### Project 1

Student ID: 518030990025

Name: Eduardo Wang Zheng

E-mail: eduardo@sjtu.edu.cn

1. Introduction

Given a dataset of drug molecules with their SMILES expressions and the binary labels indicating whether one drug molecule is toxic or not. I develop a Deep Neural Network which can learn useful patterns from the data provided and predict the toxicity of a new list of molecules based on learned knowledge using TensorFlow package.

1. Experiments
   * 1. Analysis

I divide the task into two steps (Training and testing):

1. Train the model on the data under the folder “train” and validate its performance (by calculating AUC) on the data under the folder “validation”
2. Restore model parameters and then test the model on the marking data (“../test/”), and output predictions into a file, which indicating that the probability that current drugs are toxic.
   * 1. Code

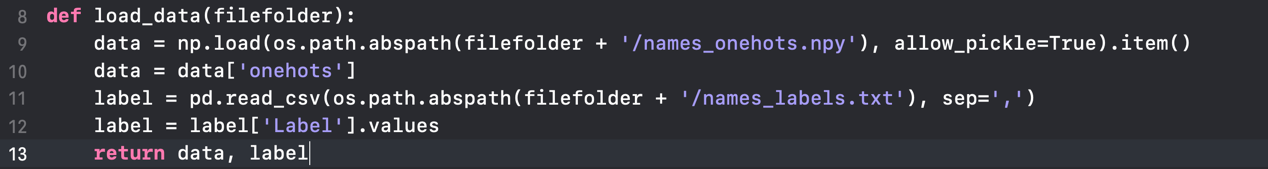


Figure 1 Load training data

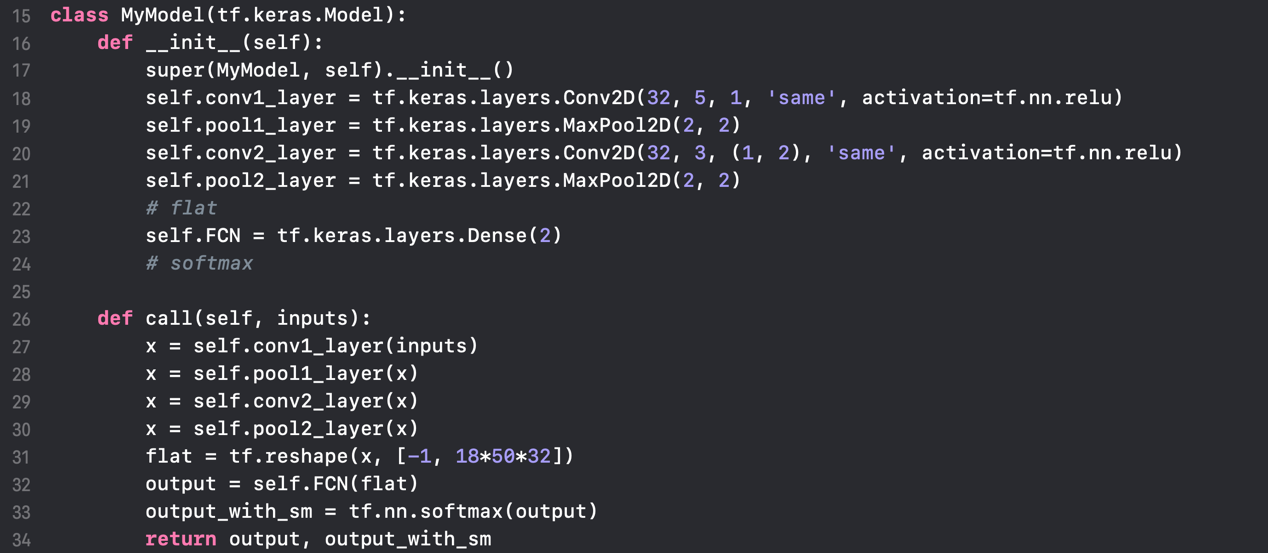
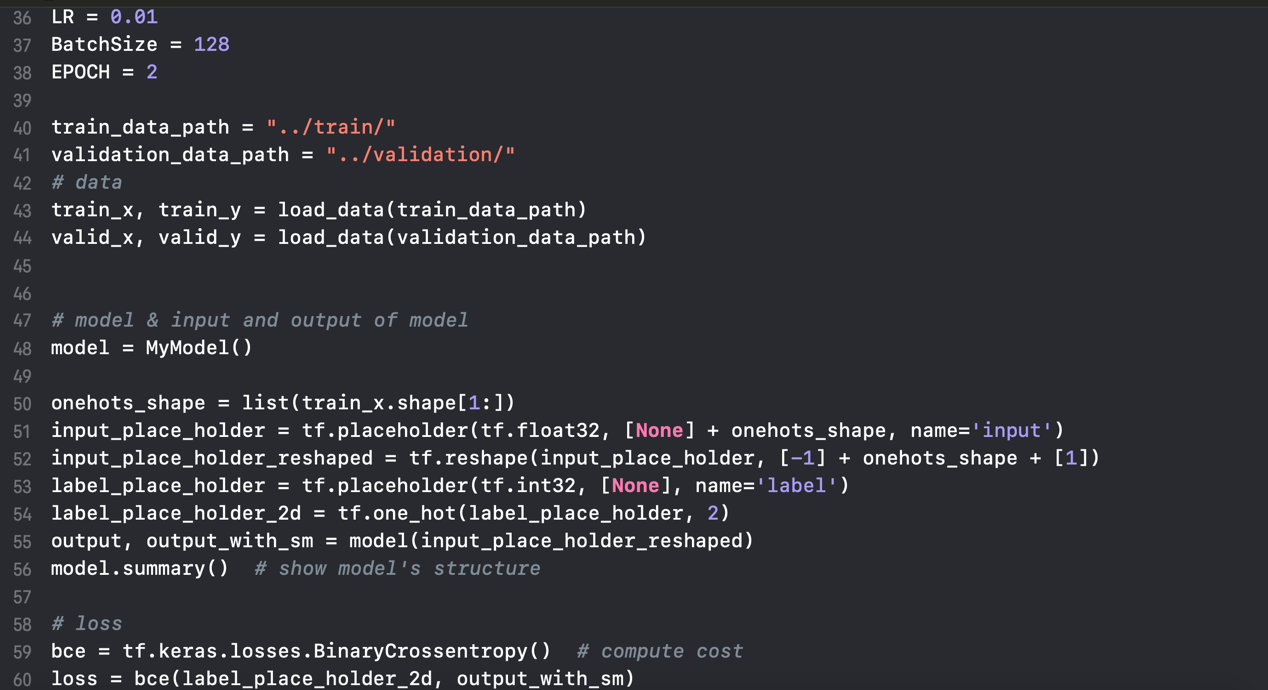
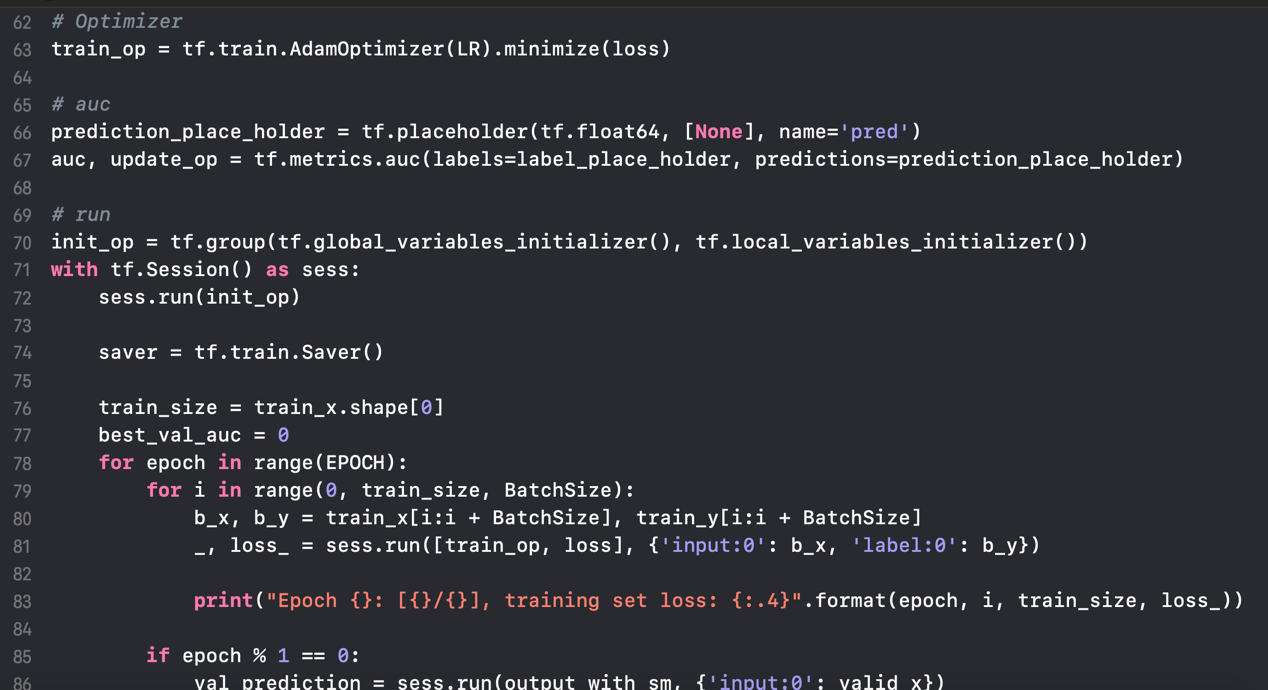


