CSE474/574: Introduction to Machine Learning(Fall 2018)

Instructor: Sargur N. Srihari Teaching

September 3, 2018

**Project 1.1: Software 1.0 Versus Software 2.0**

Due Date: Monday, September 17

Name: Alex Zhong

UBID: zechengz

Person Number: 50040948

Target

In this assignment, we need use Keras and TensorFlow to find which is Fizz, Buzz or fizzbuzz in 1 to 1000 number. The testing data have 1 to 100 number to test these two models. And training data is 101 to 1000 to training these two models. We have two outputs excel files are output 1 and ouput2.

Background

Back Propagation

It is a multi-layer feedforward neural network trained according to the error back propagation algorithm, and is the most widely used neural network. Its basic idea is the gradient descent method, which uses gradient search technology to minimize the error mean square error between the actual output value of the network and the expected output value.

**Input output**

X1 dropout Xt1 Software Y1

X2 Xt2

X3 Xt3 Y2

. .

. . Y3

. .

X10 Xt256 Y4

Overfitting：

Overfitting is an important concept in classifier operations. Overfitting is the process of making a hypothesis overly strict in order to obtain a consistent one. Avoiding overfitting is a core task in classifier design. The performance of the classifier is evaluated by increasing the amount of data and testing the sample set.

Keras:

Keras is a deep learning framework based on Theano, which references the Torch. Written in Python, it is a highly modular neural network library.

TensorFlow:

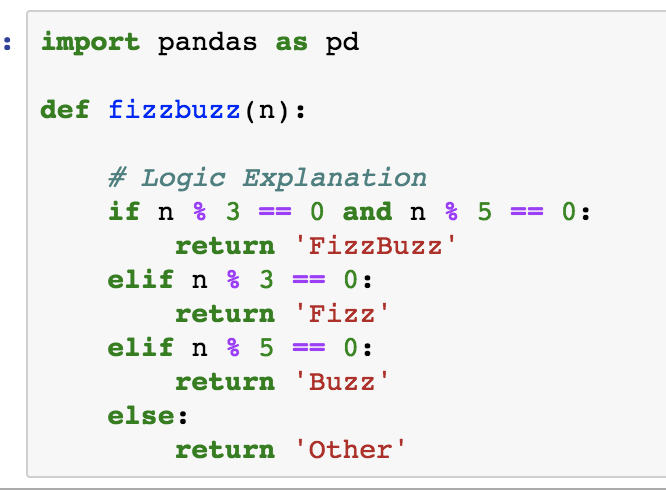
Tensor means an N-dimensional array, and Flow means a calculation based on a data flow graph. TensorFlow is a process in which tensors flow from one end of the flow graph to the other. TensorFlow is a system that transmits complex data structures to an artificial intelligence neural network for analysis and processing.

Code analysis

Keras:

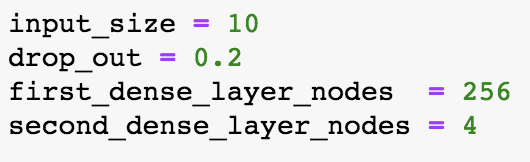
This neural network has three layers: 1. Input and dense 2 add layers and full connect 3. Output four category

1.0

1. 

First part code is activation function. And that is the main code for this project, because this part code can find which number meets the four types in the requirements and classifies. So, this part also has the role of a classifier.

1. Second part code is about input data and output data. At input data, we create two datasets are testing data and training data. Because in the following model, we need use training data to parameter training, and then use the test data to verify the performance of the model. This part use DataFrame because that can orderly storage and easily called.
2. Third part code is about data processing. This part does some processing on the data here, and it does a total of 10 bit operations to get a list, and then it forms an array. And we have input size=10 in the model defined later so we need to use number 10 at here.
3. Fourth part code is model definition.



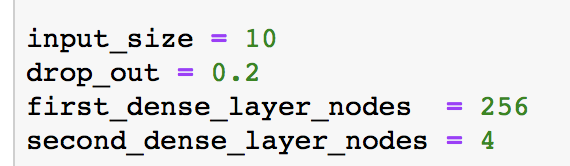
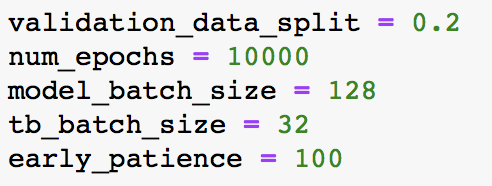
Adjustable parameter: The input size can change to 20 or 30.

