**PROGRAM DESCRIPTION**

*Program Introduction*

This is my attempt at using neural networks to predict whether an individual is looking for a job. I’ve used google colab to complete this assignment and have written the required code all in a single file, which you can see on my github in the file named code. Please note that due to my unfamiliarity with TensorFlow, I have made use of several resources all of which can be seen in my code and listed in this document under references.

*Code Description*

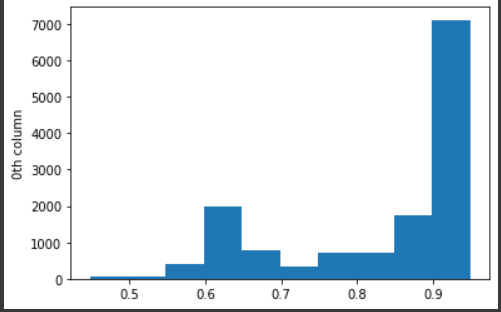
I have opted to not include the coding for cleaning the original csv file. The code has used one hot coding to replace pretty much all the values with either 0 or 1. After saving my dataframe, I saved the target value for later use, then removed it from the datafame. I then stratified the variables to prevent high bias and split the test and train data set so by .85/.15. Once that was done, then split the training file again with the validation file so the validation file would receive .15 from the remaining test set. After standardizing the variables, I began building my neural network.

**NEURAL NETWORK**

*Forward Feeding Neural Network*

I normalized my data before creating the layers for my network. It may not have been needed; however, I was hoping it would help with my accuracy a bit if done.

*Before standardization After standardization*

 Chart, histogram

Description automatically generated

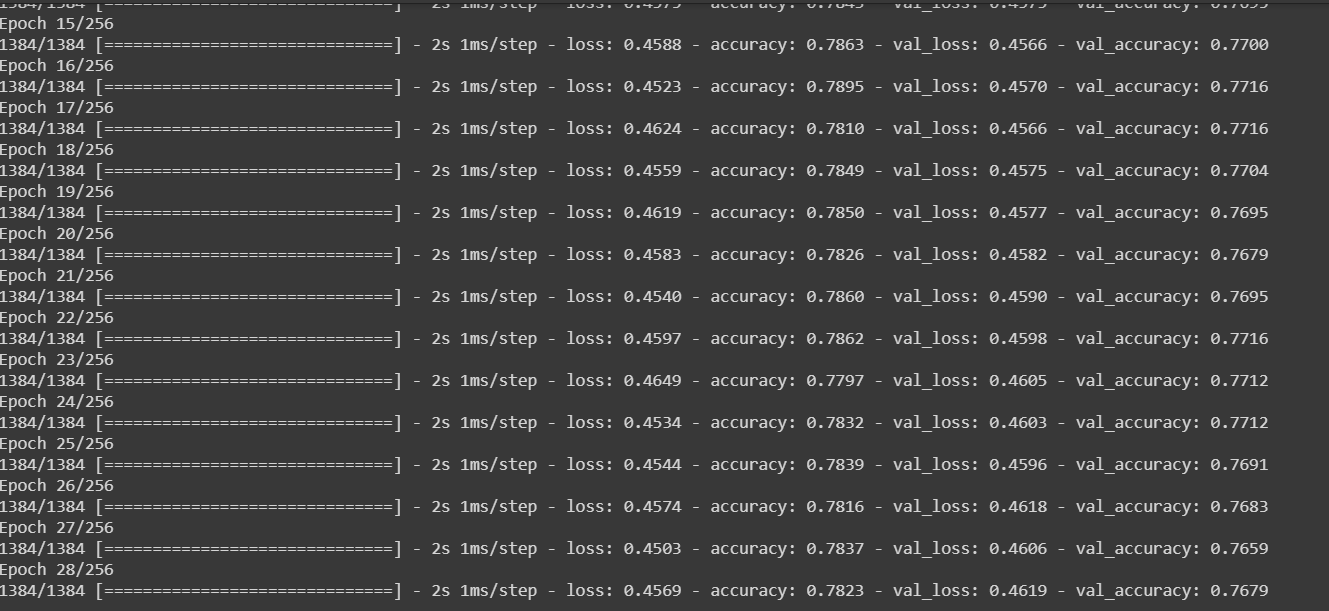
The input layer will have 8 neurons, there is one hidden layer with 4 neurons.I attempted to create an outer layer with 2 neurons using softmax, however, I was not able to get it to function, so I swapped to using 1 neuron in the output layer and using sigmoid. I then compiled the data using binary\_crossentropy and adam as the optimizer. When the model was completed, the output was an accuracy ranging between 78% - 80%.