Figure 2 Define model





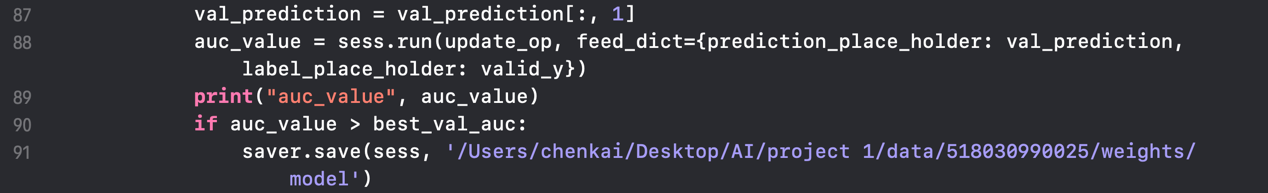


Figure 3 Train and save model

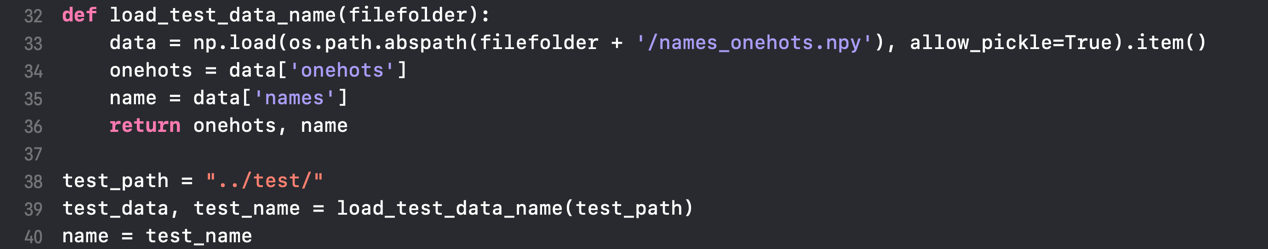


Figure 4 Load testing data