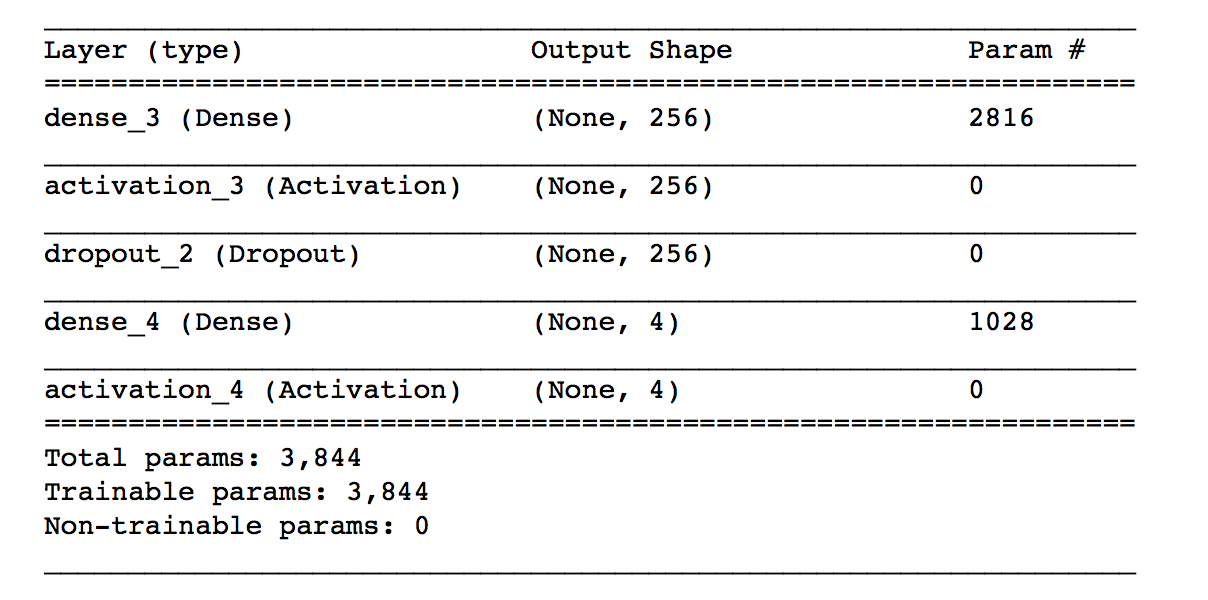
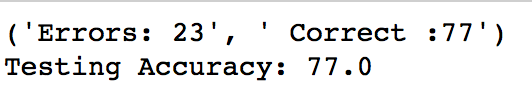
The drop out can change to 0.1-0.5. The first dense layer nodes can change to 64, 128, 512. Because we have four classify, the second dense layer nodes cannot be change.

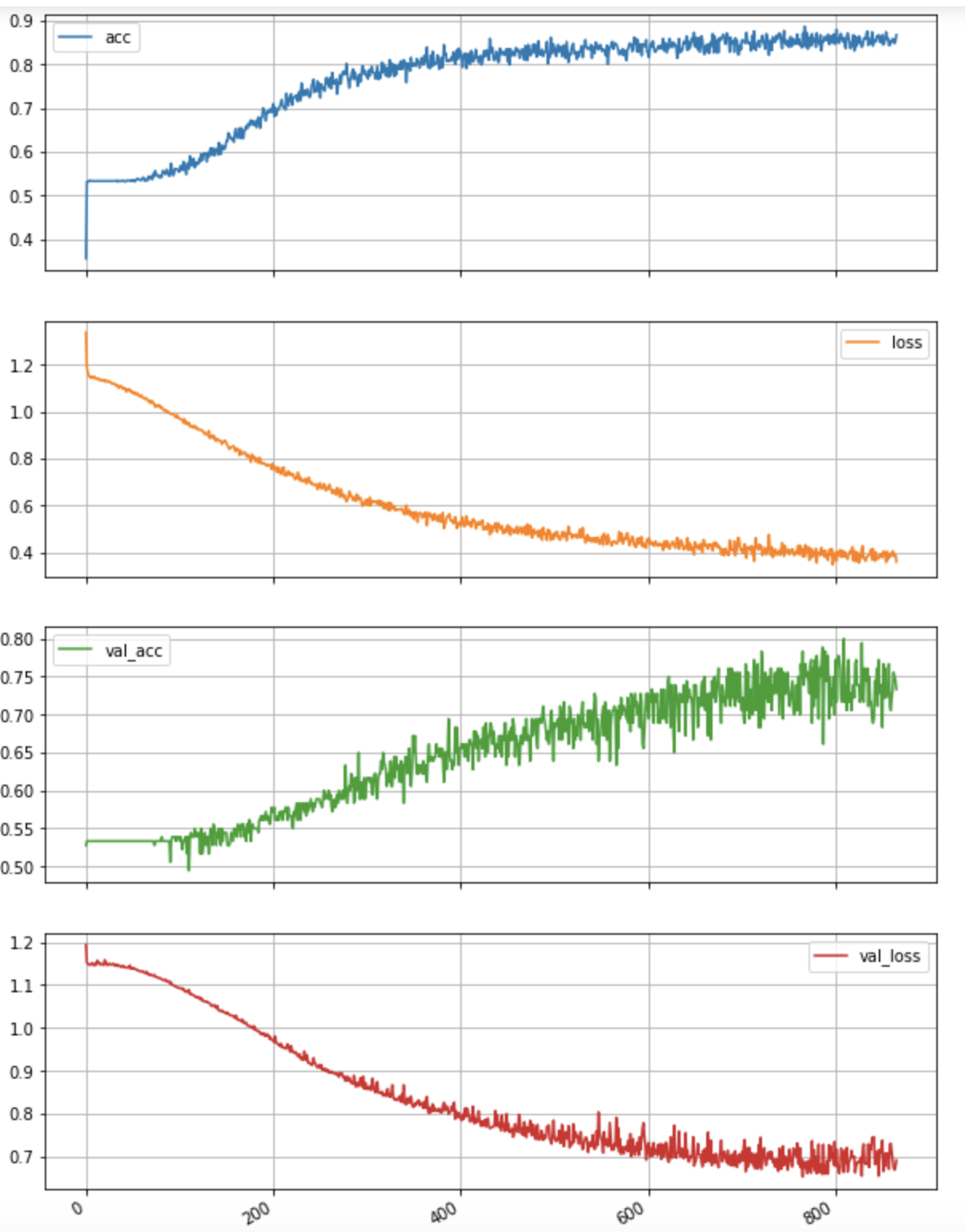
At follow code, we need define a neural network to calculate. In order to avoid overfitting, we need to set the number of network layers (Include pooling layer and enlarging layer). And I write down some background knowledge with overfitting. It uses dropout also to prevent overfitting. Because we are a major part of the classification model, so in this place use to Softmax.

