Box Office Forecast

Zhangtao Xue

Xi'an Shiyou University

Chinese Academy of Sciences

October 7, 2020

Overview

Project Overview

Project Introducing

Defn

- introduce

  With the development of the film industry, a variety of film and television companies need to predict the cost and income of shooting a film and television to reduce the amount of money spent. This software is designed to predict the movie revenue, etc.

Method Adopted

BP Neural Network Prediction Model

Back propagation network (BP network) is also known as back-propagation neural network. Through the training of sample data, the weights and thresholds of the network are constantly modified, so that the error function decreases along the negative gradient direction and approaches the expected output. It is a widely used neural network model, which is mostly used in function approximation, model recognition and classification, data compression and time series prediction. Click to open the link (example of BP neural network prediction)

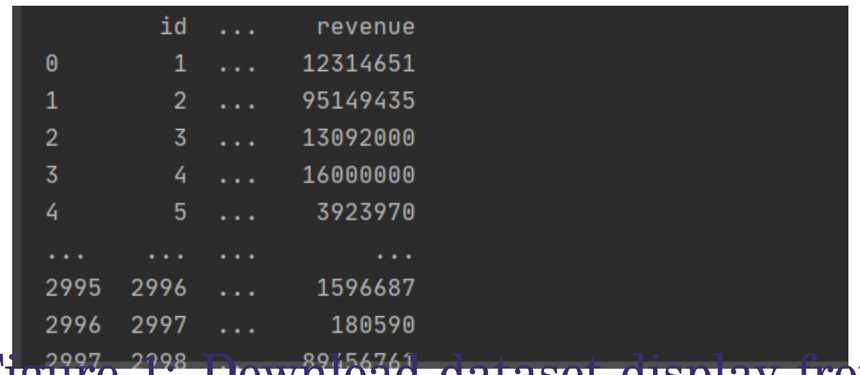
Data

Data Related Operations

- Data Collection

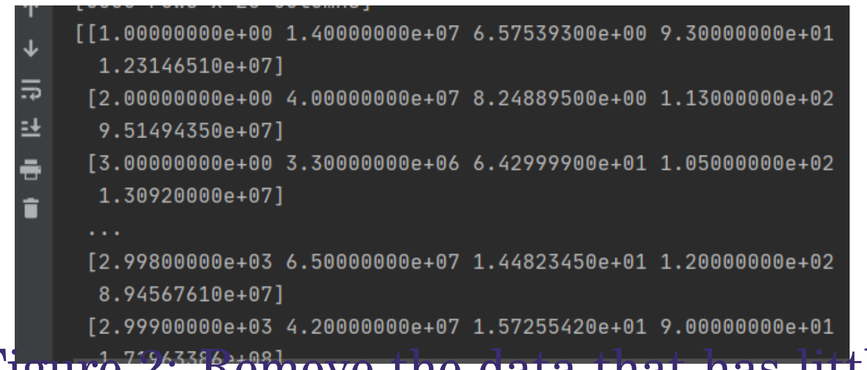
  The data set directly obtained in kaggle
- Data Processing

  The processing of useless data



	id	...	revenue
0	1	...	12314651
1	2	...	95149435
2	3	...	13092000
3	4	...	16000000
4	5	...	3923970
...	...	...	...
2995	2996	...	1596687
2996	2997	...	180590
2997	2998	...	8956761

Figure 1: Download dataset display from kaggle



```
[[1.00000000e+00 1.40000000e+07 6.57539300e+00 9.30000000e+01
  1.23146510e+07]
 [2.00000000e+00 4.00000000e+07 8.24889500e+00 1.13000000e+02
  9.51494350e+07]
 [3.00000000e+00 3.30000000e+06 6.42999900e+01 1.05000000e+02
  1.30920000e+07]
...
[2.99800000e+03 6.50000000e+07 1.44823450e+01 1.20000000e+02
  8.94567610e+07]
[2.99900000e+03 4.20000000e+07 1.57255420e+01 9.00000000e+01
```

