



Programa de disciplina

I. Identificação da disciplina

<i>Código</i>	<i>Nome da disciplina</i>	<i>Horas-aula semanais</i>		<i>Horas-aula semestrais</i>
MTMXXX	Aprendizado de Máquina	<i>Teóricas: 3</i>	<i>Práticas: 1</i>	72

II. Pré-requisito(s)

MTM3401 - Cálculo I, MTM3422 - Álgebra Linear II, MTM3520 - Laboratório de Matemática Computacional.

III. Curso(s) para o(s) qual(is) a disciplina é oferecida

Matemática.

IV. Ementa

Conceitos fundamentais e resultados clássicos em aprendizado de máquina. Técnicas de regressão, classificação, redução de dimensionalidade e de clusterização e suas aplicações. Implantação de modelos de aprendizado de máquina na nuvem. Projeto de aprendizado de máquina de ponta a ponta.

V. Objetivos

Concluindo o programa do curso, o aluno deverá ser capaz de:

- Compreender os fundamentos em aprendizado de máquina, assim como resultados teóricos essenciais sobre o assunto.
- Reconhecer conceitos matemáticos envolvidos na construção de algoritmos em machine learning, incluindo as hipóteses assumidas e suas limitações.
- Utilizar pacotes computacionais de última geração e implantar modelos de aprendizado de máquina na nuvem.
- Aplicar técnicas de machine learning para solução de problemas.

VI. Conteúdo programático

Unidade 1. Introdução ao aprendizado de máquina.

- 1.1 Conceito de aprendizado de máquina.
- 1.2 Instrumentalismo v.s. Realismo e a perspectiva do aprendizado de máquina.
- 1.3 Tipos de aprendizado.
- 1.4 Qualidade dos dados.
- 1.5 Noções básicas em sobreajuste de dados e subajuste de dados.
- 1.6 Noções gerais em resultados teóricos clássicos em machine learning.
- 1.8 Naive Bayes e aplicação na construção de um filtro de spam.
- 1.9 Projeto de aprendizado de máquina de ponta a ponta.

Unidade 2. Modelos de aprendizado de máquina e técnicas de preparação de dados

- 2.1. Introdução a regressão.
- 2.2 Gradiente descendente.
- 2.3 Modelos lineares regularizados.
- 2.4 Compensação viés/variância.
- 2.5 Métodos para validação de modelos.
- 2.6 Técnicas para ajuste e seleção de modelos.
- 2.7 K-vizinhos mais próximos (K-NN).
- 2.8 Perceptron.
- 2.9 Máquina de Vetores de Suporte (SVM).
- 2.10 Regressão logística.
- 2.11 Medidas de desempenho para classificação.
- 2.12 Classificação multiclasse.
- 2.13 Técnicas avançadas de imputação de dados.
 - 2.13.1 Imputação múltipla por equações encadeadas (MICE).

- 2.13.2 Imputação através de modelos de machine learning.
- 2.14 Técnicas de reamostragem para dados desbalanceados.
- 2.15 Árvores de decisão.
- 2.16 Métodos de Ensemble.
- 2.16.1 Classificação por votos.
- 2.16.2 Agregação com reposição amostral (Bagging).
- 2.16.3 Agregação sem reposição amostral (Pasting).
- 2.16.4 Árvores aleatórias.
- 2.16.5 Métodos de boosting.

Unidade 3. Redução de dimensionalidade e Clusterização.

- 3.1 Análise de Componentes Principais (PCA) e conexão com decomposição em valores singulares (SVD).
- 3.2 Projeções Aleatórias e Lema de Johnson-Lindenstrauss.
- 3.3 Técnicas de redução de dimensionalidade não-linear (manifold learning).
- 3.4 Aplicações de redução de dimensionalidade.
- 3.5 K-means.
- 3.6 Clusterização Hierárquica.
- 3.7 Clusterização Espacial Baseada em Densidade de Aplicações com Ruído (DBSCAN).
- 3.8 Aplicações de clusterização.

Unidade 4. Aprendizado de Máquina em produção e desenvolvimento de projeto

- 4.1 Aprendizado de máquina na nuvem.
- 4.1.1 Introdução a computação na nuvem.
- 4.1.2 Implantação de modelos.
- 4.1.3 Construção de modelos.
- 4.1.4 Ajuste a atualização de modelos.
- 4.2 Desenvolvimento e planejamento de projeto de ponta a ponta em aprendizado de máquina.

VII. Bibliografia básica

1. Gron, A. Hands-On Machine Learning with Scikit-Learn and TensorFlow 2e: Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol: O'Reilly Media, Inc, 2019.
2. Deisenroth, M. Peter; Faisal, A. Aldo; Ong, C. Soon. Mathematics for Machine Learning. Cambridge: Cambridge University Press, 2020.
3. Vapnik, V. Empirical inference science afterword of 2006. London: Springer, 2006.
4. Amazon SageMaker: Developer Guide. 2020. Disponível em:
<https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-dg.pdf>

VIII. Bibliografia complementar

1. Abu-Mostafa, Y.; Magdon-Ismael M.; Lin H.T. Learning From Data. New York: AMLBook, 2017.
2. Cunningham, J.P; Ghahramani, Z. Linear Dimensionality Reduction: Survey, Insights, and Generalizations. Journal of Machine Learning Research 16, 2859 – 2900, 2015.
3. Dasgupta, S; Gupta, A. An Elementary Proof of a Theorem of Johnson and Lindenstrauss, Random structures and algorithms 22(1), 60 – 65, 2003.
4. Hastiem, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Basiléia: Springer, 2009.
5. Murphy, K.P. Machine Learning: A Probabilistic Perspective. Cambridge: MIT press. 2012.
6. Raschka, S.; Mirjalili, V. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Birmingham: Packt Publishing, 2017.
7. VanderPlas, J. Python Data Science HandBook: Essential tools for working with Data. Sebastopol: O'Reilly Media, Inc, 2016.

Este programa foi criado pelo professor Edson Cilos Vargas Júnior em 08 de novembro de 2020.