



## Redes Neurais e Aprendizado Profundo

Prof. Diego Furtado Silva

Departamento de Ciências de Computação (SCC)

Instituto de Ciências Matemáticas e de Computação (ICMC)

Universidade de São Paulo

### Trabalho prático

Nota inicial: Apesar de possuir o título de “trabalho prático”, este documento se refere a diferentes entregas, com diferentes notas parciais. Há uma atividade não obrigatória, mas complementar a sua nota se for cumprida.

O objetivo deste trabalho é avaliar a capacidade dos alunos em implementar e avaliar arquiteturas de redes neurais para resolver problemas práticos do mundo real. Para isso, o trabalho deve ser realizado por **grupos de 5 a 7 pessoas**. O trabalho contará com duas etapas (além de uma extra descrita mais abaixo). As entregas se complementam, formando uma solução para um único problema. Inicialmente, deve-se escolher um problema com aspecto de multimodalidade ou outras características complementares em relação à descrição do fenômeno estudado. Alguns exemplos são:

- Classificação de imagem, com texto descritivo (duas modalidades: imagem e texto);
- Regressão dos eixos de emoções por meio da fala e de imagem do rosto;
- Descrição de cena, podendo passar por classificação de imagem, mas também considerando o tempo (o tempo é um exemplo de característica complementar que não outra modalidade).

A **primeira entrega** deste trabalho deve conter a solução para uma das modalidades descrita. Seu prazo de entrega é dia **22/06/2025**. Esta entrega deverá conter apenas um código para a solução adotada (preferencialmente utilizando PyTorch). O código deve estar bem estruturado e disponibilizado em um repositório público como o GitHub. Ainda, o código deve estar bem comentado, incluindo uma breve descrição do problema atacado e o conjunto de dados utilizado no início do arquivo principal. Além disso, a documentação interna deve deixar bem claro o que o código pretende fazer em termos de arquitetura, manipulação dos dados, avaliação, entre outros. Esta etapa comporá 50% da nota final.

A **segunda entrega**, limitada para o dia **07/07/2025**, deve conter a solução completa para o problema. Utilizando o exemplo da classificação multimodal, a primeira entrega pode conter uma solução para a classificação por meio da imagem e a segunda utilizando ambas as modalidades. Se for da preferência do grupo, a primeira pode conter a solução por meio da classificação do texto. O material a ser entregue nesta etapa deve conter o código fonte, com a mesma qualidade exigida na primeira entrega, além de um breve relatório (sem contar figuras, no máximo duas páginas), explicando o que foi realizado e mostrando os resultados.

É importante notar que a avaliação da primeira entrega focará mais na solução do problema e a segunda no avanço que o aspecto complementar pode oferecer. Assim, a segunda entrega pode ser mais simples em termos de solução do problema em si.

Serão considerados os seguintes critérios para compor as notas:

- Qualidade do código apresentado, incluindo sua documentação interna;
- Decisões em termos de arquitetura;
- Manipulação correta dos dados;
- Corretude na forma de avaliação;
- Qualidade do relatório entregue.

Por fim, há uma **terceira entrega**, não obrigatória. Essa entrega poderá complementar em até dois pontos a nota final de cada integrante do grupo. Ela está relacionada a possibilidade dessa disciplina oferecer horas de extensão nos próximos semestres, por meio da apresentação de conceitos relacionados ao trabalho à comunidade, no dia **11/06**. Nela, o grupo deve apresentar um pôster sobre o problema que irá atacar e como ele pode ser resolvido. Esta etapa não precisa conter os resultados finais, apenas informar o público sobre o que pode ser feito com redes neurais. Lembrando que o público será majoritariamente composto por alunos de ensino médio. Portanto, não deve ser uma apresentação demasiadamente técnica. Os pôsteres poderão ser impressos na gráfica do ICMC.

Outras informações:

- Os trabalhos podem utilizar quaisquer recursos, incluindo modelos pré-treinados.
- Idealmente, o trabalho (em especial, a primeira parte) deve comparar mais de uma arquitetura, uma vez que muitas decisões nesse contexto são empíricas.
- Os conjuntos de dados podem vir de fontes como Kaggle (ex: <https://www.kaggle.com/search?q=multimodal+in%3Adatasets>) ou Papers with Code (<https://paperswithcode.com/datasets?q=multimodal&v=lst&o=match>);
- Não haverá apresentação, pois não sobrou tempo hábil para isso.