Figure 5 Load previously saved model

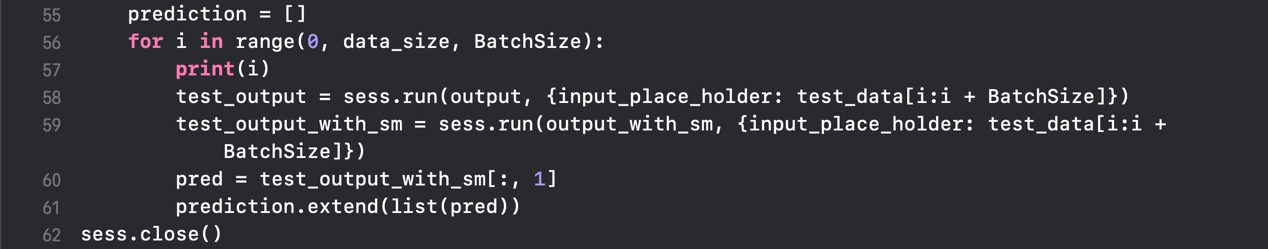


Figure 6 Make prediction

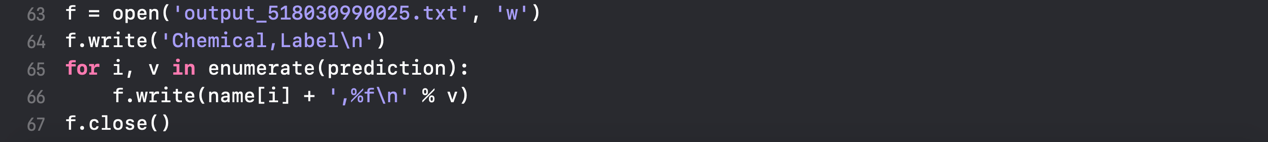
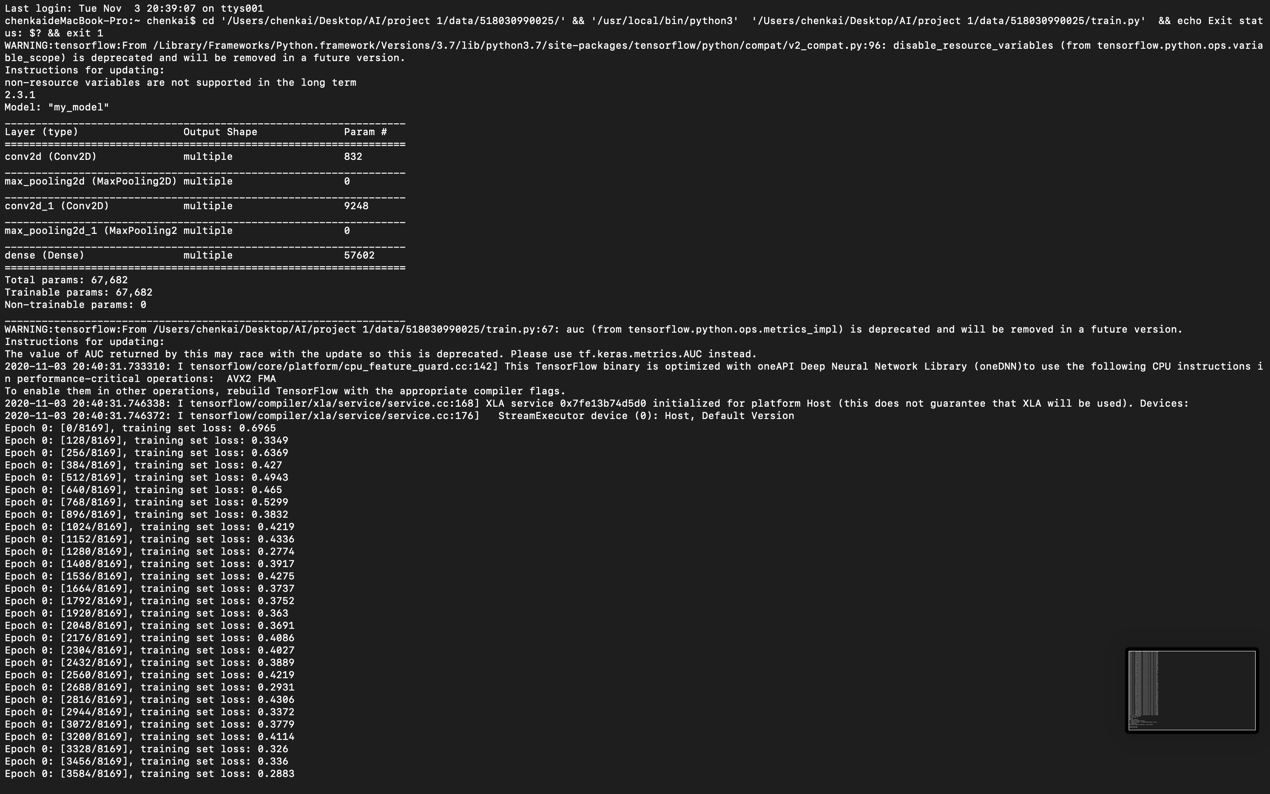


