Real estate warning based on BP neural network

Jiajun Hu[1], Yixiang Wang[2], Shengquan Wang[3], Ang Li[3, 4]

[1]School of economics, Central University of Finance and Economics,Beijing China.

[2]Guangxi University, Nanning, Guangxi China

[3]Institute of electronic engineering and Optoelectronic Technology, Nanjing University of Science and Technology ZiJin College, Nanjing City, China

[4]College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing City, China)

Keywords: BP neural network; intelligent algorithm; house price forecast; matlab

Abstract: Nowadays the real estate industry is the pillar industry in many areas. Therefore, experts from all walks of life have come up with some convenient and accurate methods for forecasting real estate prices. However, they are generally manual analysis methods, and there are many subjective factors in it, so they are very inaccurate. The method based on BP neural network introduced in this paper can avoid this, and has a very good effect.

1. introduction

This paper estimates the house price by macro factors, and constructs a comprehensive index system. Using neural network theory, we can estimate the average price of any real estate (including the upcoming real estate) in the third quarter of 2018. Take a real estate as an example. According to its previous real estate price and the application of neural network algorithm, it can predict the real estate price of this property in the third quarter of 2018. This method can be used to evaluate and forecast any real estate.

2. Bp neural network

2.1 General bp neural network

The structure of multi-layer feedforward network based on bp algorithm is shown in Figure 1.

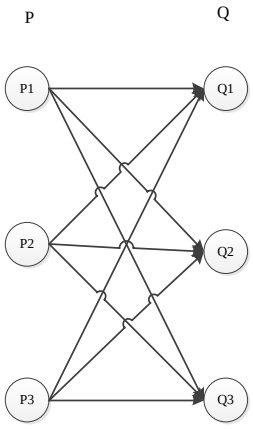

Figure 1. Bp neural network

When the kth sample is input, the input of node j is

$$net_{ij}^{l} = \sum_{j} w_{ij}^{l} o_{jk}^{l-1} \tag{1}$$

$$o_{jk}^{l} = f(net_{jk}^{l}) \tag{2}$$

The error function used is

$$E_k = \frac{1}{2}\sum_{l}(y_{lk} - \bar{y}_{lk})^2 \tag{3}$$

among them $\bar{y}_{lk}$ Is the actual output of unit j. The total error is

$$E = \frac{1}{2N}\sum_{k=1}^{N} E_k \tag{4}$$

definition $\delta_{jk}^{l} = \frac{\partial E_k}{\partial net_{jk}^{l}}$

$$\text{then } \frac{\partial E_k}{\partial w_{ij}^{l}} = \frac{\partial E_k}{\partial net_{jk}^{l}}\frac{\partial net_{jk}^{l}}{\partial w_{ij}^{l}} = \frac{\partial E_k}{\partial net_{jk}^{l}} o_{jk}^{l-1} = \delta_{jk}^{l} o_{jk}^{l-1} \tag{5}$$

The following two situations are discussed:
(1) If node j is an output unit, then $o_{jk}^{l} = \bar{y}_{jk}$

$$\delta_{jk}^{l} = \frac{\partial E_k}{\partial net_{jk}^{l}} = \frac{\partial E_k}{\partial \bar{y}_{jk}}\frac{\partial \bar{y}_{jk}}{\partial net_{jk}^{l}} = -(y_k - \bar{y}_k) f'(net_{jk}^{l}) \tag{6}$$

(2) If node j is not an output unit, then

$$\delta_{jk}^{l} = \frac{\partial E_k}{\partial net_{jk}^{l}} = \frac{\partial E_k}{\partial \bar{y}_{jk}}\frac{\partial o_{jk}^{l}}{\partial net_{jk}^{l}} = \frac{\partial E_k}{\partial o_{jk}^{l}} f'(net_{jk}^{l}) \tag{7}$$

At the mth unit of the (l+1) layer

$$\frac{\partial E_k}{\partial o_{jk}^{l}} = \sum_{m}\frac{\partial E_k}{\partial net_{mk}^{l=1}}\frac{\partial net_{mk}^{l+1}}{\partial o_{jk}^{l}} = \sum_{m}\frac{\partial E_k}{\partial net_{mk}^{l+1}} w_{mj}^{l+1} = \sum_{m}\delta_{mk}^{l+1} w_{mj}^{l+1} \tag{8}$$

Substituting equation (8) into equation (7),

$$\delta_{jk}^{l} = \sum_{m}\partial_{mk}^{l+1} w_{mj}^{l+1} f'(net_{jk}^{l}) \tag{9}$$

Summarize the above results, there are

$$\begin{cases} \delta_{jk}^{l} = \sum_{m}\delta_{mk}^{l+1} w_{mj}^{l+1} f'(net_{jk}^{l}) \\ \frac{\partial E_k}{\partial w_{ij}^{l}} = \delta_{jk}^{l} o_{jk}^{l-1} \end{cases} \tag{10}$$

