

## TensorFlow Project

In this project you are asked to run experiments on the Wisconsin Breast Cancer dataset. There are 569 examples, each labeled as 0 or 1. Classical approaches achieve accuracy of over 98%. You are asked to train two classifier for this problem using TensorFlow. The first classifier should be trained using only 8% of the data (46 examples). The second classifier should be trained using only 16% of the data (91 examples).

1. Your programs must set the random seeds of python and tensorflow to 1 to make sure that your results are reproducible.
2. Your programs will be tested by training them on randomly selected fractions of the dataset. The testing data will be the entire dataset.
3. The training and testing of each program should not take more than 5 minutes.

## Provided programs and data

1. The dataset is given in the files **x\_test.csv** and **y\_test.csv**
2. A random subset of 46 training examples in **x\_train8.csv** and **y\_train8.csv**
3. A random subset of 91 training examples in **x\_train16.csv** and **y\_train16.csv**
4. An example program **proj\_example.py**.
5. A program that can extract a random fraction from the training data is available as **fraction\_xy.py**.

## What you need to do

Design a network to solve this problem. You can use all the functionality of tensorflow, not only the parts that were described in class. But your programs cannot read additional material the hard drive.

## Grading

We will generate random subsets of training examples by running the program **fraction\_xy.py** with a seed that is kept secret. If, for example, the seed is 7, generating a fraction of 8% can be done as follows:

```
python3 fraction_xy.py x_test.csv y_test.csv 0.08 7
```

This creates the files **x\_test\_7\_8.csv** and **y\_test\_7\_8.csv** that should be renamed to **x\_train.csv** and **y\_train.csv**

Your grade will be based on the accuracy of your models trained with the generated examples and tested on the entire testing data.

## What you need to submit

1. Source code of the python script. They should be named as follows:

```
userid-8.py  
userid-16.py
```

where userid is your user id.

2. Documentation describing your network, and the results of experiments/accuracy that your program achieves on the provided data.