Figure 7 Output predictions into a file

* + 1. Evaluation



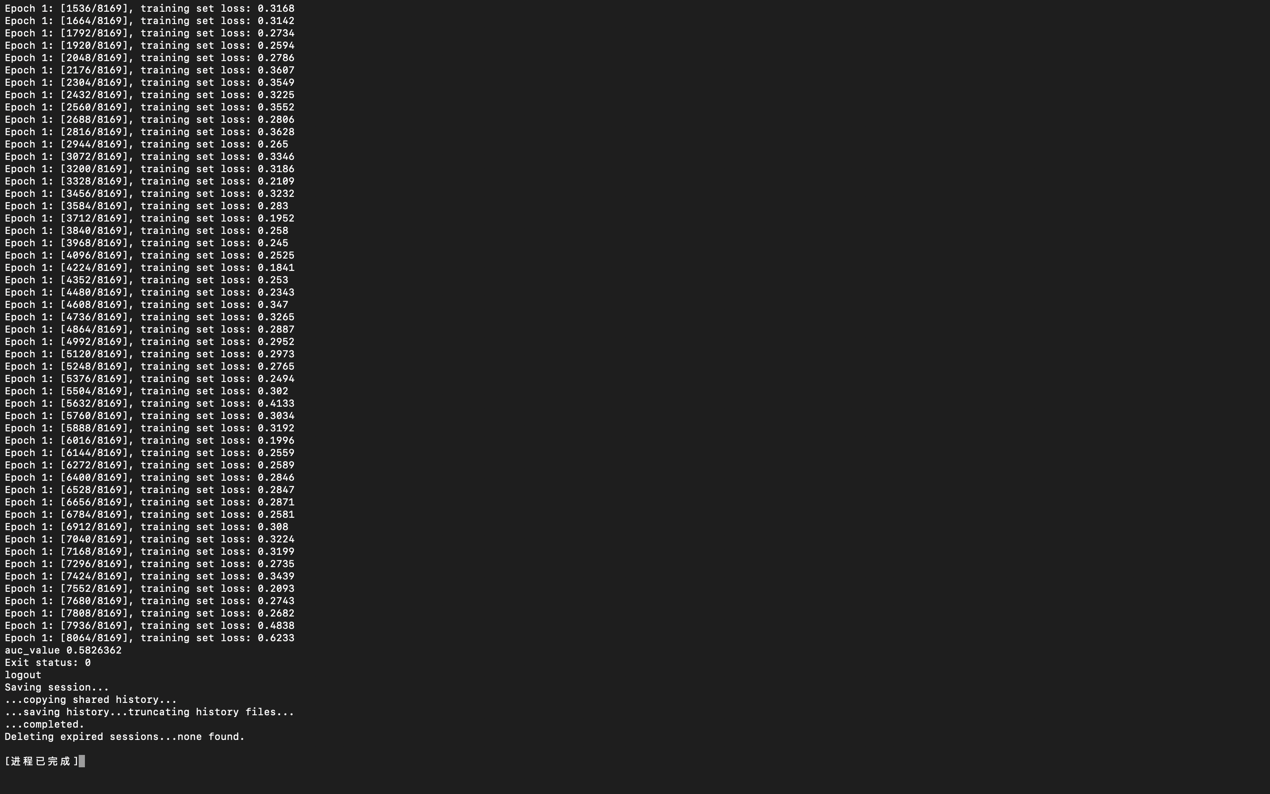


