

Disciplina: Aprendizagem de Máquina Probabilística

Período: 2025.2

Professor: César Lincoln Cavalcante Mattos

Projeto Final

Trabalhando em **trios**, desenvolva um projeto de aprendizagem de máquina probabilística contendo as seguintes partes:

- Um problema de estimação de densidade de probabilidade **ou** redução de dimensionalidade **ou** dados faltantes **ou** geração de dados;
 - Use modelos de GMM, PPCA, FA, GPLVM, VAE, NF, FM, Difusão, etc.
- Um problema de regressão com previsões probabilísticas;
 - Use modelos de GP, regressão linear/polinomial/RBF Bayesiana, etc.
- Ajuste de hiperparâmetros via otimização Bayesiana.

Alunos da graduação: Escolha **2** dos pontos acima para trabalhar.

Alunos da pós-graduação: Todos os pontos devem ser trabalhados.

O projeto deve ser documentado a partir dos entregáveis abaixo:

1. **Relatório técnico:** Resumo do projeto, contextualização do problema, descrição dos dados (atributos, saídas, etc), detalhamento da metodologia de aplicação do(s) modelo(s) de aprendizagem e de análise dos resultados.
2. **Códigos:** Jupyter notebook e/ou repositório de códigos usados nas simulações.
3. **Apresentação:** Sumarização do trabalho para a turma.

O projeto possui as seguintes datas relevantes previstas:

- **11/12/2025 a 18/12/2025:** Acompanhamento da finalização dos projetos.
- **07/01/2026:** Entrega da versão final da documentação via SIGAA.
- **08/01/2026, 13/01/2026, 15/01/2026 e 20/01/2026:** Apresentações das equipes.

Observações:

- Tirem dúvidas sobre o projeto com antecedência ao longo das aulas ou por e-mail.
- A equipe deve indicar como cada participante contribuiu para o projeto.

- O projeto deverá usar dados reais (não sintéticos).
- Use mais de um modelo de aprendizagem de máquina para o problema escolhido, permitindo comparações de desempenho.
- As implementações podem ser feitas em qualquer linguagem de programação.
- Será permitido o uso de pacotes de software já existentes.
- Recomenda-se o uso de L^AT_EX. Você pode seguir o modelo de um dos principais congressos da área: Neurips, ICLR, ICML, AAAI, AISTATS, UAI, etc.

Sugestões de fontes de dados reais:

- Dados do Íris Lab: <https://irislabs.ce.gov.br/>
- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/index.php>
- OpenML: <https://www.openml.org/search?type=data>
- Kaggle: <https://www.kaggle.com/datasets>
- Google Dataset Search: <https://toolbox.google.com/datasetsearch>
- Carnegie Mellon University Machine Learning and AI Datasets: <https://guides.library.cmu.edu/machine-learning/datasets>
- Portal Brasileiro de Dados Abertos: <https://dados.gov.br/>
- U.S. Government's Open Data: <https://www.data.gov/>
- EU Open Data Portal: <https://data.europa.eu/euodp/en/home>
- UK Data Service: <https://www.ukdataservice.ac.uk/>
- Awesome Public Datasets: <https://github.com/awesomedata/awesome-public-datasets>