

# IA048 – Aprendizado de Máquina

## Exercício de Fixação de Conceitos (EFC) 2 – Classificação

Turma A – 2º semestre de 2022

Prof: Levy Boccato Email: lboccato@dca.fee.unicamp.br

Prof: Romis Attux Email: attux@dca.fee.unicamp.br

## Introdução

Nesta atividade, vamos abordar o problema de classificação de imagens de ultrassom mamário em duas categorias: (1) câncer (classe negativa, com rótulo 0) e (2) normal (classe positiva, com rótulo 1). A base de dados **breastmnist** [Gomaa et al., 2020, Yang et al., 2021] (<https://medmnist.com/>) possui 780 imagens representadas em níveis de cinza de dimensão  $28 \times 28$  pixels. Algumas amostras da base são exibidas na Figura 1.

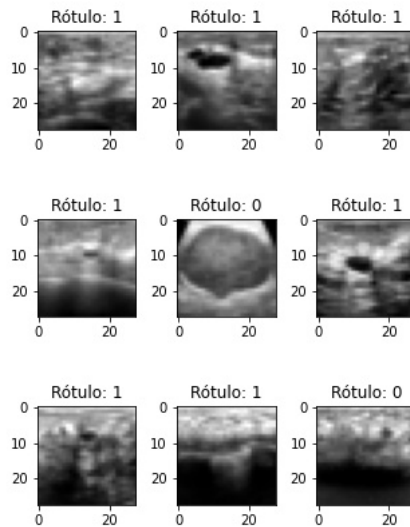


Figura 1: Amostras do conjunto de dados **breastmnist**.

Vamos trabalhar com dois modelos de classificação: (1) regressão logística e (2)  $k$  vizinhos mais próximos (kNN, do inglês *k-nearest neighbors*).

## Atividades

- Analise o balanceamento das classes nos conjuntos disponíveis.
- Faça a aplicação do modelo de regressão logística ao problema de classificação de imagens de ultrassom mamários. Descreva de forma clara a metodologia empregada para as etapas de treinamento e validação (o que inclui, por exemplo, a busca pelos hiperparâmetros). Tendo em mãos a versão final do modelo, avalie o desempenho do classificador no conjunto de teste considerando (pelo menos): (1) a matriz de confusão, (2) a acurácia, (3) a  $F_1$ -medida, (4) a acurácia balanceada e (5) a curva ROC.
- Considere, agora, a técnica kNN. Mostre como o desempenho desse classificador junto aos dados de validação varia conforme o parâmetro  $k$  é modificado. De posse da melhor versão do modelo, avalie o seu desempenho no conjunto de teste, explorando as mesmas métricas do item anterior (exceto a curva ROC), e faça uma comparação entre os dois modelos.

**Observações:** No relatório, não é necessário descrever a teoria sobre os modelos de regressão logística e kNN. Não obstante, todas as escolhas feitas com respeito às características dos modelos, dos dados e dos experimentos devem ser apresentadas e justificadas, de modo a possibilitar o entendimento e a reprodução da metodologia.

## Referências

- [Yang et al., 2021] J. Yang, R. Shi, D. Wei, z. Liu, L. Zhao, B. Ke, H. Pfister, B. Ni, *MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification*. arXiv preprint arXiv:2110.14795, 2021.
- [Gomaa et al., 2020] W. A.-D., M. Gomaa, H. Khaled, A. Fahmy, *Dataset of Breast Ultrasound Images*, Data in Brief, vol. 28, 2020.