Until the error indicator meets the accuracy requirement [11], ie:
$E = \frac{1}{2N}\sum_{k=1}^{N} E_k < \varepsilon, \varepsilon$ Calculate the unit of each floor o_{jk}^{l-1}, net_{jk}^{l} with $\bar{y}_k$,k=2, ..., N.
Correction weight

$$w_{ij} = w_{ij} - \mu\frac{\partial E}{\partial w_{ij}}\ \mu > 0 \tag{11}$$

3. Simulation

Buyers are asked to describe their dream houses, they may not start from the height of the basement ceiling or near the East-West railway. But the data set for this playground game proves that the price negotiation is more than the number of bedrooms or white fences.

There are 79 explanatory variables describing (almost) all aspects of the home in Ames, Iowa, which are used to predict the final price of each household

Table 1. Housing price data (from the kaggle platform)

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope
1461	20	RH	80	11622	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl
1462	20	RL	81	14267	Pave	NA	IR1	Lvl	AllPub	Corner	Gtl
1463	60	RL	74	13830	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl
1464	60	RL	78	9978	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl
1465	120	RL	43	5005	Pave	NA	IR1	HLS	AllPub	Inside	Gtl
1466	60	RL	75	10000	Pave	NA	IR1	Lvl	AllPub	Corner	Gtl
1467	20	RL	NA	7980	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl
1468	60	RL	63	8402	Pave	NA	IR1	Lvl	AllPub	Inside	Gtl
1469	20	RL	85	10176	Pave	NA	Reg	Lvl	AllPub	Inside	Gtl
1470	20	RL	70	8400	Pave	NA	Reg	Lvl	AllPub	Corner	Gtl

3.1 Model Solution

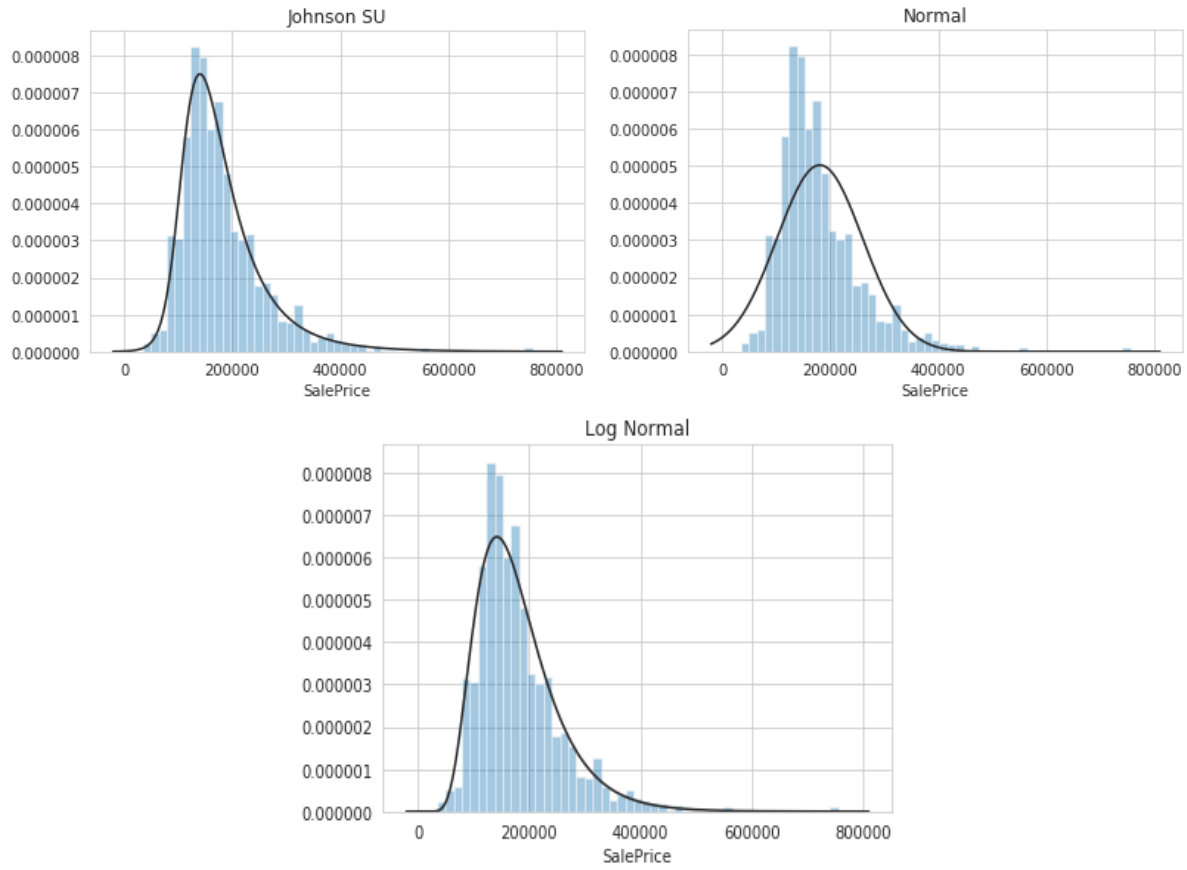

Figure 2. SalePrice

Obviously, SalePrice does not follow a normal distribution, so it must be converted before performing the regression. Although the log conversion does a good job, the most appropriate is the Johnson infinite distribution.

