



PUC Minas

Pós-graduação Ciência de Dados e Big Data

Aprendizado de Máquina

Sérgio M. Dias

1º semestre de 2018

Apresentação

- **Disciplina**

- AM- Aprendizado de Máquina

- **Datas:**

Junho / 2018					
Dia da Semana	Sab.	Sab.	Sab.	Sab.	Sab.
Horário/ Data	02	09	16	23	30
08:00 - 08:50	Feriado	AM	AM	AM	AM
08:50 - 09:40	Feriado	AM	AM	AM	AM
09:40 - 10:30	Feriado	AM	AM	AM	AM
10:50 - 11:40	Feriado	AM	AM	AM	AM
11:40 - 12:30	Feriado	AM	AM	AM	AM
12:30 - 13:20	Feriado	AM	AM	AM	AM

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 Boscaroli. 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
 - 60%
 - Trabalho 1 - 40%
 - Projeto prático - 60%

Trabalho 1 – 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 sete, soluções (bibliotecas, linguagens e ferramentas de alto nível) para aplicação de técnicas de aprendizado de máquina
 - 2 - Escolher um “toy example” em regressão, associação, agrupamento ou classificação e demonstrar o uso em cada uma das soluções identificadas no item 1

Trabalho 1 – 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 1 – Soluções tecnológicas para aprendizado de máquina

• Trabalho individual

- 40% da nota total

• Entrega

- Relatório técnico conforme template por e-mail
 - sergiomariano@gmail.com
 - Assunto:
 - “[pos-pucminas-AM] título do relatório”
- Data: Sábado, dia 30 de junho de 2018

Projeto prático

• Objetivos

- Experimentar um projeto completo em aprendizado de máquina

• Descrição

- Escolher um **problema real**
 - O problema deve fazer uso de pelo menos duas técnicas de aprendizado de máquina
 - Acessar os dados relacionados ao problema selecionado
 - Levantar um conjunto de perguntas e hipóteses minimamente complexas
 - Ao modelar o problema, aplicar diferentes algoritmos
- Pode ser utilizado qualquer ferramenta

Projeto prático

- Trabalho em dupla
 - 60% da nota total
- Entregas
 - 1 - Proposta de projeto
 - Resumo da proposta de projeto prático (introdução, motivação, perguntas e hipóteses e técnicas candidatas).
 - Máximo de duas laudas
 - Data: Sábado, dia 16 de junho de 2018
 - sergiomariano@gmail.com

Projeto prático

- Trabalho em dupla
 - 60% da nota total
- Entregas
 - 2 - Preparar e entregar apresentação de 15 minutos do projeto desenvolvido
 - sergiomariano@gmail.com
 - Assunto:
 - “[pos-pucminas-AM] título da apresentação”
 - Apresentação: oportunidade de conhecer diferentes técnicas de aprendizado de máquina aplicada em diferentes problemas reais.
 - Data: Sábado, dia 30 de junho de 2018

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/>



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