

# MC886A - Aprendizado de Máquina - Turma A

## Plano de Desenvolvimento da Disciplina

**Docente:** Marcelo da Silva Reis

**Monitores PED:** Décio Gonçalves de Aguiar Neto  
Hudson Martins S. Bruno

Campinas, 12 de março de 2022

Este documento descreve, de forma sucinta, o plano de desenvolvimento da disciplina (PDD) de MC886 - Aprendizado de Máquina, mais especificamente da turma A do primeiro semestre de 2022. Em particular, são destacados, de acordo com os requisitos do Regimento Geral de Graduação, o cronograma de atividades, os critérios de avaliação, punição para fraudes e plágios e a bibliografia a ser utilizada ao longo do semestre. Este PDD também contempla a forma de oferecimento da disciplina, por conta da pandemia de COVID-19 ainda estar em andamento.

### Sumário

|  |          |
|--|----------|
| <b>1 Programa da disciplina</b>              | <b>1</b> |
| <b>2 Cronograma de atividades</b>            | <b>2</b> |
| <b>3 Critérios de avaliação</b>              | <b>3</b> |
| <b>4 Punição para fraudes e plágios</b>      | <b>3</b> |
| <b>5 Forma de oferecimento da disciplina</b> | <b>3</b> |
| 5.1 Protocolos de COVID-19 . . . . .         | 4        |
| <b>6 Bibliografia a ser utilizada</b>        | <b>4</b> |

### 1 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 e a dimensão VC;
- Sobreajuste, regularização e validação;
- Redes neurais;
- Máquinas de vetores de suporte;
- Árvores de decisão e métodos *ensemble*;
- Redução de dimensionalidade;
- Agrupamento (*clustering*);
- Redes Bayesianas;
- Detecção de anomalias (*outliers*).

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

## 2 Cronograma de atividades

Esta disciplina não terá exames individuais, e os alunos serão avaliados através de (quatro) trabalhos realizados em duplas e também de um projeto final elaborado em grupos de 4-5 alunos.

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 monitores e as soluções deverão ser implementadas em Python, utilizando Jupyter Notebook. As soluções não poderão utilizar bibliotecas de Aprendizado de Máquina tal como a scikit-learn, devendo a dupla implementar por conta própria os algoritmos de aprendizado. Os enunciados de cada trabalho e suas respectivas datas de entrega serão distribuídos de forma aproximadamente uniforme ao longo do semestre letivo.

Já nos trabalhos em grupo, os alunos deverão escolher um *dataset* de Aprendizado de Máquina, definir um problema a ser respondido com o mesmo utilizando algoritmos de aprendizado, resolver o problema com esses algoritmos e analisar os resultados. A definição do *dataset* e o enunciado do problema deverão ser feitos até aproximadamente o meio do semestre letivo, em data a definir, através da entrega de uma proposta de projeto. Já a resolução e a análise de resultados deverão ser entregues no final do semestre, através da escrita de um short paper, da gravação de um vídeo de aproximadamente 5 minutos apresentando esse trabalho e da disponibilização do código-fonte elaborado utilizando Jupyter notebook.

### 3 Critérios de avaliação

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

$$MF = \begin{cases} 0,6P + 0,4 \sum_{i=1}^4 \alpha_i T_i, & \text{se } \sum_{i=1}^4 \alpha_i T_i \geq 5 \\ \sum_{i=1}^4 \alpha_i T_i, & \text{caso contrário,} \end{cases}$$

onde  $\alpha_1 = 0,1$ ,  $\alpha_2 = 0,2$ ,  $\alpha_3 = 0,4$  e  $\alpha_4 = 0,3$ . Já a nota do projeto final  $P$  será composta por uma média ponderada na qual cada um dos quatro elementos (proposta, vídeo, short paper e código-fonte) valem de zero a dez e têm o mesmo peso para compor essa média.

### 4 Punição para fraudes e plágios

Os trabalhos em dupla devem ser resolvidos apenas pelos componentes da mesma, sem a consulta 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). Reincidência implica em  $MF = 0$  para todos os envolvidos.

No caso do projeto final, detecção de fraude e/ou plágio (da internet e/ou de projetos de colegas) em qualquer um dos elementos do mesmo (proposta, vídeo, short paper ou código-fonte) resulta em  $P = 0$  para todos os envolvidos.

### 5 Forma de oferecimento da disciplina

Por conta da pandemia de COVID-19, até segunda ordem a disciplina será oferecida num regime híbrido, no qual as aulas serão lecionadas presencialmente em sala de aula para alunos até o limite da lotação máxima permitida de acordo com o escritório de dados da UNICAMP e ao mesmo tempo transmitidas ao vivo pelo Google Meet. Em dias em que o número de alunos matriculados (de MC886 e também de MO444, que é oferecida conjuntamente) exceder a capacidade da sala de aula com distanciamento social, metade da turma precisará necessariamente assistir à aulas de casa, enquanto que a outra metade poderá assistir presencialmente. Na aula seguinte em que a capacidade for novamente excedida, haverá alternância de alunos (i.e., quem assistiu presencialmente terá que acompanhar online e vice-versa).

As aulas serão transmitidas ao vivo no Google Meet e serão gravadas para que possam ser acompanhadas de forma assíncrona. Em dias em que o(a) aluno(a) tiver direito a assistir presencialmente lhe será cobrada presença, que será verificada por lista de presença ou então pelo registro no relatório de participação ao vivo que é emitido ao término de uma sessão de Google Meet.

## 5.1 Protocolos de COVID-19

Todos os alunos que assistirem às aulas presencialmente deverão estar imunizados contra o SARS-CoV-2 e seguir os protocolos de orientação da UNICAMP. O uso de máscaras é obrigatório durante toda a permanência em sala de aula, até que o Governo do Estado de São Paulo e a UNICAMP determinem que tal uso não seja mais obrigatório.

Durante o curso da pandemia, alunos(as) que apresentem sintomas gripais e/ou testem positivo para o SARS-CoV-2 não deverão ir às aulas, devendo resguardar em casa o devido período de quarentena, sendo facultativo acompanhar de forma remota se for possível. Se for o docente quem apresentar sintomas gripais e/ou testar positivo, o mesmo cancelará a aula ou então a ministrará de forma totalmente remota, de acordo com a situação em questão.

## 6 Bibliografia a ser utilizada

O docente não deverá se basear em um único livro-texto para cumprir o cronograma proposto na seção 2; ao invés disso, material didático sobre os assuntos listados nessa seção deverão ser retirados de artigos e também dos seguintes livros:

- “An Introduction to Statistical Learning” [1];
- “Diving into Deep Learning” [2];
- “Introduction to Data Mining” [3];
- “Learning from Data” [4];
- “Pattern Classification” [5];
- “The Elements of Statistical Learning” [6].

Alguns dos livros acima (as referências [1, 2, 6]) foram disponibilizados gratuitamente na internet pelos autores. Para o projeto final também será recomendada a referência “Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems” [7].

## Referências

- [1] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.
- [2] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.
- [3] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Education India, 2016.

- [4] Yaser S. Abu-Mostafa, Malik Magdon-Ismael, and Hsuan-Tien Lin. *Learning from data*, volume 4. AMLBook New York, 2012.
- [5] Peter E. Hart, David G. Stork, and Richard O. Duda. *Pattern classification*. Wiley Hoboken, 2000.
- [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.