1. Adjust the parameters to test model performance

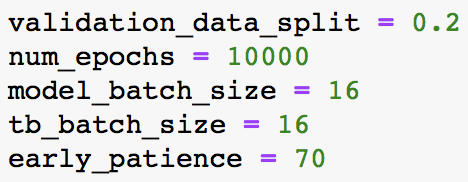
The first:

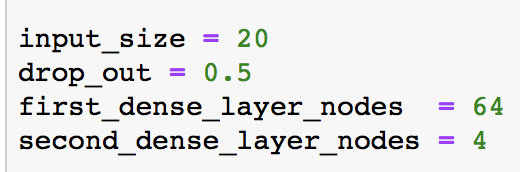


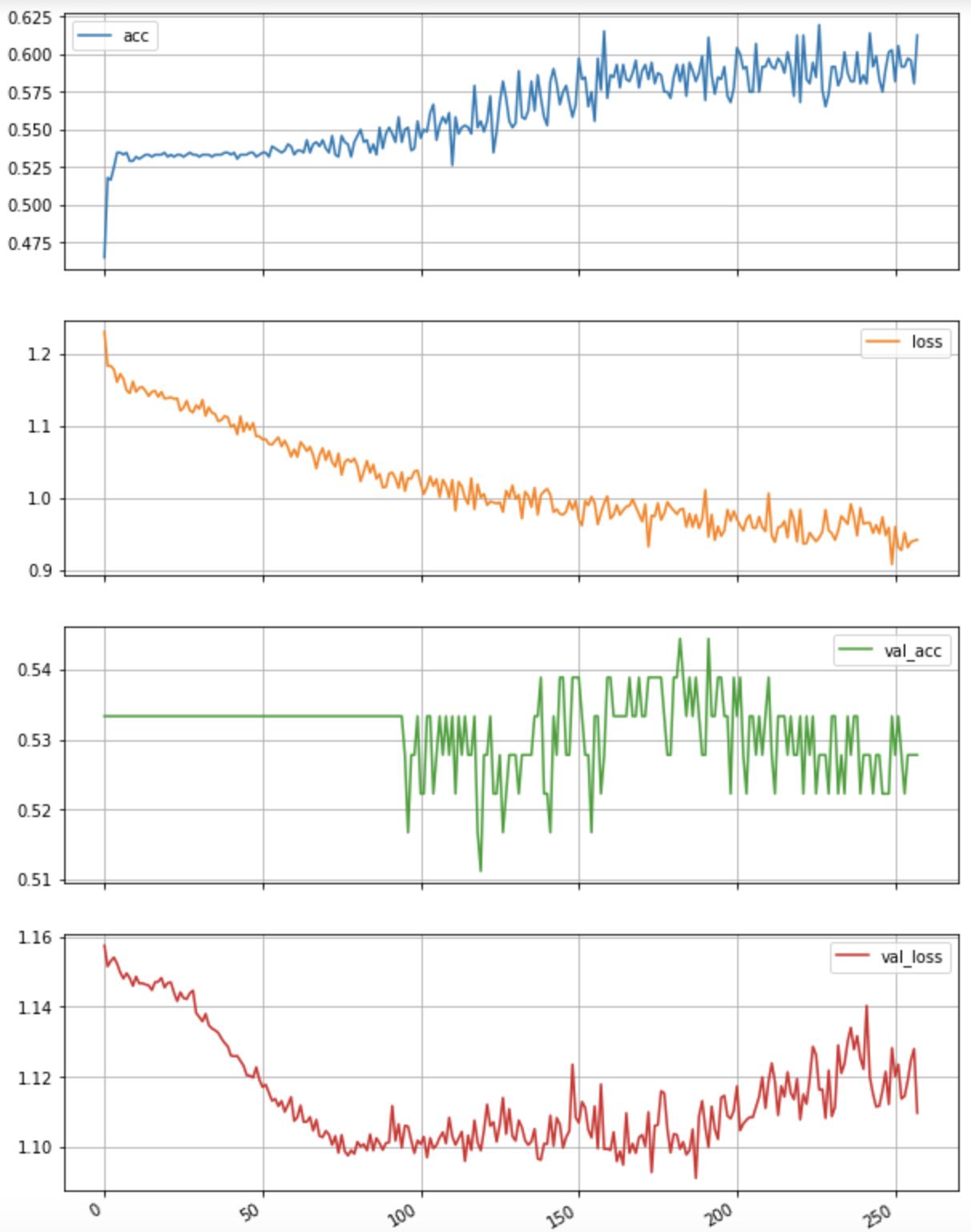
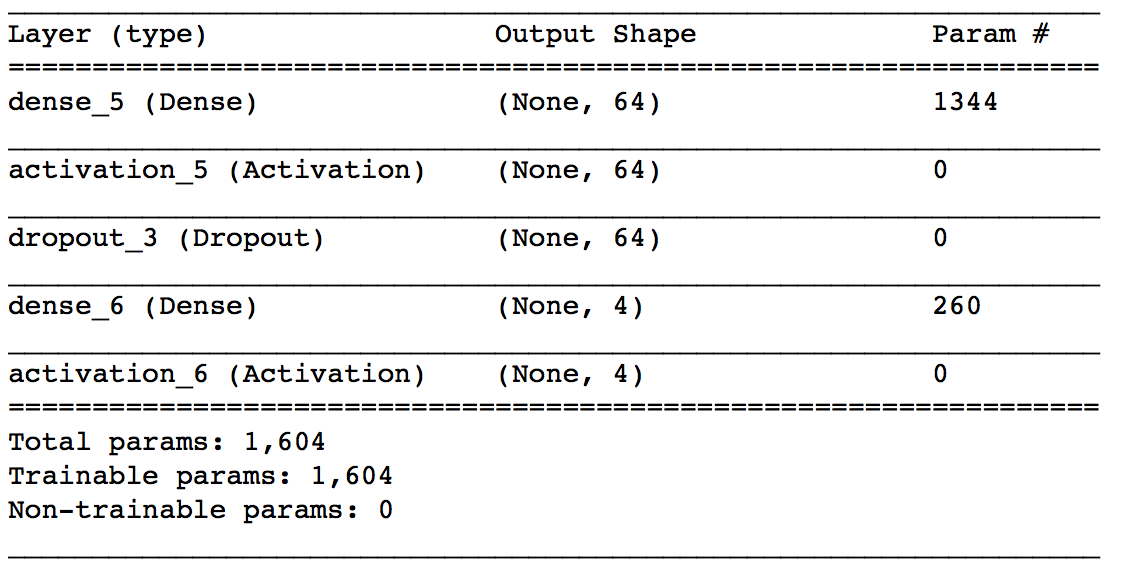
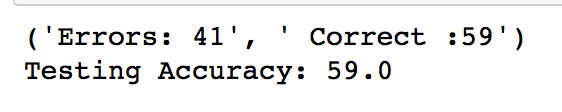




