Box Office Forecast

Zhangtao Xue

Xi'an Shiyou University

Chinese Academy of Sciences

October 8, 2020

Overview

Project Overview

Project Introducing

Method Adopted

BP Neural Network Prediction Model

Data

Model Training And Testing

Training And Testing

conclusion

conclusion

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                              belongs_to_collection    budget  \
0   1  [{'id': 313576, 'name': 'Hot Tub Time Machine ...  14000000
1   2  [{'id': 107674, 'name': 'The Princess Diaries ...  40000000
2   3                                                NaN   3300000
3   4                                                NaN   1200000
4   5                                                NaN         0

                                              genres  \
0                     [{'id': 35, 'name': 'Comedy'}]
1  [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...
2                      [{'id': 18, 'name': 'Drama'}]
3  [{'id': 53, 'name': 'Thriller'}, {'id': 18, 'n...
4  [{'id': 28, 'name': 'Action'}, {'id': 53, 'nam...

                            homepage    imdb_id original_language  \
0                                NaN  tt2637294                en
1                                NaN  tt0368933                en
2  http://sonyclassics.com/whiplash/  tt2582802                en
3         http://kahaanithefilm.com/  tt1821480                hi
4                                NaN  tt1380152                ko

                            original_title  \
0                     Hot Tub Time Machine 2
1  The Princess Diaries 2: Royal Engagement
2                                 Whiplash
3                                  Kahaani
4                                     마린보이

                                            overview  popularity    ...     \
0    who has become the "father of the In...    6.575393    ...
1  mopolis is now a college graduate and ...    8.248895    ...
2  Under the direction of a ruthless instructor, ...   64.299990    ...
3  Vidya Bagchi (Vidya Balan) arrives in Kolkata ...    3.174936    ...
4  Marine Boy is the story of a former national s...    1.148070    ...