*Model Summary*

*Graphical user interface, text

Description automatically generated*

*Accuracy*

**

*Neural Network Layout*

*Chart

Description automatically generated*

*Cross Validation*

Unfortunately, I was unable to get my code for cross\_validation to work. My intention was to use K-fold to use my input features, create 2 hidden layers with 8 nodes each and create 2 output nodes that use softmax activation function. After running the model my goal was to use a batch\_szie of 10 and run 150 epochs, after which I would try to use scikit-learn to calculate acracies via cross\_val\_score.

I would then take the mean and std of the accuracies and continue formatting from there. You can see the attempt at the very bottom of the code.

**PROGRAM ISSUES**

*Possible Inaccuracies*

This value could come from several things and I am believing the values may be incorrect. Some issues that lead me to believe that excluding my inability to create 2 neurons in the output layer, may include how I formatted my files. I began with the one hot coding and worked from there. My inexperience or incorrect values may have caused inaccuracies very early on. My target values may have been saved incorrectly and produced a false accuracy score.

*How To Fix Inaccuracies*

If the accuracy score is closer to correct than I am led to believe, then there are a few issues that could be adjusted to achieve a better score. High variance is an issue where overfitting your model may seem to give good values but may not perform as well in other examples. Comparing the training set (which may contain 70% of the data) and measuring its error rate to the remaining 30 percent may allow you to see if the model has low error in the training set but high values in the test. This is an indication of high variance and should be fixed. Having low error rates for both the testing and training indicate the the dataset is nicely balanced.

**REFERENCES**

*Sources Used*

Keras - models. (n.d.). Retrieved May 09, 2021, from https://www.tutorialspoint.com/keras/keras\_models.htm

Khandelwal, R. (2020, November 11). Building neural network using Keras for classification. Retrieved May 09, 2021, from https://medium.datadriveninvestor.com/building-neural-network-using-keras-for-classification-3a3656c726c1

Live coding : Pytorch feed forward neural network on diabetes dataset. (2020, June 03). Retrieved May 09, 2021, from https://www.youtube.com/watch?v=hPOiKNADhaQ

MidiMidi 14311 gold badge11 silver badge77 bronze badges, Zipazipa 24.2k66 gold badges2929 silver badges4949 bronze badges, FlairFlair 80111 gold badge99 silver badges2222 bronze badges, & Dr\_dronychdr\_dronych 5133 bronze badges. (1966, January 01). How can I split a dataset from a .csv file for training and testing? Retrieved May 09, 2021, from https://stackoverflow.com/questions/43697240/how-can-i-split-a-dataset-from-a-csv-file-for-training-and-testing

Split dataframe Into train, Test, validation sets using scikit-learn. (2020, October 05). Retrieved May 09, 2021, from https://www.youtube.com/watch?v=ZHPwPHkrT4I&t=408s

TensorFlow tutorial 3 - neural networks with sequential and Functional API. (2020, August 10). Retrieved May 09, 2021, from https://www.youtube.com/watch?v=pAhPiF3yiXI&t=207s

Tutorial 14: K-fold cross validation using Keras Python | k-fold Cross-Validation in neural network. (2020, January 01). Retrieved May 09, 2021, from https://www.youtube.com/watch?v=lFRuAAxkz7A&t=242s

Tensoflow Playground. (n.d), Retrieved May 09, 2021 from

https://playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=circle&regDataset=reg-plane&learningRate=0.03&regularizationRate=0&noise=0&networkShape=8&seed=0.99168&showTestData=false&discretize=false&percTrainData=80&x=true&y=true&xTimesY=false&xSquared=false&ySquared=false&cosX=false&sinX=false&cosY=false&sinY=false&collectStats=false&problem=classification&initZero=false&hideText=false