In this case, Spearman correlation is better, because even if they are non-linear, it will acquire the relationship between variables. OverallQual is the main criterion for establishing house prices. The neighborhood has great influence, and part of it has its own Some intrinsic value, but houses in some areas tend to have the same characteristics (confusion) leading to similar valuation factors.

	Id	SalePrice
0	1461	106479.0
1	1462	141435.0
2	1463	160652.0
3	1464	172714.0
4	1465	166140.0

Figure 3. results

After the calculation of the kaggle platform, the result is only 0.10649 deviation.

4. Conclusion

RF-BP neural network can reduce people's subjectivity, and it is effective for real estate development enterprises to avoid risks. Real estate development companies can implement risk

management by means of risk avoidance and insurance. The opacity of real estate market is in its agent-based simulation. There are some challenges. Although some real estate websites openly offer a large number of homes, the rest are not available. These price estimates are necessary to simulate their evolution from a complete initial home. In addition, this estimate can also be used for other purposes. , for example, to evaluate the home, let the buyer know which is the best price (ie the lowest price compared to the assessment) and advise the buyer to set the initial price. This work suggests combining the dimension reduction method with machine learning techniques to Obtaining the estimated price. In particular, this work analyzes the use of non-negative decomposition, recursive feature elimination, and feature selection with variance thresholds as a dimensionality reduction method. It will linear regression, support vector regression, k-nearest neighbors, and multi-layer sensing. The application of neural networks was compared as machine learning techniques. This work used ten times cross-validation for comparisonEstimate and error, and evaluate the improvement of the basic estimator commonly used at the beginning of the simulation. For reproducibility and support for other researchers, the software developed and the dataset used can be obtained free of charge from the data research library. Current work The problem of estimating the price of an unknown house using BP neural network has been solved.

Acknowledge

Jiangsu University Natural Science Foundation of Jiangsu Province, project number: 18KJD510004. Scientific Research Innovation Program for academic degree Graduate students in Colleges and Universities in Jiangsu Province, Project No: KYLX160661.

References

[1] Sun Xibo. Research and application of bp neural network algorithm and other algorithms [d]. Chongqing University, 2011.

[2] Chen Jianting. Multivariable time series prediction method based on improved bp neural network [j/ol]. Electronic Technology and Software Engineering, 2019 (05): 163-165 [2019-03-21].

[3] Yan Feiyu, Li Weizhao, Yang Weiwei, He Yaling. Fuel cell performance prediction method based on Bagging neural network integration [J/OL]. Chinese Science: Technical Science: 1-11 [2019-03-21].

[4] Wang Yang, Wang Feifan, Zhang Shuyi, Huang Shaofen, Xu Shuang, Zhao Chenxi, Zhao Chuanxin. Social media spam filtering based on tf-idf and improved bp neural network [j]. Computer System Application, 2019 (03): 126-132.

[5] Wang Gusen, Gao Bo. Fire Point Location Model Based on Adaptive Learning Rate bp Neural Network [j]. Computer System Application, 2019 (03): 250-254.

[6] Feng Yufang,Lu Houqing,Yin Hong,Cao Lin.Research on fault diagnosis model based on bp neural network [J].Computer Engineering and Applications, 2019 (06): 24-30.

[7] However, Song Lin, Liu Shangqi, Luo Yanyan, Liang Guangyue, Yang Zhaopeng. Prediction of the effect of high aquifer on sagd development based on bp neural network [j/ol]. Daqing Petroleum Geology and Development: 1-8 [2019-03-21].

[8] Zhang Wei,Qian Yuliang,Qiu Zheng,Zhang Yun.Gas turbine fault diagnosis based on dragonfly algorithm for optimizing bp neural network [j/ol]. Thermal Energy Engineering, 2019(03):26-32 [2019-03-21].

[9] Xu Liang, Zhao Songbo, Gao Qiang, Mo Jiaqing, Lu Xiaoyi, Guo Xiang.Application of pso-bp neural network in the evaluation of speech interference effects [j/ol]. Modern electronic technology, 2019(06): 43-46+50 [2019-03-21].