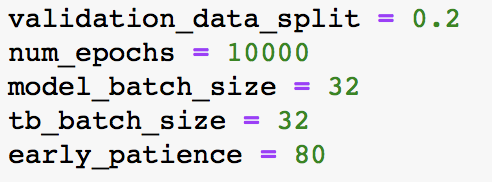
这个是第一次的模型参数结构，比较合理，其中训练的误差是逐渐降低的，直到600次训练次数以后就开始波动，趋于稳定了，准确率也在85%左右就上下波动。

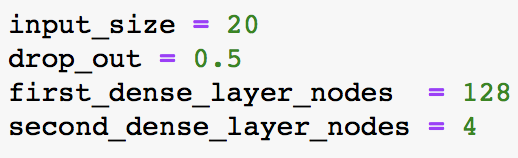
The Second:

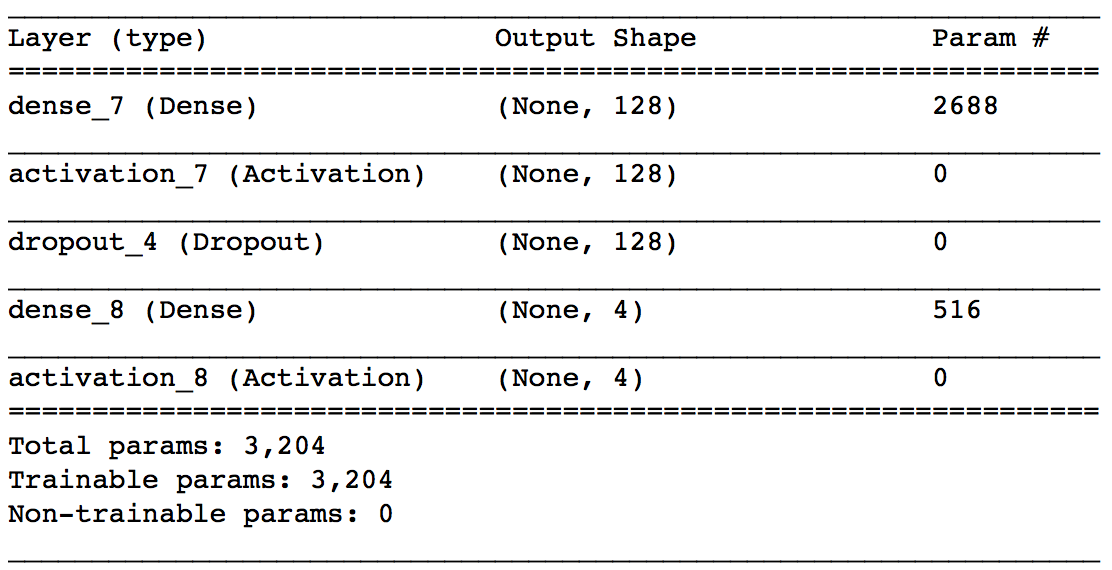
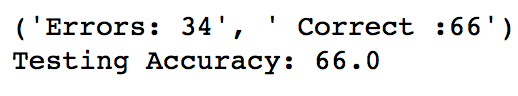


