



6

Projeto Final

Objetivo: Abordar um problema real de engenharia usando processamento de sinais. Desenvolver autonomia para definir a melhor estrutura de rede neural para resolver o problema proposto, configurando os hiperparâmetros da rede e analisando criticamente os resultados.

O projeto final pode ser desenvolvido em grupo de até 3 alunos.

Para desenvolver esta atividade o grupo deve escolher um problema de aprendizado de máquinas associado a uma base de dados, que pode ser própria ou pública.

Há várias opções de bases de dados públicas nos links:

- UCI: <https://archive.ics.uci.edu/>
- Kaggle: <https://www.kaggle.com/datasets>
- Wikipedia: https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research
- Papers with code: <https://paperswithcode.com/datasets>

Visualize a base de dados e certifique-se que ela é bem documentada, completa e adequada para o escopo da disciplina.

O grupo é livre para solucionar o problema empregando as técnicas e metodologias que preferir, se guiando pelo que foi apresentado ao longo da disciplina. As dificuldades, decisões e pressupostos que forem assumidos ao longo do desenvolvimento do projeto devem constar no relatório final e serem comunicadas na apresentação do Seminário.

6.1 ROTEIRO

1. Defina as variáveis de entrada e saída do seu problema.
2. Verifique se há variáveis categóricas, dados faltantes ou entradas inválidas na base de dados. Faça os ajustes necessários.
3. Realize uma análise estatística das variáveis de entrada e saída e verifique se eles seguem uma distribuição/comportamento esperado e se há *outliers*.
4. Verifique a necessidade de aplicação de técnicas de pré-processamento, como filtros, normalização ou padronização.



5. Escolha a técnica de redes neurais mais adequada para o problema e base de dados escolhidos.
6. Verifique se é possível extrair características dos sinais que permitam reduzir a dimensão dos dados de entrada.
7. Verifique se é necessário realizar uma seleção das variáveis de entrada.
8. A base de dados é genérica e representativa? É balanceada? Faça a partição dos dados de treinamento/validação/teste (a partição de testes é opcional) mantendo em todos os conjuntos as características de proporção de dados e representatividade inicial da base.
9. Ajuste os parâmetros do modelo, apresentando todos os ranges e configurações testadas, com os respectivos desempenhos. Considere alterar: número de camadas, número de neurônios, função de ativação, modelos da função de perda e de otimização, taxa de aprendizado, etc.
10. Analise criticamente o resultado e certifique-se de que não esteja ocorrendo *overfitting*.
11. Após finalizada a estrutura da rede neural, avalie o desempenho com a função '*evaluate*' com a partição de teste, ou apresente o desempenho obtido usando a técnica de *k-fold cross-validation*, com $k \geq 5$.
12. Compare o seu resultado com o de outros trabalhos com propostas semelhantes. Indique as vantagens e desvantagens entre as diferentes abordagens.

6.2 PRIMEIRO BIMESTRE: PROPOSTA DE PROJETO

O documento a ser entregue deve utilizar o primeiro template da IEEE, disponível em: <https://www.overleaf.com/latex/templates?q=ieee>

A proposta de projeto deve conter:

1. Título do projeto.
2. Integrantes do grupo.
3. Introdução: Introduzir o problema a ser abordado, a motivação e os objetivos.
4. Metodologia: Apresentar a base de dados: variáveis de entrada/saída, número de amostras e todas as informações que forem pertinentes para sua especificação. Se a base de dados se for pública, indicar o link para acesso.

6.3 EXAME

6.3.1 Seminário: apresentação oral

O seminário deve ser apresentado em sala por todos os integrantes do grupo no dia combinado, durante o horário de aula.

A avaliação da apresentação será individualizada, portanto todos os membros devem participar ativamente da apresentação.

Cada grupo terá de 10 a 20 minutos para expor o trabalho.

É permitido utilizar slides, relatórios ou códigos que o grupo julgue necessário para uma melhor explanação sobre o projeto.

O seminário deve necessariamente abordar:



- i. Apresentação dos membros do grupo.
- ii. Descrição do problema escolhido, motivação e objetivos.
- iii. Apresentação da base de dados.
- iv. Escolhas relevantes que guiaram o projeto.
- v. Dificuldades encontradas e como o grupo as contornou.
- vi. Apresentar a estrutura final do modelo proposto para solucionar o problema e o desempenho.
- vii. Analisar criticamente a solução encontrada. Alguns dos questionamentos que devem ser considerados: Ela soluciona de maneira satisfatória o problema proposto? Quais as limitações do problema/solução? Etc.
- viii. Trabalhos futuros que poderiam ser encaminhados para melhoria da solução, melhor abrangência do projeto ou implementações práticas.

6.3.2 Manuscrito

Entregar todo o código gerado com os devidos comentários (ou o link do Colab, contendo as devidas seções comentadas).

Também deve ser entregue um breve relatório (máximo 5 páginas) complementando o documento 'Proposta de Projeto' elaborada no primeiro bimestre. Utilizar o primeiro template da IEEE, disponível em: <https://www.overleaf.com/latex/templates?q=ieee> ou o template do SIGE <https://www.sige.ita.br/homesubmissoes/>.

1. Título do projeto.
2. Integrantes do grupo.
3. **Introdução:** Introduzir o problema a ser abordado, a motivação e os objetivos.
4. **Metodologia:** Apresentar a base de dados: variáveis de entrada/saída, número de amostras e todas as informações que forem pertinentes para sua especificação. Se a base de dados for pública, indicar o link para acesso.
Breve descrição de todas as técnicas empregadas para o desenvolvimento do trabalho.
5. **Resultados:** Apresentação dos principais resultados. Discutir brevemente o desempenho do modelo proposto, realizando uma análise crítica sobre a viabilidade da solução e suas limitações. Se possível, comparar com as soluções propostas em outros trabalhos de mesmo escopo.
6. **Conclusão:** Considerações finais e trabalhos futuros.