Figure 2: Remove the data that has little influence on the weight

Data Related Operations

- Data Collection

  The data set directly obtained in kaggle
- Data Processing

  The processing of useless data

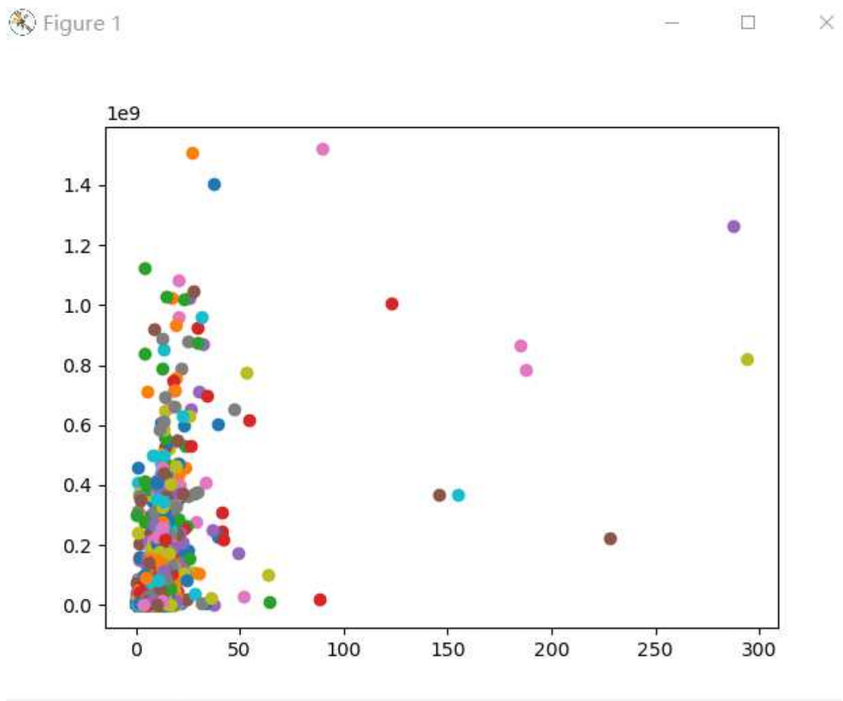


Figure 3: Download dataset display from kaggle

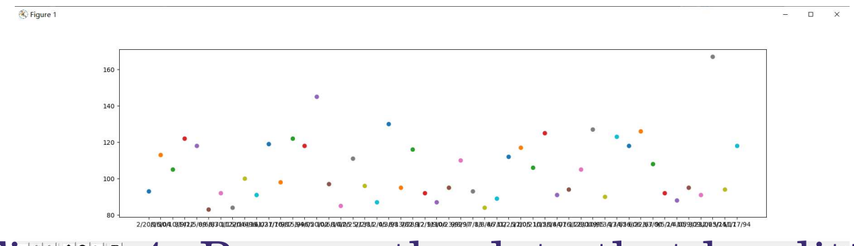


Figure 4: Remove the data that has little influence on the weight

Data Related Operations

- Data Collection

  The data set directly obtained in kaggle
- Data Processing

  The processing of useless data

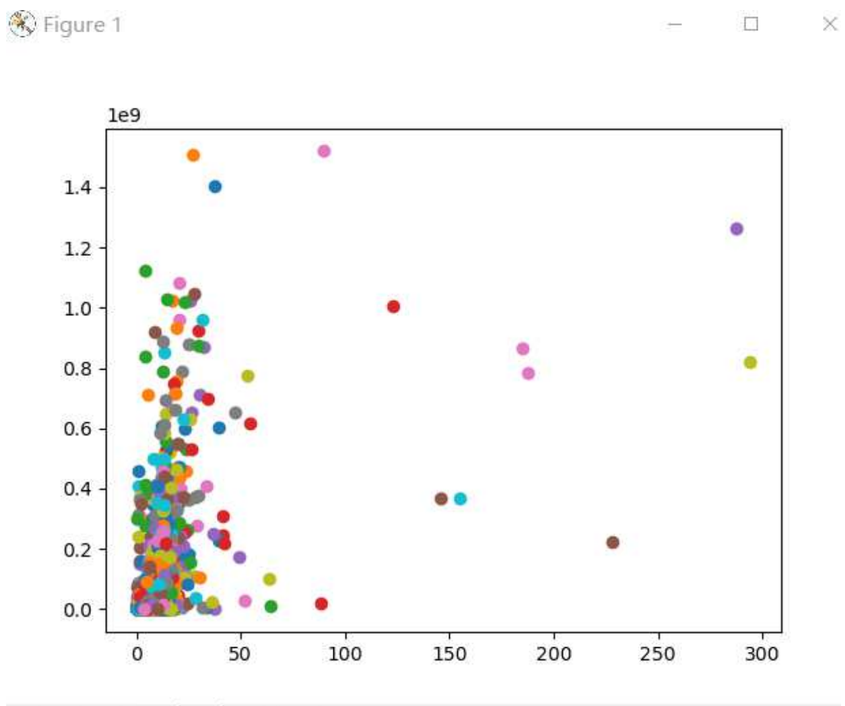


Figure 5: Download dataset display from kaggle

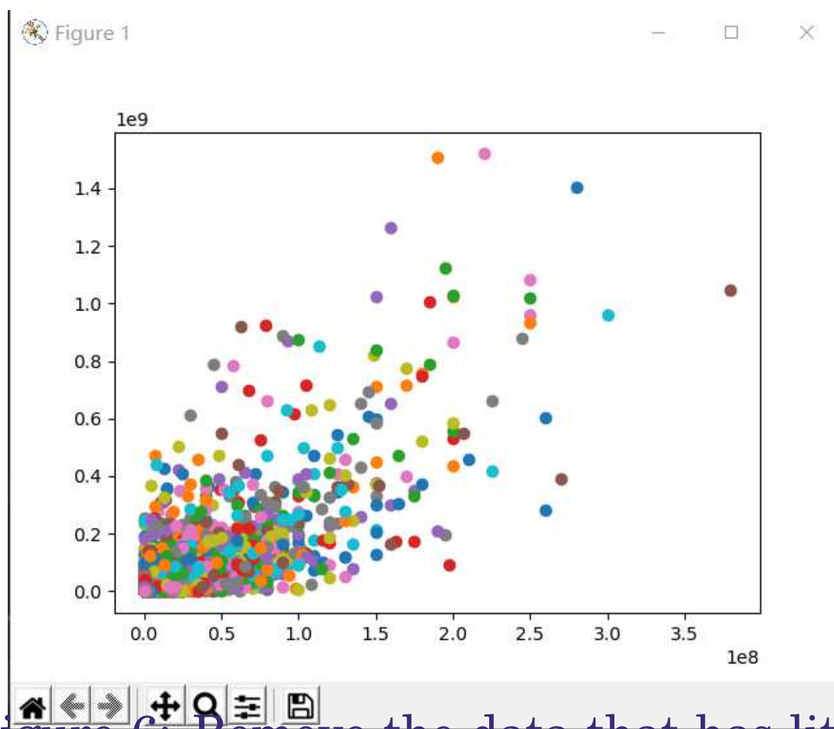