  release_date runtime                                   spoken_languages  \
0      2/20/15    93.0           [{'iso_639_1': 'en', 'name': 'English'}]
1       8/6/04   113.0           [{'iso_639_1': 'en', 'name': 'English'}]
2     10/10/14   105.0           [{'iso_639_1': 'en', 'name': 'English'}]
3       3/9/12   122.0  [{'iso_639_1': 'en', 'name': 'English'}, {'iso...
4       2/5/09   118.0           [{'iso_639_1': 'ko', 'name': '한국어/조선말'}]
```

Figure 1: Download dataset display from kaggle

	id	budget	popularity	release_date	runtime	revenue
0	1	14000000	6.575393	2/20/15	93.0	12314651
1	2	40000000	8.248895	8/6/04	113.0	95149435
2	3	3300000	64.299990	10/10/14	105.0	13092000
3	4	1200000	3.174936	3/9/12	122.0	16000000
4	5	0	1.148070	2/5/09	118.0	3923970

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

Data Related Operations

- The relationship between time and business income
- Relationship between runtime and business income

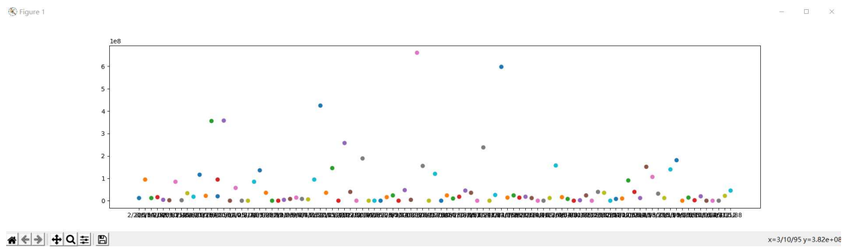

Figure 3: time

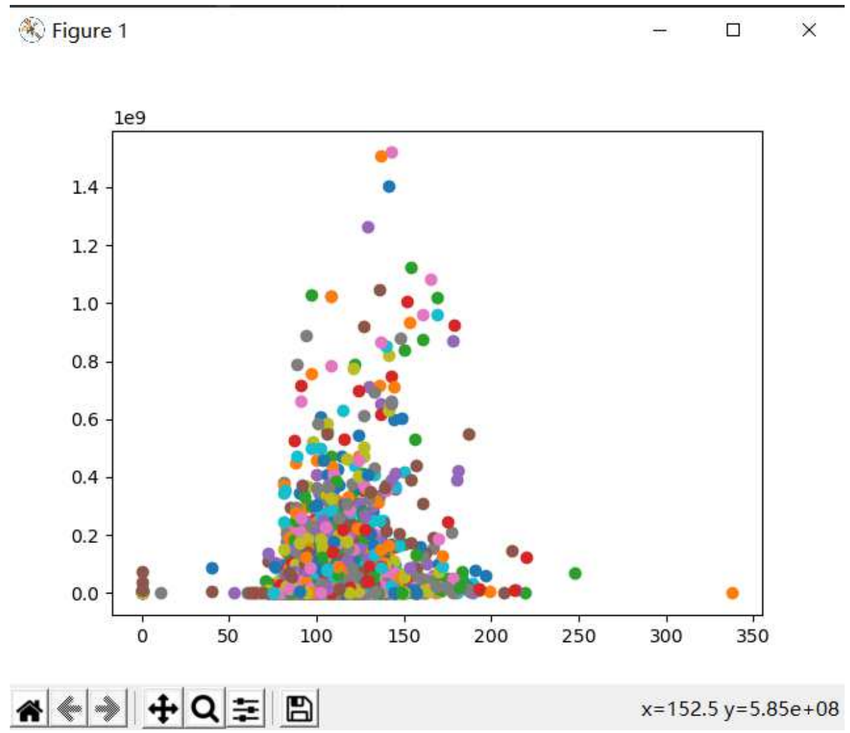

Figure 4: runtime

Data Related Operations

- The relationship between the prophase investment and business income of films
- The relationship between popularity and business income

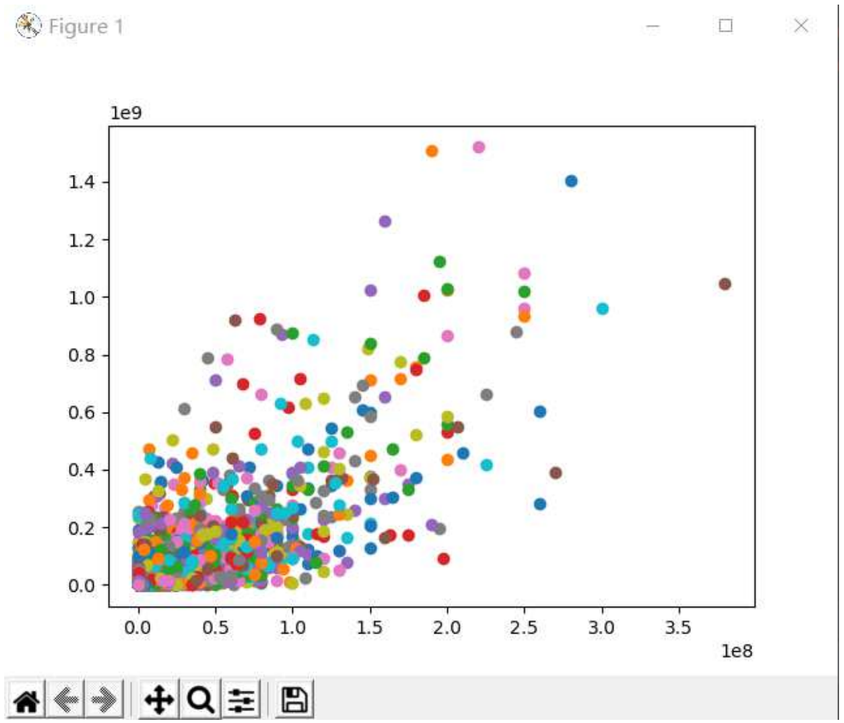

Figure 5: investment

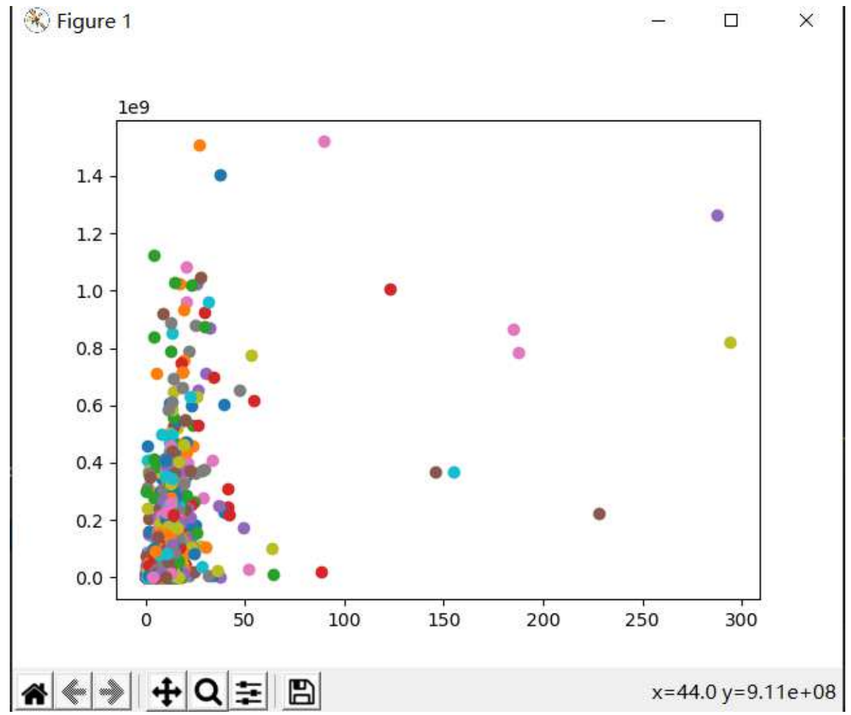

Figure 6: popularity

Model Training And Testing

Training And Testing

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

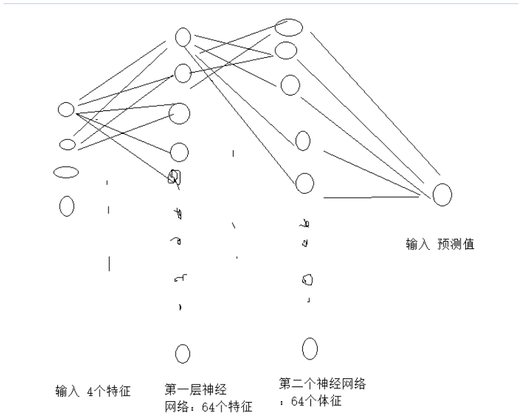

Figure 7: precdtion

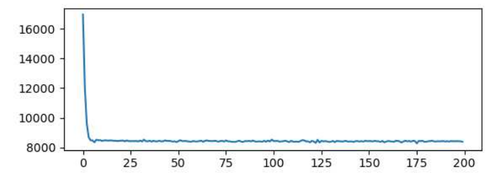

Figure 8: loss upate

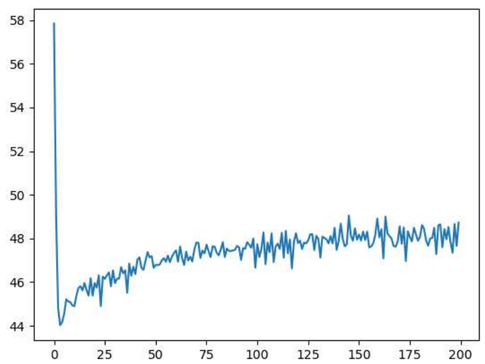

Figure 9: mean_absolute_error

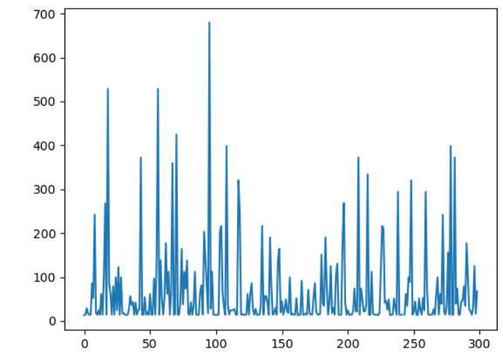

Figure 10: precdtion

conclusion

conclusion

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