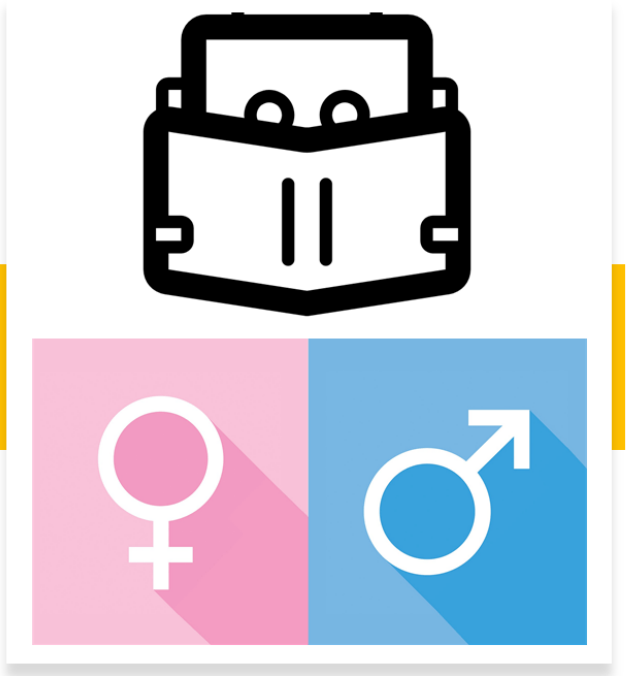

Gender Classification Model with Deep Learning

Yaara Barak & Matan-Ben Nagar

DEEP LEARNING
&
NATURAL LANGUAGE PROCESSING

PROPSPAL

Using the Cardiovascular Disease dataset from Kaggle (provided in this link: https://www.kaggle.com/sulianova/cardiovascular-disease-dataset) we will be trying to solve a classification problem.

The problem:

Gender classification based on person's height, weight, age, whether he smokes or not. How well will we be able to predict if a person is a male or a female based only on his phsycial attributes?

Divding the Dataset:

We will be splitting the dataset to 70% training and 30% testing as suggested in most common experiments of predicting a classifcation problems.

Work Progress

In this assignment, my partner and I have tried to establish a connection between different physical attributes of a person, to predict his sex. We were using the Cardiovascular Disease dataset from Kaggle (provided in this link: https://www.kaggle.com/sulianova/cardiovascular-disease-dataset).

Dataset:

- Our dataset contains 70,000 examples
- Each example holds five features: [age, gender, height, weight, smoke]

Training and Testing

- 50K examples are used as a training set
- 20K are being used for testing

```
(             age  gender  height  weight  smoke
 id
 0       51.091667       1     168    62.0      0
 1       56.188889       0     156    85.0      0
 2       52.380556       0     165    64.0      0
 3       48.952778       1     169    82.0      0
 4       48.538889       0     156    56.0      0
 ...           ...     ...     ...     ...    ...
 99993   53.444444       1     168    76.0      1
 99995   62.780556       0     158   126.0      0
 99996   52.961111       1     183   105.0      0
 99998   62.308333       0     163    72.0      0
 99999   57.055556       0     170    72.0      0

 [70000 rows x 5 columns],
 0    45530
 1    24470
 Name: gender, dtype: int64)
```

```
training_set = df2.iloc[-20000:]
testing_set = df2.iloc[50000:]
```

Logistic Regression Results

CNN Results

Comparison