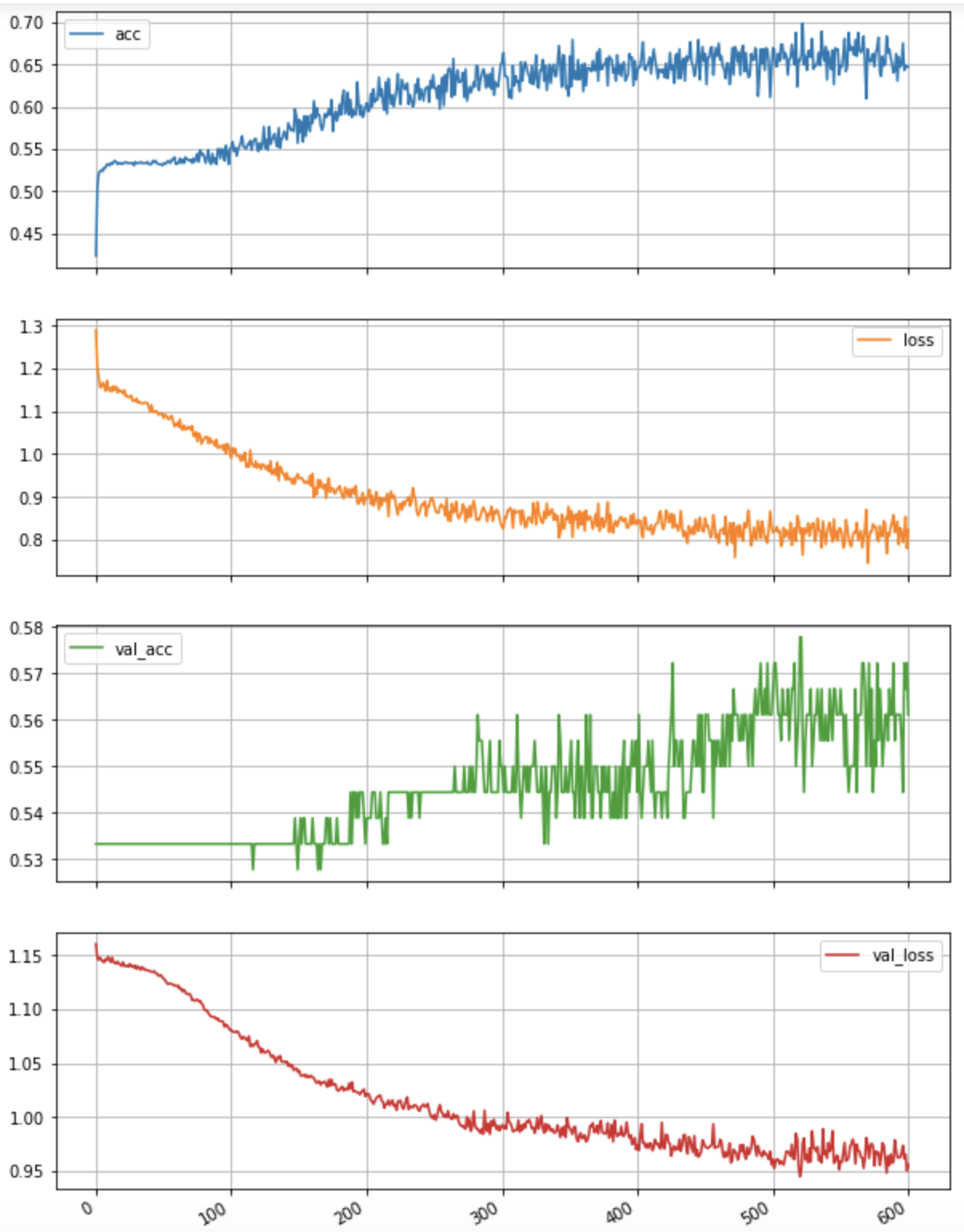


第二次的参数可以发现，loss 没有一个收敛的趋势，尤其是红色这个曲线，loss还在比较明显大幅度的波动，而准确率也是只有0.6左右，说明训练次数不够250次太少了。相比第一次的800次 迭代训练，还不不够。

