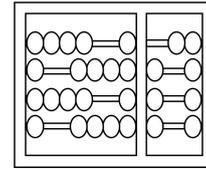




MC959 - TÓPICOS  
EM INTELIGÊNCIA  
ARTIFICIAL



PROF. MARCELO REIS  
msreis@unicamp.br

# INTRODUÇÃO AO APRENDIZADO AUTO-SUPERVISIONADO

Campinas, 17 de julho de 2024

## Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Pré-requisitos</b>	<b>2</b>
<b>3</b>	<b>Ementa</b>	<b>2</b>
<b>4</b>	<b>Local e horários de aulas</b>	<b>2</b>
<b>5</b>	<b>Avaliação</b>	<b>3</b>
5.1	Trabalho prático . . . . .	3
5.2	Seminário . . . . .	3
5.3	Liberação de notas . . . . .	3
<b>6</b>	<b>Medidas em casos de fraude e/ou plágio</b>	<b>3</b>
<b>7</b>	<b>Bibliografia</b>	<b>3</b>

## 1 Introdução

Nos últimos anos, o aprendizado auto-supervisionado de representação (SSRL, do inglês *Self-Supervised Representation Learning*) tem atraído cada vez mais atenção dos acadêmicos e profissionais de aprendizado de máquina, uma vez que permite o aproveitamento de grandes quantidades de dados não rotulados para aprender representações de dados que são úteis para uma ampla gama de tarefas posteriores, tais como as existentes em visão computacional, processamento de linguagem natural, bioinformática, entre muitas outras áreas. Neste curso, pretendemos apresentar o paradigma de SSRL, incluindo aspectos conceituais e técnicos/práticos, e sua utilização

em diferentes modalidades e aplicações de dados. Pretendemos também mostrar como os modernos grandes modelos de linguagem (LLM, do inglês *Large Language Models*) são construídos a partir do SSRL, bem como os aspectos éticos que envolvem seu treinamento e uso.

## 2 Pré-requisitos

Não há pré-requisitos formais. Todavia, espera-se que os alunos já estejam familiarizados com os fundamentos de aprendizado supervisionado e não supervisionado, que são ensinados em disciplinas tais como MC886/MO444 (“Aprendizado de Máquina e Reconhecimento de Padrões”).

## 3 Ementa

- Revisão de arquiteturas de deep learning: CNNs, RNNs, auto-atenção e transformadores;
- Revisão da aprendizado por transferência;
- Princípios de aprendizado auto-supervisionado;
- Principais componentes de uma abordagem de aprendizado auto-supervisionado: tarefas de pretexto e tarefas-alvo (*downstream tasks*);
- Tipos de tarefas de pretexto: predição mascarada, predição de transformação, discriminação de instâncias, abordagens contrastivas, agrupamento;
- Adaptação supervisionada de tarefas-alvo;
- Métodos SSRL relevantes para diferentes modalidades de dados (e.g., imagens, texto, séries temporais) e também para multimodalidade;
- Adaptação de domínio e ajuste fino auto-supervisionado;
- Aprendizado auto-supervisionado de LLMs;
- Aprendizado em contexto (*in-context learning*);
- Engenharia de prompt (*prompt engineering*);
- *Retrieval-Augmented Generation* (RAG);
- Aspectos éticos do treinamento SSRL em *big data* (público) não rotulado;
- Desenvolvimento de projetos de IA utilizando SSRL.

## 4 Local e horários de aulas

- Segundas, das 16 às 18 horas, na CC51;
- Quartas, das 16 às 18 horas, na CC51.

## 5 Avaliação

A avaliação consistirá em um trabalho prático e em um seminário.

### 5.1 Trabalho prático

O trabalho prático valerá 50 % da nota, e deverá ser desenvolvido em grupos de 2 a 3 pessoas. Esse trabalho envolverá o desenvolvimento de uma solução de IA auto-supervisionada para um problema prático a ser escolhido pelo grupo. O grupo deverá entregar o código-fonte da solução desenvolvida e um relatório descrevendo o contexto do problema, a metodologia empregada e os resultados obtidos. Essa entrega deverá ocorrer na segunda quinzena de novembro.

### 5.2 Seminário

Um seminário deverá ser apresentado pelo mesmo grupo do trabalho prático, em horário de aula no mês de novembro. Essa apresentação valerá os 50 % restantes da nota. Na apresentação o grupo deverá mostrar o trabalho desenvolvido e responder a dúvidas do ministrante. Perguntas conceituais das aulas teóricas poderão eventualmente ser feitas pelo ministrante.

### 5.3 Liberação de notas

As notas deverão ser liberadas em até uma semana após a entrega da última atividade avaliativa por parte do último grupo (seminário e trabalho prático).

## 6 Medidas em casos de fraude e/ou plágio

Qualquer tipo de fraude na metodologia ou nos resultados apresentados, ou de plágio no código-fonte e/ou no relatório e/ou nos slides de apresentação, resultará em nota 0 (zero) na disciplina para todos os envolvidos.

## 7 Bibliografia

A maior parte do material do curso será baseada em artigos (recentes) sobre SSRL, que serão indicados pelo ministrante no decorrer da disciplina. No entanto, alguns livros podem ser úteis para os conteúdos pré-requisitos e fundamentos de aprendizado de máquina em geral [1, 2, 3].

### Referências

- [1] 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.
- [2] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021. d2l.ai.

- [3] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. *An Introduction to Statistical Learning with Applications in Python*. Springer, 2013. [www.statlearning.com](http://www.statlearning.com).