



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>
MTMXXXX	Laboratório de Aprendizado de Máquina	<i>Teóricas: 2</i>	<i>Práticas: 2</i>	72

II. Pré-requisito(s)

MTM 3101 - Cálculo 1 (ou equivalentes), MTM 3112 - Álgebra Linear (ou equivalentes), INE5201 - Introdução à Ciência da Computação (ou equivalentes)

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

Engenharias, Física, Química e Matemática

IV. Ementa

Revisão de Álgebra Linear. Decomposição matricial; Decomposição em valores singulares. Pacotes computacionais. Aprendizado supervisionado; aprendizado não-supervisionado; subajuste e sobreajuste de dados; técnicas de preparação de dados; Ajuste de Modelos. Algoritmos de Regressão (linear, polinomial, logística e outros); Algoritmos de Classificação (K vizinhos mais próximos, SVM); Florestas aleatórias; Métodos de Ensemble. Redução de dimensionalidade; PCA.

V. Objetivos

- Construir pilares sólidos dos principais conceitos em machine learning e ilustrando, em problemas relevantes e atuais, tanto de interesse acadêmico quanto da indústria, a importância do raciocínio matemático não apenas como uma coleção de ferramentas, mas sobretudo uma maneira de pensar.
- Desenvolver a habilidade do aluno em pacotes computacionais de última geração, mas, ao mesmo tempo, fazê-lo compreender que machine learning não se resume a bibliotecas prontas, sendo, antes de tudo, técnicas de extrair padrões e conceitos dos dados, gerando modelos que podem ser representados em um computador para tomada de decisões automatizada.
- Objetivos específicos:
 - Aprofundar conhecimentos em álgebra linear, em especial decomposição matricial.
 - Compreender a diferença entre aprendizado supervisionado e não-supervisionado, suas principais diferenças, dificuldades e aplicações.
 - Entender a importância da qualidade dos dados e do pré-processamento de dados.
 - Compreender os principais conceitos matemáticos de algoritmos em machine learning.
 - Fornecer uma noção de resultados teóricos importante em Aprendizado de Máquina, por exemplo, Eficácia não razoável de dados, Teorema do Não Existe Almoço grátis, consistência universal do classificador Knn, dentre outros. A ideia é que o aluno relacione a relevância desses resultados para projetos práticos.
 - Aprender a utilizar pacotes computacionais de machine learning.
 - Identificar, compreender e aplicar as etapas de um projeto de machine learning.
 - Desenvolver um projeto prático de machine learning de ponta a ponta.
 - Conhecer com detalhes a matemática da análise de componentes principais e aplicar em um projeto de reconhecimento facial (ou outro projeto de escolha do aluno).

VI. Conteúdo programático

Unidade 1. Conceitos básicos de Aprendizado de Máquina e decomposição em valores singulares.
1.1. Tipos de aprendizado. Aprendizado não-supervisionado e supervisionado.
1.2. Qualidade dos dados. Sobreajuste de dados, subajuste de dados.
1.3. Revisão de álgebra linear. Independência e dependência linear, subespaços, base, dimensão, imagem, núcleo, posto, inversa matricial, ortogonalidade, determinante, transformações lineares, espaços afins, hiperplanos, normas, autovalores e autovetores.
1.4. Decomposição em valores singulares.

Unidade 2. Aprendizado de máquina na prática: aprenda fazendo.

- 2.1. Sistemas de regressão.
- 2.2. Visualização de dados e preparação de dados.
- 2.3. Criação de modelos, validação cruzada e GridSearch.
- 2.4. Sistemas de classificação.
- 2.5. Construção de classificador binário.
- 2.6. Medidas de desempenho. Precisão e Revocação. Matriz de confusão. Curva ROC.
- 2.7. Classificação multiclasse, multilabel e multioutput.

Unidade 3. Treinamento de modelos.

- 3.1. Regressão Linear. Método dos mínimos quadrados. Complexidade computacional.
- 3.2. Gradiente descendente. Gradiente descendente estocástico.
- 3.3. Regressão polinomial. Curvas de Aprendizado. Compensação viés/variância.
- 3.4. Modelos lineares regularizados. Regressão de Ridge. Regressão Lasso. Elastic Net.
- 3.5. Regressão logística. Função Custo. Fronteiras de Decisão. Regressão Softmax.
- 3.6. K vizinhos mais próximos.
- 3.7. Máquinas de Vetores de Suporte (SVM). Núcleos.
- 3.8. Árvores de decisão. Florestas aleatórias. Métodos de Ensemble. Entropia e coeficiente de Gini.

Unidade 4. Redução de dimensionalidade

- 4.1. Maldição de dimensionalidade.
- 4.2. Análise de componentes principais e aplicações.

VII. Bibliografia básica

1. Gron, A. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol: O'Reilly Media, Inc, 2017.
2. Deisenroth, M. Peter; Faisal, A. Aldo; Ong, C. Soon. Mathematics for Machine Learning. Cambridge: Cambridge University Press, 2020.

VIII. Bibliografia complementar

1. VanderPlas, J. Python Data Science HandBook: Essential tools for working with Data. Sebastopol: O'Reilly Media, Inc, 2016.
2. Hastiem T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Basiléia: Springer, 2009.

Este programa foi criado pelo professor Edson Cilos Vargas Júnior em 20 de Janeiro de 2020.