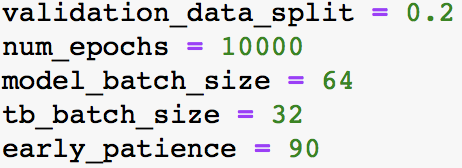
The Third one:

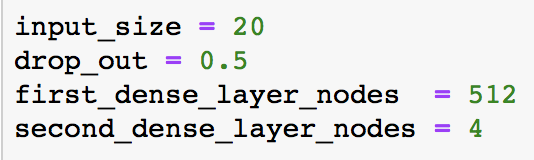


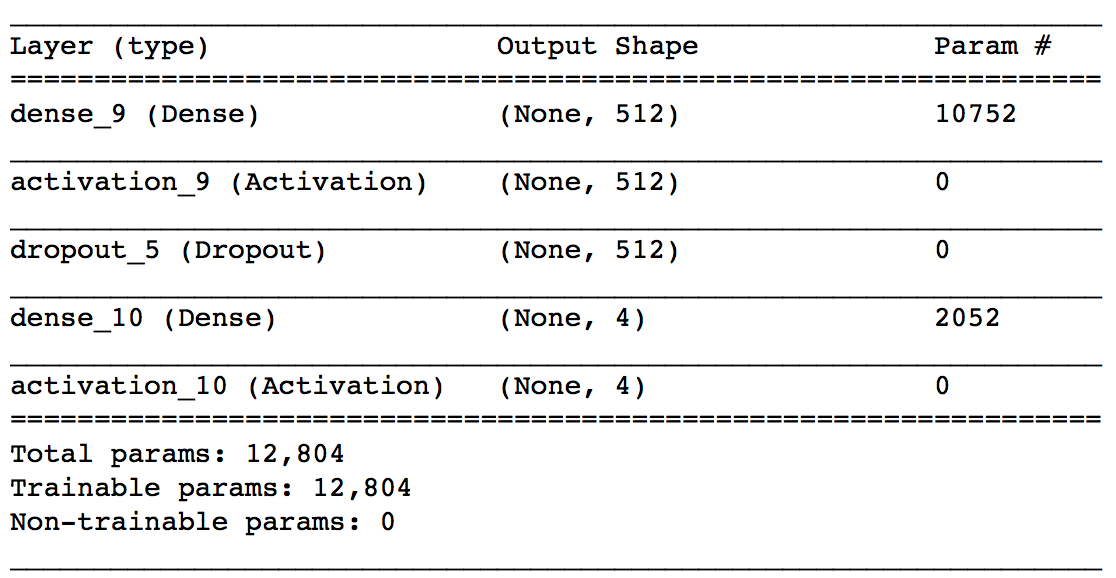
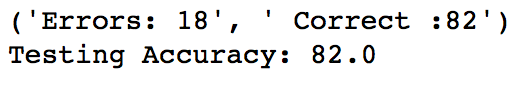