Figure 6: Remove the data that has little influence on the weight

Model Training And Testing

Training And Testing

- Cost function selection and neural network structure selection
- Display of test data

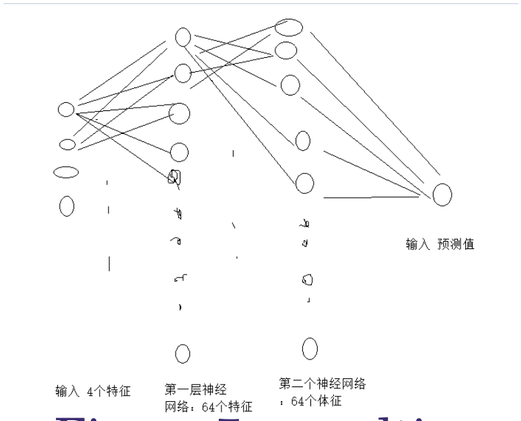


Figure 7: precdtion

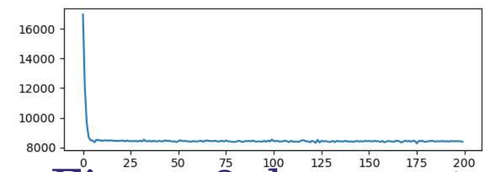


Figure 8: loss upate

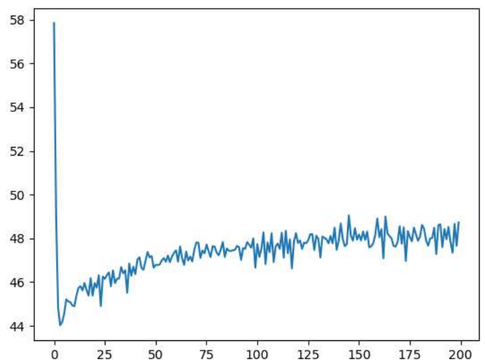


Figure 9: mean_absolute_error

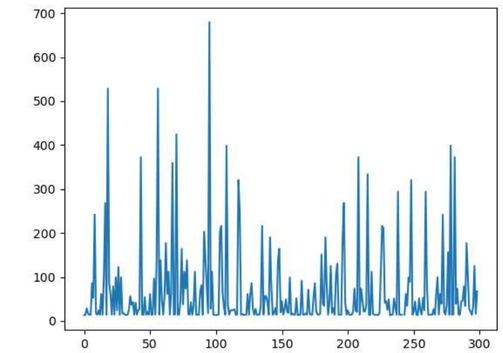


Figure 10: precdtion

conclusion

conclusion

Summary: problem: in the actual test process: in the process of numerical calculation, the large value results in the program running error. // Harvest: a deeper understanding of logical regression