



## Plano de Desenvolvimento da Disciplina

MC886A/MO444B - Tópicos em Aprendizado de Máquina

Docente: Prof. Marcos M. Raimundo

### 1. Pré-requisitos

Esta disciplina exige conceitos básicos de matemática e computação vistos nas disciplinas pré-requisitadas: MA327 (Álgebra Linear); ME323 (Introdução aos Métodos Probabilísticos) ou ME210 (Probabilidade I); e MC102 (Algoritmos e Programação de Computadores). Exige-se também alguma experiência com a linguagem de Programação Python (utilizada em MC102) e conhecimentos básicos de Cálculo (MA111 e MA211). Uma boa referência bibliográfica para os pré-requisitos matemáticos é a parte 1 do livro “Mathematics for Machine Learning” [1].

### 2. Programa da disciplina

O programa desta turma da disciplina cobrirá, em diferentes níveis de profundidade, todos os tópicos previstos na ementa da disciplina, que envolvem “Técnicas de aprendizado de máquina estatístico para classificação, agrupamento e detecção de outliers.” Em particular, serão abordados os seguintes assuntos:

- Introdução ao Aprendizado de Máquina;
- Modelos lineares;
- O dilema entre viés e variância;
- Sobreajuste, regularização e validação;
- Redes neurais;
- Máquinas de vetores de suporte;
- Árvores de decisão e métodos *ensemble*;
- Redução de dimensionalidade;
- Agrupamento (*clustering*);
- Detecção de anomalias (*outliers*).

Cada assunto será coberto com um determinado número de aulas, que deverá variar entre uma e seis aulas.

### 3. Cronograma de atividades com nota

Os alunos serão avaliados através de (quatro) trabalhos realizados em duplas, além de duas provas individuais.

Os trabalhos em dupla deverão avaliar o domínio dos alunos dos conceitos e técnicas aprendidos em aula, através de exercícios práticos. Esses exercícios utilizarão conjuntos de dados escolhidos pelo docente e pelo monitor e as soluções deverão ser implementadas em Python, utilizando Jupyter Notebook. Para resolver as atividades propostas, poderá ser solicitado o uso de bibliotecas clássicas de Aprendizado de Máquina, tais como scikit-learn e PyTorch, a implementação de algoritmos “do zero”, estratégias que envolvam o uso de Inteligência Artificial (IA) generativa, ou mesmo uma combinação dessas técnicas. Os enunciados de cada trabalho e suas respectivas datas de entrega serão distribuídos de forma aproximadamente uniforme ao longo do semestre.

Já as provas individuais serão dadas nos dias **30 de setembro** e **25 de novembro** (ambos segunda-feira). Cada prova será aplicada em horário de aula, sendo vedada a consulta a qualquer material impresso ou eletrônico e tendo a duração máxima de duas horas. Essas provas cobrarão aspectos teóricos de Aprendizado de Máquina que serão estudados ao longo da disciplina.

### 4. Critérios de avaliação

Sejam  $T_i$  a nota do  $i$ -ésimo trabalho,  $i \in \{1, 2, 3, 4\}$ , e  $P_j$  a nota da  $j$ -ésima prova,  $j \in \{1, 2\}$ . Tanto  $T_i$  quanto  $P_j$  valem de 0 (zero) a 10 (dez). A média final  $MF$  de cada aluna(o) será calculada da seguinte maneira:

$$MF = 0.2P_1 + 0.4P_2 + 0.4 \sum_{i=1}^4 \alpha_i T_i, \quad (1)$$

onde  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$  e  $\alpha_4 = 0.4$ .

Para a pós-graduação (MO444) e alunos de graduação cursando simultaneamente, o sistema de conceitos irá se basear nas seguintes notas:

- A - Acima de 8,5.
- B - Entre 7 e 8,5.
- C - Entre 5,5 e 7.
- D - Abaixo de 5,5.

### 5. Punição para fraudes e plágios

Os trabalhos em dupla devem ser resolvidos apenas pelos componentes da mesma, sem o auxílio de terceiros. Detecção de fraude ou plágio em um trabalho  $i$  implica em  $T_i = 0$  para todos os envolvidos (quem recebeu ajuda e também quem ajudou, se aplicável).

No caso das provas, detecção de fraude ou plágio ou de consulta a colegas durante a realização dos exames resultará em  $P_j = 0$  para todos os envolvidos.

Reincidência nos casos acima implica em  $MF = 0$  para todos os envolvidos.

## 6. Forma de oferecimento da disciplina

As aulas serão presenciais, e ocorrerão sempre no **CB12**:

- Segundas, das 21 às 23 horas;
- Quartas, das 19 às 21 horas.

## 7. Bibliografia a ser utilizada

O docente deverá seguir majoritariamente como livro-texto a edição em Python do “Introduction to Statistical Learning” [2], que foi lançada recentemente e cujo pdf está disponível gratuitamente em: [www.statlearning.com](http://www.statlearning.com).

Além disso, material didático sobre os assuntos listados na seção 2 deverão ser retirados de artigos e também dos seguintes livros:

- “Dive into Deep Learning” [3];
- “Introduction to Data Mining” [4];
- “Learning from Data” [5];
- “The Elements of Statistical Learning” [6];
- “Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems” [7].

Alguns dos livros acima (as referências [3, 6]) também foram disponibilizados gratuitamente na internet pelos autores.

## Referencias

- [1] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020.
- [2] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. *An Introduction to Statistical Learning with Applications in Python*. Springer, 2023.
- [3] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.
- [4] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Education India, 2016.
- [5] Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. *Learning from data*, volume 4. AMLBook New York, 2012.
- [6] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.
- [7] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O’Reilly Media, Inc., 2019.