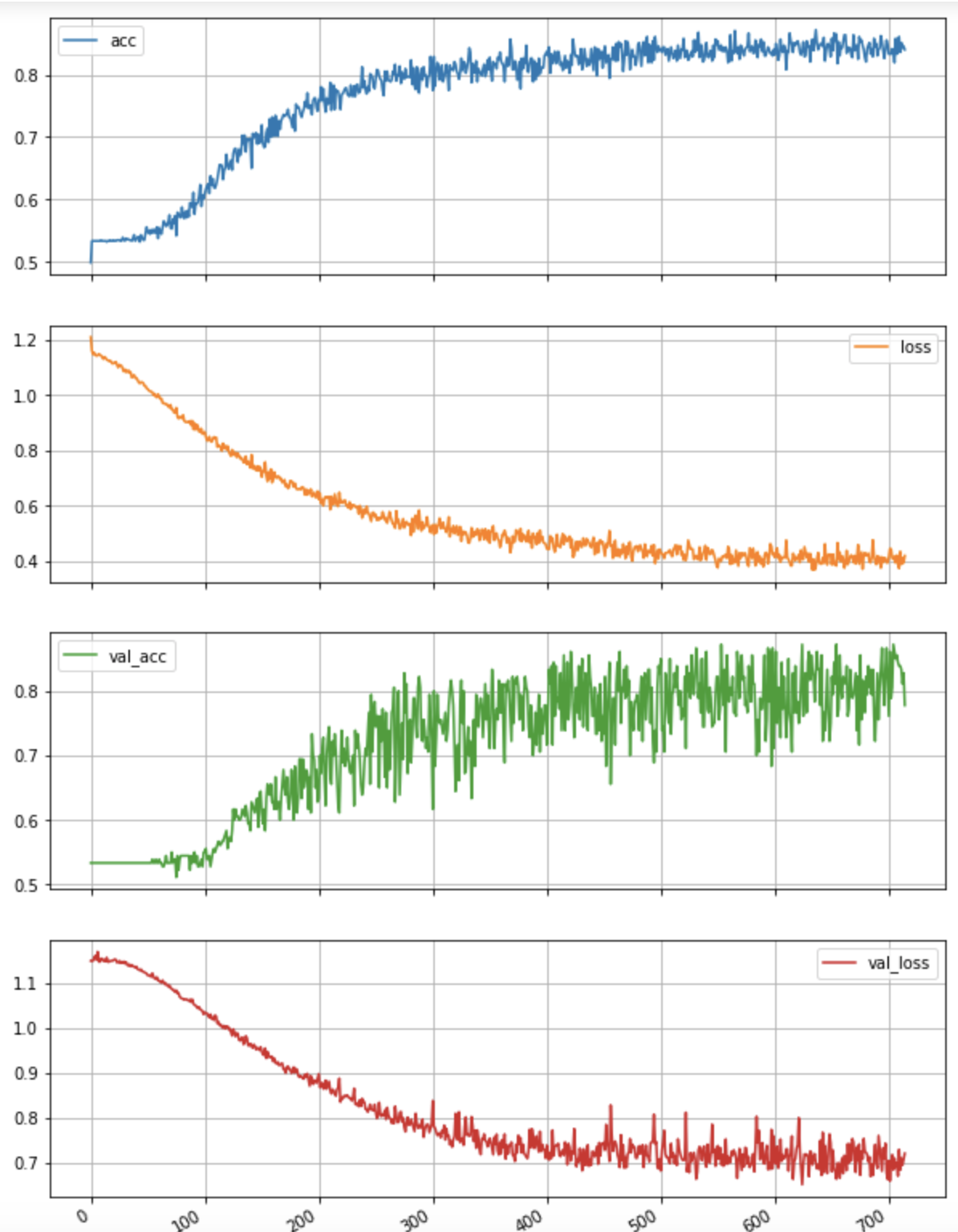


第三次这个模型，准确率和loss都趋于收敛，但是模型的参数可能有些不太对，比如dropout=0.5 太大，可能就导致了效果没有第一个模型那么好，以及第一层layers=128，比第一次的模型少了128，这些参数的改变都会影响模型的最终精度。

The fourth:







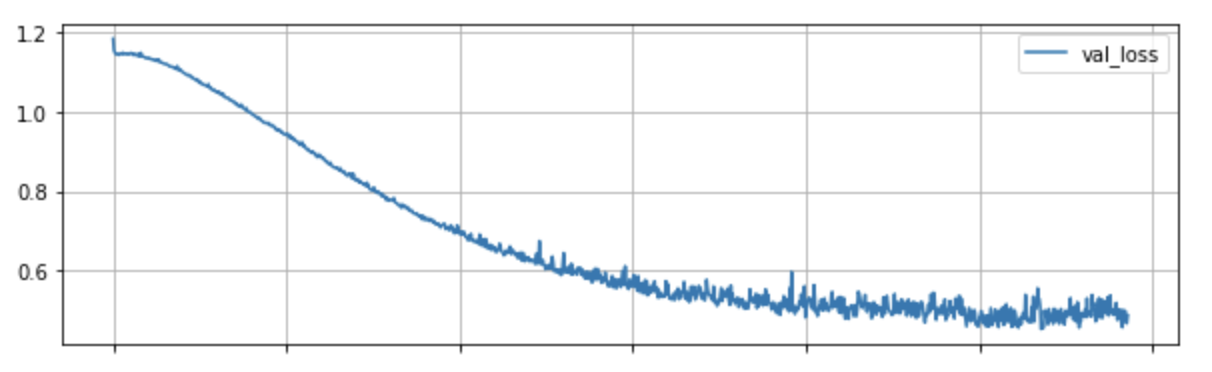
第四次，的训练结果，综合前面的参数调整以及分析，最终选择了这样的参数，得到了准确率为82%，是四个模型里面最好的。

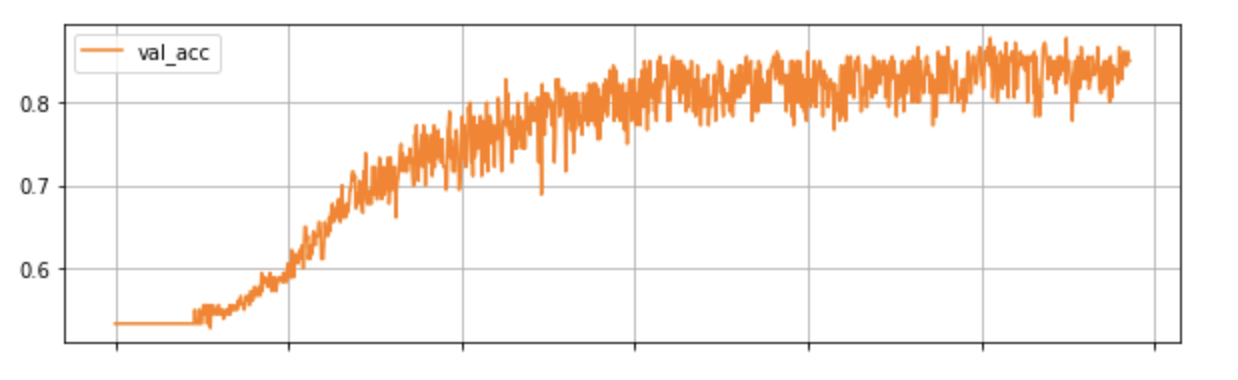
通过这四个模型参数的调整，可以发现神经网络模型的难点就在于调整和训练得到不同的参数，用来提高模型预测的准确率，合理设置模型层数，以及连接的属性是很重要的事情。

