



PUC Minas

Pós-graduação Ciência de Dados e Big Data Aprendizado de Máquina

Sérgio M. Dias

2º semestre de 2010

Apresentação

- Disciplina
 - AM- Aprendizado de Máquina
- Datas:

Curso: CIÊNCIA DE DADOS E BIG DATA

Oferta: 2

Unidade: Contagem

Disciplinas	Cód.	CH	Professor(a) – NOME COMPLETO
Técnicas Estatísticas de predição: Teoria e Aplicações	AP	24	Danilo Costa
Processamento de Linguagem natural	PLN	24	Geanderson Esteves dos Santos
Integração e Fluxo de Dados	ETL	24	Leonardo Maurício Cerqueira
Machine Learning	ML	24	Sérgio Mariano Dias

Agosto/2019

Horário/ Data	Sab.	Sab.	Sab.	Sab.	Sab.
08:00 - 08:50	AP	AP	AP	AP	PLN
08:50 - 09:40	AP	AP	AP	AP	PLN
09:50 - 10:40	AP	AP	AP	AP	PLN
10:40 - 11:30	AP	AP	AP	AP	PLN
11:40 - 12:30	AP	AP	AP	AP	PLN
12:30 - 13:20	AP	AP	AP	AP	PLN

Setembro/2019

Horário/ Data	Sab.	Sab.	Sab.	Sab.
08:00 - 08:50	PLN	PLN	ETL	
08:50 - 09:40	PLN	PLN	ETL	
09:50 - 10:40	PLN	PLN	ETL	
10:40 - 11:30	PLN	PLN	ETL	
11:40 - 12:30	PLN	PLN	ETL	
12:30 - 13:20	PLN	PLN	ETL	

Outubro/2019

Horário/ Data	Sab.	Sab.	Sab.	Sab.
08:00 - 08:50	PLN		ETL	ETL
08:50 - 09:40	PLN		ETL	ETL
09:50 - 10:40	PLN		ETL	ETL
10:40 - 11:30	PLN		ETL	ETL
11:40 - 12:30	PLN		ETL	ETL
12:30 - 13:20	PLN		ETL	ETL

Novembro/2019

Horário/ Data	Sab.	Sab.	Sab.	Sab.	Sab.
08:00 - 08:50		ETL		ML	ML
08:50 - 09:40		ETL		ML	ML
09:50 - 10:40		ETL		ML	ML
10:40 - 11:30		ETL		ML	ML
11:40 - 12:30		ETL		ML	ML
12:30 - 13:20		ETL		ML	ML

Dezembro/2019

Horário/ Data	Sab.	Sab.
08:00 - 08:50	ML	ML
08:50 - 09:40	ML	ML
09:50 - 10:40	ML	ML
10:40 - 11:30	ML	ML
11:40 - 12:30	ML	ML
12:30 - 13:20	ML	ML

Ementa

- AM - Aprendizado de Máquina (NC)
 - Metodologia para descoberta de conhecimento em banco de dados. Exploração do espaço problema e espaço solução. Técnicas de aprendizado supervisionado e não-supervisionado. Regras de associação, agrupamento (clustering) e classificação. Rede neural, Agrupamento com K Means. Classificador Naïve Bayesian. Árvore de decisão. Outros algoritmos

Ementa

- **Principais tópicos:**
 - Metodologia para descoberta de conhecimento
 - Técnicas de aprendizado supervisionado e não-supervisionado.
 - Regressão
 - Associação
 - Agrupamento
 - Classificação
- **Atividades prática**
 - R
 - Knime

Bibliografia

- Marsland, Stephen. Machine learning: an algorithmic perspective. CRC press, 2015.
- Zaki, Mohammed J., Wagner Meira Jr, and Wagner Meira. Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, 2014.



