasil-2010, and a B.E. in System Engineering from the National University of San Agustín (UNSA), Peru-2008. As part of my professional life I worked in zAgile Inc as a software engineer on 2008. Ive done internships at Google Inc. (2008 and 2010), Kitware Inc (2011), Oak Ridge National Laboratory (2012) and Xerox Research (2013). My research has focused on data visualization. I have participated in projects on information visualization, scientific visualization, and visual analytics. I was also involved in interdisciplinary collaborations that focused on the development of novel visualization methods to enable both climate and urban data analysis.

8. Link para o Currículo Lattes

<http://lattes.cnpq.br/1565709349354834>