Figure 8 Training output

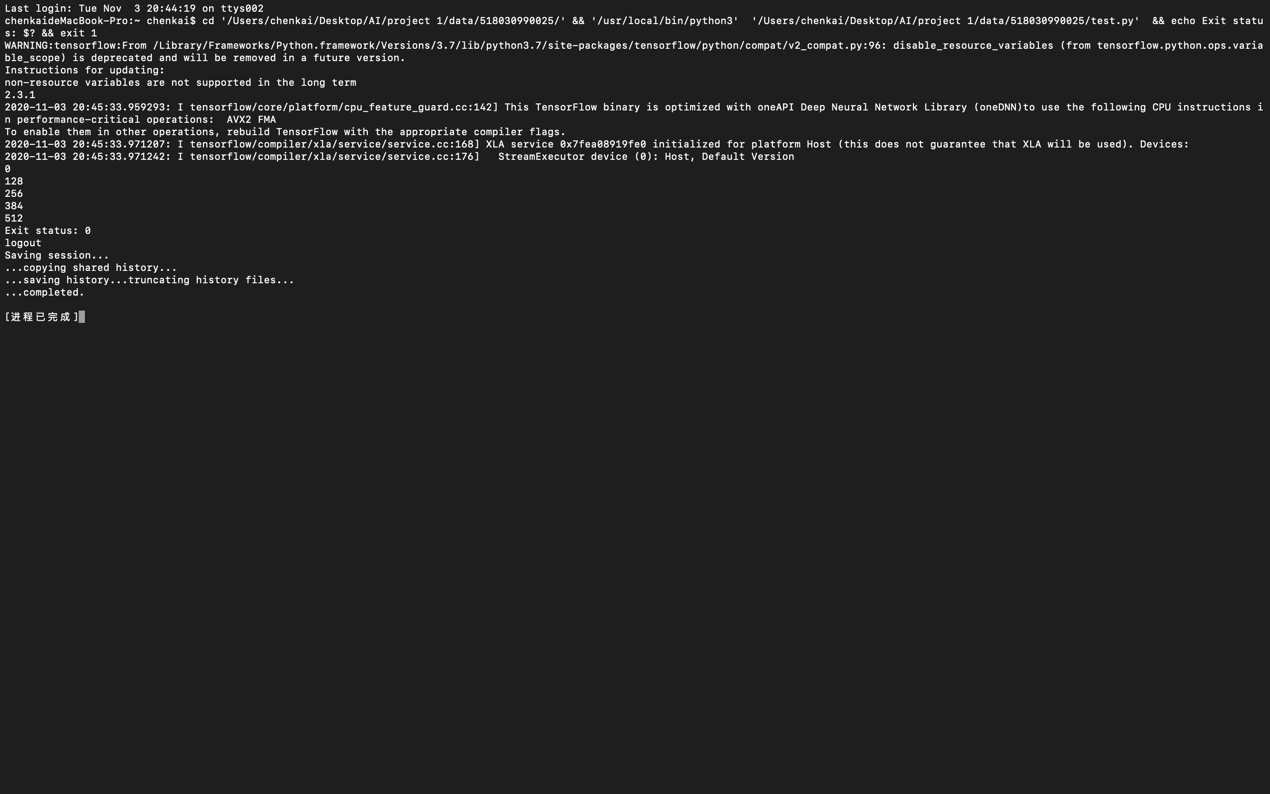
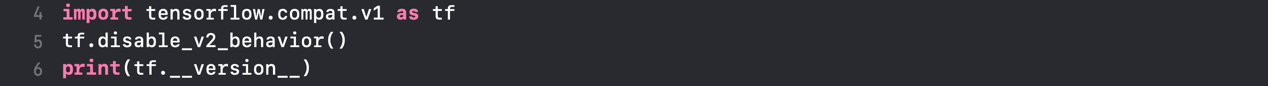