这两个结果不知道怎么解释。。。

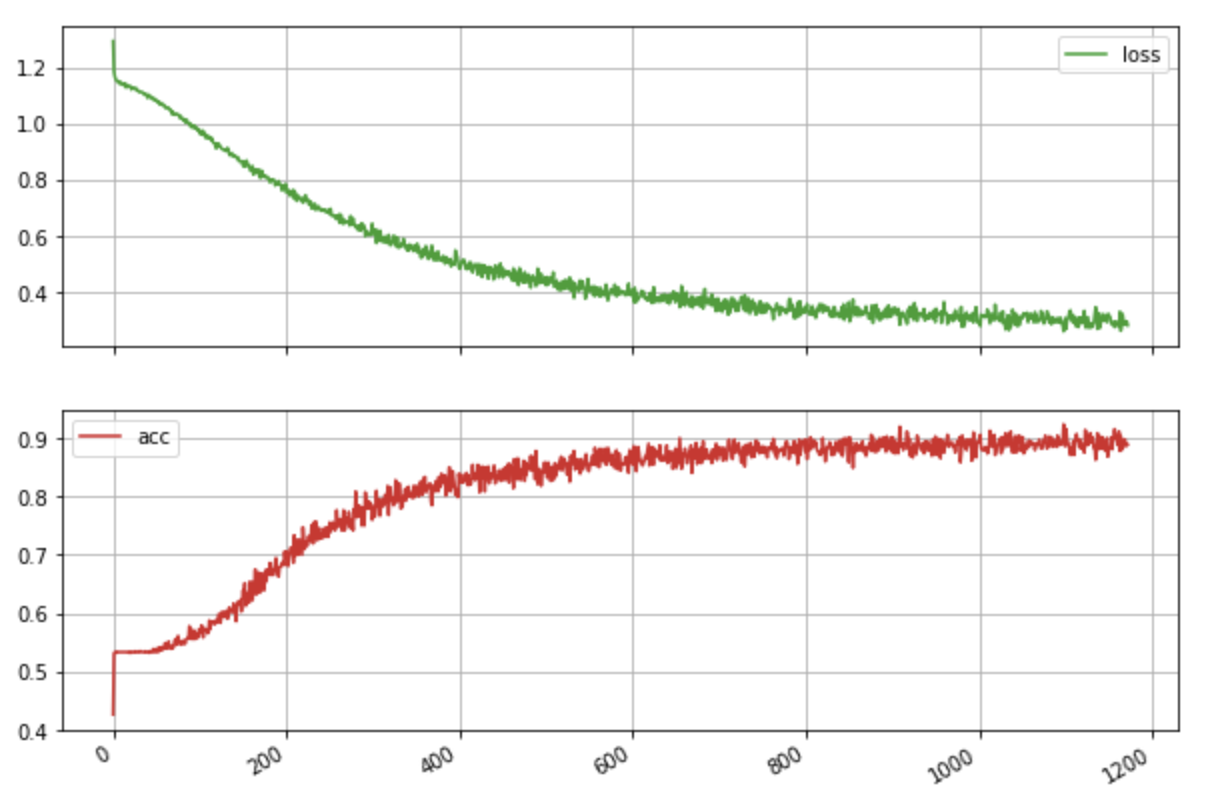
1. Training and validation graph

This code use RMSE to do the evaluation. RMSE(Root Mean Square Error): 1. It is the square root of the ratio of the square of the deviation between the observed value and the true value and the number m of observations. 2. Is used to measure the deviation between the observed value and the true value.



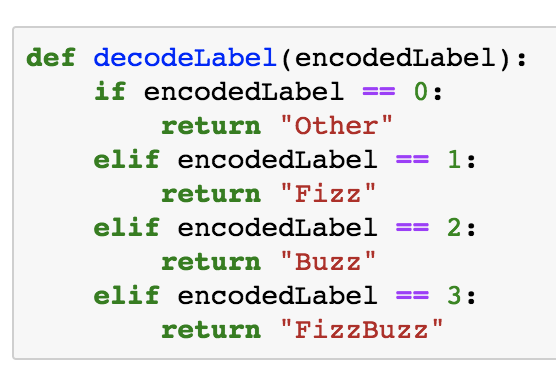


验证集



测试集

运算正确率显示。分类运算结果在output文档中



This part code gives each case label, the other label is 0, the fizz label is 1, the buzz label is 2, the fizzbuzz label is 3.

TensorFlow：

1.0

1. logic based fizzbuzz function, this part will use if and elif to calculate which is fizz, buzz or fizzbuzz. Create training and Testing datasets in csv format, this part we can get two datasets input data and output data. The data processing and give label was same with keras

2. TensorFlow model definition.

这个部分开始就不太会写了需要帮助

Reference：

<https://baike.baidu.com/item/BP%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C/4581827?fr=aladdin>

<https://baike.baidu.com/item/TensorFlow/18828108?fr=aladdin>