

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%, using cross validation.

You are asked to create a classifier to this problem using TensorFlow. Your classifier should be using only 10% of the data (57 examples) for training. It is expected that the accuracy in this case would be significantly lower.

1. Your program must set the random seeds of python and tensorflow to 1 to make sure that your results are reproducible.
2. Your program will be tested by training on a fraction of 0.1 of the standard training set. The testing data will be the entire dataset.
3. The training and testing of your 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 57 training examples **x_train.csv** and **y_train.csv**
3. An example program **proj1.py**.
4. 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.

Grading

We will generate a random set of 57 training examples by running the program **fraction_xy.py** with a seed that is kept secret. If, for example, the seed is 7, the program is run as follows:

```
python3 fraction\_xy.py x\_test.csv y\_test.csv 0.1 7
```

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

What you need to submit

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