As bibliografias complementares são:

- Han, Jiawei, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, 2011.
- da Silva, Leandro Augusto, Sarajane Marques Peres, and Clodis Boscarioli. Introdução à mineração de dados: com aplicações em R. Elsevier Brasil, 2017.
- Braga, LUIS PAULO VIEIRA BRAGA. Introdução à Mineração de Dados-2a edição: Edição ampliada e revisada. Editora E-papers, 2005.
- Amaral, Fernando. Introdução à Ciência de Dados: mineração de dados e big data. Alta Books Editora, 2016.
- Barber, David. Bayesian reasoning and machine learning. Cambridge University Press, 2012.



As bibliografias complementares são:

- O curso também será baseado em artigos

Aprovação

- Frequência
 - 60%
- Avaliações
 - 100%
 - Trabalho prático

Trabalho – Soluções tecnológicas para aprendizado de máquina

- Objetivos:
 - Identificar as principais soluções livres e proprietárias para aprendizado de máquina
 - Proporcionar ao aluno o uso de diferentes soluções tecnológicas
- Descrição:
 - 1 - Identificar, no mínimo seis, soluções (bibliotecas, linguagens e ferramentas de alto nível) para aplicação de técnicas de aprendizado de máquina
 - 2 - Escolher um exemplo (“toy example”) em regressão, associação, agrupamento ou classificação e demonstrar o uso em cada uma das soluções identificadas no item 1
 - Apesar de ser uma base de exemplo, procure dados que necessitem passar por um processo de adequação dos dados.

Trabalho – Soluções tecnológicas para aprendizado de máquina

- Objetivos:
 - Identificar as principais soluções livres e proprietárias para aprendizado de máquina
 - Proporcionar ao aluno o uso de diferentes soluções tecnológicas
- Descrição:
 - 3 – Escrever um relatório técnico, em formato de artigo, apresentando as vantagens e desvantagens de cada solução livre. Considerar que nas organizações existem diferentes perfis de profissionais que atuam em ciência de dados. Ademais, refletir a respeito da viabilidade computacional de cada solução.
 - <http://www.sbc.org.br/documentos-da-sbc/category/169-templates-para-artigos-e-capitulos-de-livros>

Trabalho – Soluções tecnológicas para aprendizado de máquina

- Trabalho individual
 - 100% da nota total
 - 20% para complexidade da base escolhida
 - 40% para as ferramentas selecionadas
 - 40% para a qualidade do relatório entregue
- Entrega
 - Relatório técnico conforme template por e-mail
 - sergiomariano@gmail.com
 - Assunto:
 - “[pos-pucminas-AM] título do relatório”
 - Data limite: Sábado, dia 21 de Dezembro de 2019 até as 23:59
 - **Atenção:** Para dia atraso o trabalho será penalizado em 2 pontos.

Recursos

- UCI Machine Learning Repository
 - archive.ics.uci.edu/ml/
- Ferramentas básicas
 - The R Project for Statistical Computing
 - <https://www.r-project.org/>
 - Rstudio
 - <https://www.rstudio.com/>
 - Knime
 - <https://www.knime.com/>

Recursos

- Ferramentas adicionais
 - Rapidminer
 - <https://rapidminer.com/>
 - Azure machine learning
 - <https://azure.microsoft.com/en-us/services/machine-learning-studio/>
 - Cloud Machine Learning - google.com
 - <https://cloud.google.com/ml-engine>

Recursos

- Recursos na web
 - Machine Learning, Data Science, Big Data, Analytics, AI
 - <https://www.kdnuggets.com/>
 - The Home of Data Science & Machine Learning
 - <https://www.kaggle.com/>
 - https://pt.wikibooks.org/wiki/Processamento_de_Dados_Massivos
 - <http://posgraduando.com/>
 - <http://phdcomics.com/>



PUC Minas

Obrigado pela atenção!

Sérgio M. Dias | www.sergiomdias.com

Professor e Pesquisador | PUC Minas

<http://www.icei.pucminas.br/projetos/dsrgroup/>

Doutor em Ciência da Computação | UFMG

www.dcc.ufmg.br

Analista Sênior - Cientista de dados | SERPRO

www.serpro.gov.br

Contato: sergiomariano@gmail.com