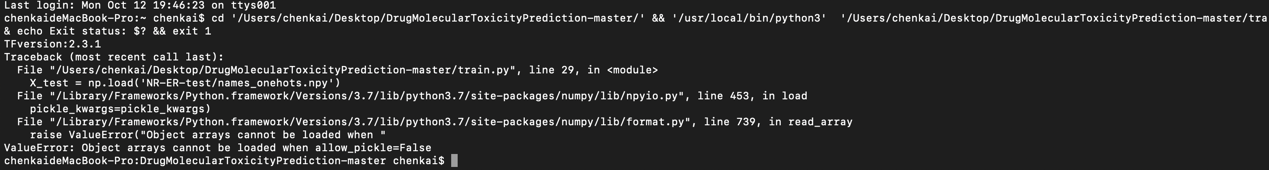


Figure 9 Testing output

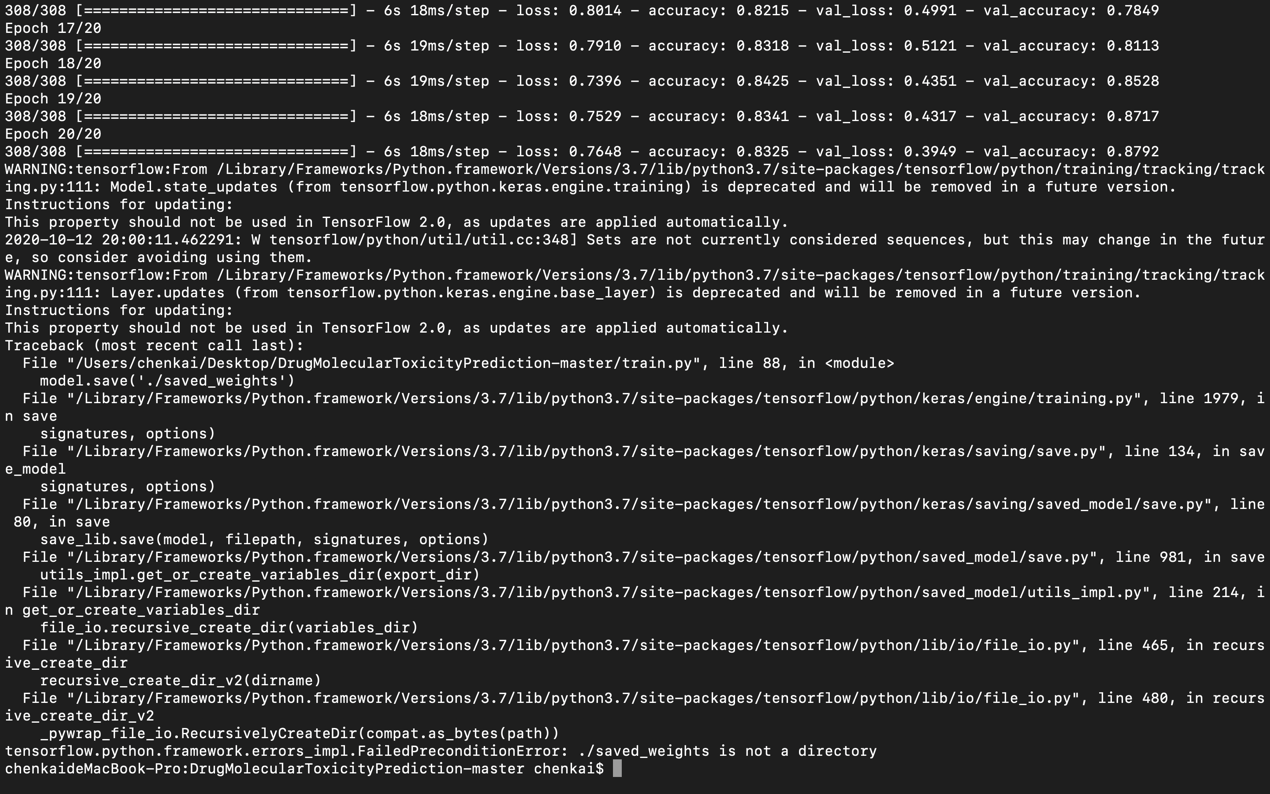
1. Conclusion
   1. Problems
2. **Problem:** Though I import the following package to make sure TensorFlow version is 1. The output is still 2.3.1, which might be a problem.



1. **Problem:** When I first learn to write the code totally by myself, I encounter the following problems:



**Solution:** Set **allow\_pickle=true**



**Solution:** Correctly save the model in the appropriate position.

* 1. Achievements

From this project, I achieve the following achievements:

1. Learn the normal procedure of deep learning:
2. Load training data
3. Define model
4. Train and save model
5. Load testing data
6. Load previously saved model
7. Do the task
8. Get more experience in writing and debugging python code
9. Learn how to use the model to make predictions and output predictions into a file, which indicating that the probability that current drugs are toxic.
10. Learn how to optimize a model.
11. Learn how to compute AUC
12. Reference
13. <https://www.kaggle.com/qizhili/project1-cnn-demo>