Classification of buyers

by: Leonardo Vázquez Rodríguez

Abstract

In this project I am going to show one classification method, using a basic deep learning model, to define if a possible purchaser either buy or not. For this I used a data set from Kaggle [1] to extract all the possible features used by the model and try to make the best prediction, without go to over fitting.
This model works well and give me an accuracy of **90 %** and to see with more detail I am going to show the confusion matrix of the model.

1 Objective

- Be able to determinate based on the features if the people going to buy or not.
- Determinate the features what make the have more likelihood to the people buy.

2 Analysis

In this project I am going to use a deep learning model and the sequential core have 4 dense with 3 relu activation and 1 sigmoid activation.

Is important to see that before to do the model implementation is necessary to clean, restructure and split the data.
Clean the data is important because maybe not all the columns in our data frame have a features that are useful to build our model work and make the best predictions so, we need to know how is the data to make this choose.
Restructure is really important to seek good result of the model because the data seldom have a good structure to be implemented in the model, so before to start I usually normalize the data to have all the features values in the range of zero and one.
Slip the data in train and test is necessary when we train the model, because we don't use all the data to make this train, as it will be over fitting and actually don't work to make predictions, so we need to split the data. In this case I decided to split in 30% to train and 70% to test.

Results

This is the model summary.

```
Model: "sequential"

Layer (type)                  Output Shape             Param #
=================================================================
dense (Dense)                 (None, 32)               128

dense_1 (Dense)               (None, 32)               1056

dense_2 (Dense)               (None, 32)               1056

dense_3 (Dense)               (None, 16)               528

dense_4 (Dense)               (None, 1)                17
=================================================================
Total params: 2,785
Trainable params: 2,785
Non-trainable params: 0
```

After to training and testing the model I obtained 90% of accuracy, to that I make the plot of confusion matrix to see this with more detail.
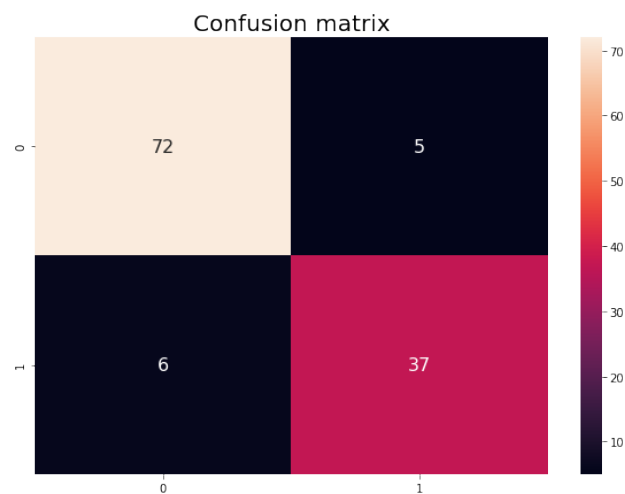

Figure 1: Zero indicate the gender male and one female

With the before figure we can see that the 77 people that didn't buy, 72 was good predicted and 5 was wrong. And 43 people that did buy, 37 was good predicted and 6 was wrong.

3   Conclusions

We can see that the model works well can be useful to market strategy and also this model can be enhanced with more buyers information to work. This is a easy model to get start in deep learning and maybe this is not sophisticated but it is works and have a lot things that make a useful model.

References

[1] https://www.kaggle.com/datasets?search=buyers&datasetsOnly=true

[2] https://jupyter.org

[3] https://matplotlib.org

[4] https://pythonhosted.org/uncertainties/

[5] https://pandas.pydata.org/docs/

[6] https://numpy.org/doc/