Project Proposal
Privacy Preserving Binary Classifier for Breast Tumors

Alessandro Stanghellini, Anastasiya Merkushova, Yannick Martin

University of Basel
Privacy-Preserving Methods for Data Science and Distributed Systems (PDS) course
Spring Semester 2024

1 Goal

We propose to conduct an experiment wherein one participant will undertake the training of a binary classification machine learning model with a focus on privacy preservation, employing the Stochastic Gradient Descent (SGD) algorithm and Federated Learning (FL) methodology. Subsequently, a participant will transmit solely the trained model along with the type of mechanism applied. Two additional participants will then endeavor to perform model unlearning, leveraging existing pipelines and methodologies, with the exploration and identification of said methodologies constituting a fundamental aspect of the project.

2 Libraries

The classic libraries are used to create Python classifiers (PyTorch, Numpy, Panda). For the privacy preserving model we want to use:

- PyVacy: Privacy Algorithms (DPSGD) for PyTorch
- Flower: A Friendly Federated Learning Framework

3 Dataset

Breast Cancer Wisconsin (Diagnostic): Dataset contains extracted features of fine needle aspirate breast mass images. The dataset has 569 instances. The dataset can be found here `https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic`