Consumer Loan Credit Risk Analyzer using Artificial Neural Networks

Shilpa Laddha (Kabra)
Assistant Professor (IT),
Government Engg College,
Station Rd, Aurangabad (MS)

Shubhangi Sapkal,
Assistant Professor (MCA),
Government Engg College,
Station Rd, Aurangabad (MS)

Anjali Kulkarni
Assistant Professor,
CKT College
Pavai, New Mumbai

ABSTRACT

Neural networks provide an obvious technique for classification. In this paper, neural network approach is used for classification of Loan applications. Data has been collected from publicly available source, Alyuda Research, Inc.[1].It does not guarantee the accuracy of the data. This data is intended solely for experimental purposes. FFNN and RBN are used for classification. PREMNMX matlab Function Preprocesses data so that minimum is -1 and maximum is 1

General Terms

Neural Network, classification, Radial Basis Function, Backpropogation Learning.

Keywords

Feedforward Neural Network,Radial Basis Network,Loan approval.

1. INTRODUCTION

Consumer Loan credit risk analyzer is a method of evaluating approval/disapproval of loan applications. Using historical data and statistical techniques, it tries to isolate the effects of various applicant characteristics on delinquencies and defaults. The method produces a "Output" that a bank can use to approve/reject loan to the applicants. Information on borrowers is obtained from their loan applications and from credit bureaus.In most (but not all) decision systems, a higher score indicates lower risk, and a lender sets a cutoff score based on the amount of risk it is willing to accept. Strictly adhering to the model, the lender would approve Loan to the applicants with scores above the cutoff and deny applicants with scores below (although many lenders may take a closer look at applications near the cutoff before making the final credit decision). To build a good scoring model, developers need sufficient historical data, which reflect loan performance in periods of both good and bad economic conditions [2].

2. NEURAL NETWORKS (NN) FOR CLASSIFICATION

For classification, following methods are used,

• Feed forward Neural Network

• Radial Basis Neural Network

- Input Data Set → Pre-Processing →

Classifier → Classification

Fig 1: Steps for classification

3. SYSTEM DEVELOPMENT

3.1 Preprocessing:

PREMNMX matlab Function Preprocesses Input data set so that minimum is -1 and maximum is 1.

3.2 Data Transformation:

The datasets for training & testing are converted into numerical values by assigning the following values.

1) Sex: Male: 1

Female: 0.9

2) Age is taken as it is.

3) Address time is taken as it is.

4) Marital status:

Married and age=>20 &<=45 then 1 else 0.5

Single and age>20 &<=45 then 1 else 0.5

5) Occupation:

Blue Collar: 1 ,Professional: 1

Semi-Professional: 0.5, Retired: 0.5

Unemployed: 0, Office: 1

Manager: 1, Principal: 1

6) Job Time is taken as it is.

7) Checking: Yes : 1, No : 0.5

8) Savings: Yes: 1 No : 0.5

9) Payment History is taken as it.

10) Home Ownership: Own: 1, Rent: 0.5

11) Fin Ratio 1 is taken as it is.

12) Fin Ratio 2 is taken as it is.

10 Sample training records & 23 testing records after data transformation are shown in table 2.

Table 1: Reconstructed dataset used for training and testing before data transformation [1]

Historical data for neural network training

Inputs												Target
Sex	Age	Address Time	Marital Status	Occupation	Job Time	Checking	Savings	Payment History	Home Ownership	FinRatio 1	FinRatio 2	Credit Risk
male	30.83	0	Married	Blue Collar	1.25	Yes	Yes	1	Rent	202	0	Low
female	58.67	4.46	Married	Professional	3.04	Yes	Yes	6	Rent	43	560	Low
female	24.5	0.5	Married	Professional	1.5	Yes	No	0	Rent	280	824	Low
female	27.17	1.25	Married	Semi-profes	0	No	Yes	1	Rent	92	300	High
male	25.92	0.875	Married	Blue Collar	0.375	No	Yes	2	Own	174	3	High
male	23.08	0	Married	Blue Collar	1	No	Yes	11	Rent	0	0	High
male	39.58	5	Married	Semi-profes	0	No	Yes	2	Rent	17	1	High
male	30.58	2.71	Single	Blue Collar	0.125	No	No	0	Own	80	0	High
male	17.25	3	Married	Blue Collar	0.04	No	No	0	Own	160	40	High
female	17.67	0	Single	Semi-profes	0	No	No	0	Rent	86	0	High

Only Sample 10 records are shown above.The network is trained with 500 such records using backpropogation feedforward neural network & Radial basis network.
The trained networks are then used to classify the new data.Following 23 new data records are used for testing.

New data for forecasting

Inputs												Output		
Sex	Age	Address Time	Marital Status	Occupation	Job Time	Checking	Savings	Payment	Home Own	FinRatio 1	FinRatio 2	Feedforward	Radial Basis	Actual Targe
male	51.33	10	Married	Manager	0	Yes	Yes	11	Rent	0	1249	Low	Low	Low
male	34.75	15	Married	Principal	5.375	Yes	Yes	9	Own	0	134	Low	Low	Low
male	23.08	11.5	Married	Professional	2.125	Yes	Yes	11	Own	290	284	Low	Low	Low
male	30.17	0.5	Married	Blue Collar	1.75	Yes	Yes	11	Rent	32	540	Low	Low	Low
male	25.17	6	Married	Blue Collar	1	Yes	Yes	3	Rent	0	0	Low	Low	Low
male	39.17	1.625	Married	Blue Collar	1.5	Yes	Yes	10	Rent	186	4700	Low	High	Low
male	29.83	1.25	Single	Blue Collar	0.25	No	No	0	Rent	224	0	High	Low	High
male	20.08	0.25	Married	Blue Collar	0.125	No	No	0	Rent	200	0	High	High	High
female	29.58	1.75	Single	Blue Collar	1.25	No	No	0	Own	280	0	High	Low	High
male	32.33	3.5	Married	Blue Collar	0.5	No	No	0	Own	232	0	High	High	High
male	22	0.79	Married	Blue Collar	0.29	No	Yes	1	Rent	420	283	High	High	High
male	19.33	10.915	Married	Manager	0.585	No	Yes	2	Own	200	7	High	High	High
male	29.42	1.25	Married	Professional	0.25	No	Yes	2	Own	400	108	High	High	High
male	32.25	14	Single	Semi-profes	0	No	Yes	2	Rent	160	1	High	High	High
male	36.08	2.54	Married	Semi-profes	0	No	No	0	Rent	0	1000	High	High	High
male	29.25	13	Married	Professional	0.5	No	No	0	Rent	228	0	High	High	High
female	27.25	0.29	Married	Professional	0.125	No	Yes	1	Own	272	108	High	High	High
female	38.75	1.5	Married	Semi-profes	0	No	No	0	Rent	76	0	High	Low	High
male	32.42	2.165	Single	Semi-profes	0	No	No	0	Rent	120	0	High	High	High
male	34.17	2.75	Married	Manager	2.5	No	No	0	Own	232	200	High	High	High
male	31.58	0.75	Single	Blue Collar	3.5	No	No	0	Own	320	0	High	High	High
female	52.5	7	Married	Professional	3	No	No	0	Rent	0	0	High	High	High
male	36.17	0.42	Single	Blue Collar	0.29	No	No	0	Own	309	2	High	High	High

Table 2: Reconstructed dataset used for training and testing after data transformation [1]

Historical data for neural network training after data transformation.

Inputs											Output		Target
Sex	Age	Address Time	Marital Status	Occupation	Job Time	Checking	Savings	Payment History	Home Ownership	FinRatio 1	FinRatio 2	Feedforward	Radial Basis	Actual Target For Comparison
1	30.83	0	1	1	1.25	1	1	1	0.5	202	0	0	0	0
0.9	58.67	4.46	0.5	1	3.04	1	1	6	0.5	43	560	0	0	0
0.9	24.5	0.5	1	1	1.5	1	0.5	0	0.5	280	824	0	0	0
0.9	27.17	1.25	1	0.5	0	0.5	1	1	0.5	92	300	1	1	1
1	25.92	0.875	1	1	0.375	0.5	1	2	1	174	3	1	1	1
1	23.08	0	1	1	1	0.5	1	11	0.5	0	0	1	1	1
1	39.58	5	1	0.5	0	0.5	1	2	0.5	17	1	1	1	1
1	30.58	2.71	1	1	0.125	0.5	0.5	0	1	80	0	1	1	1
1	17.25	3	0.5	1	0.04	0.5	0.5	0	1	160	40	1	1	1
0.9	17.67	0	0.5	0.5	0	0.5	0.5	0	0.5	86	0	1	1	1

New data for forecasting after data transformation.

Inputs											Output		Target
Sex	Age	Address Time	Marital Status	Occupation	Job Time	Checking	Savings	Payment History	Home 1ership	FinRatio 1	FinRatio 2	Feedforward	Radial Basis	Actual Target For Comparison
1	51.33	10	0.5	1	0	1	1	11	0.5	0	1249	0	0	0
1	34.75	15	1	1	5.375	1	1	9	1	0	134	0	0	0
1	23.08	11.5	1	1	2.125	1	1	11	1	290	284	0	0	0
1	30.17	0.5	1	1	1.75	1	1	11	0.5	32	540	0	0	0
1	25.17	6	1	1	1	1	1	3	0.5	0	0	0	0	0
1	39.17	1.625	1	1	1.5	1	1	10	0.5	186	4700	0	1	0
1	29.83	1.25	1	1	0.25	0.5	0.5	0	0.5	224	0	1	0	1
1	20.08	0.25	1	1	0.125	0.5	0.5	0	0.5	200	0	1	1	1
0.9	29.58	1.75	1	1	1.25	0.5	0.5	0	1	280	0	1	0	1
1	32.33	3.5	1	1	0.5	0.5	0.5	0	1	232	0	1	1	1
1	22	0.79	1	1	0.29	0.5	1	1	0.5	420	283	1	1	1
1	19.33	10.915	0.5	1	0.585	0.5	1	2	1	200	7	1	1	1
1	29.42	1.25	1	1	0.25	0.5	1	2	1	400	108	1	1	1
1	32.25	14	1	0.5	0	0.5	1	2	0.5	160	1	1	1	1
1	36.08	2.54	1	0.5	0	0.5	0.5	0	0.5	0	1000	1	1	1
1	29.25	13	1	1	0.5	0.5	0.5	0	0.5	228	0	1	1	1
0.9	27.25	0.29	1	1	0.125	0.5	1	1	1	272	108	1	1	1
0.9	38.75	1.5	1	0.5	0	0.5	0.5	0	0.5	76	0	1	0	1
1	32.42	2.165	1	0.5	0	0.5	0.5	0	0.5	120	0	1	1	1
1	34.17	2.75	1	1	2.5	0.5	0.5	0	1	232	200	1	1	1
1	31.58	0.75	1	1	3.5	0.5	0.5	0	1	320	0	1	1	1
0.9	52.5	7	0.5	1	3	0.5	0.5	0	0.5	0	0	1	1	1
1	36.17	0.42	1	1	0.29	0.5	0.5	0	1	309	2	1	1	1

3.3 Results :

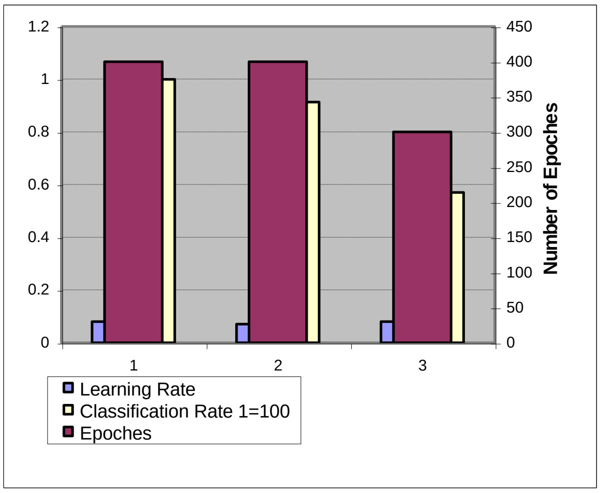
Classification rate is obtained as, Classification rate = Number of input records classified correctly * 100 / total number of input records used for classification.

3.4 Classifying With Feed Forward Neural Network

500 records are trained and other 23 records are used for testing. Different number of epochs and learning rate are used as shown in table 3 and corresponding classification rate is shown in graph 1. Feed forward Neural Network is used as a classifier.

Table 3: Classification rate by FFNN

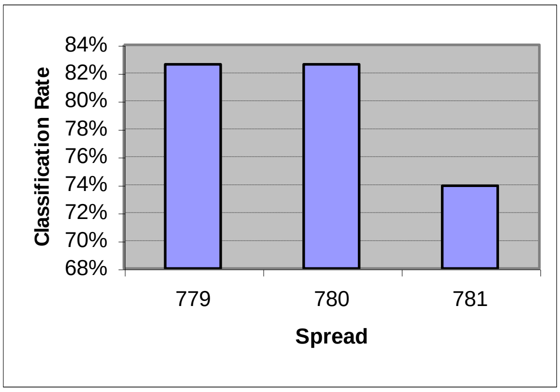
Classification Rate By Feedforward Neural Network			
Learning Rate	0.08	0.07	0.08
No. of epoches	400	400	300
Classification rate	100%	91%	57%



Graph 1: Classification rate by FFNN

3.5 Classifying with Radial Basis Neural Network

500 records are trained and other 23 records are used for testing. Different values of spread are used as shown in table 4 and corresponding classification rate is shown in graph 2. Radial Basis Neural Network is used as a classifier.



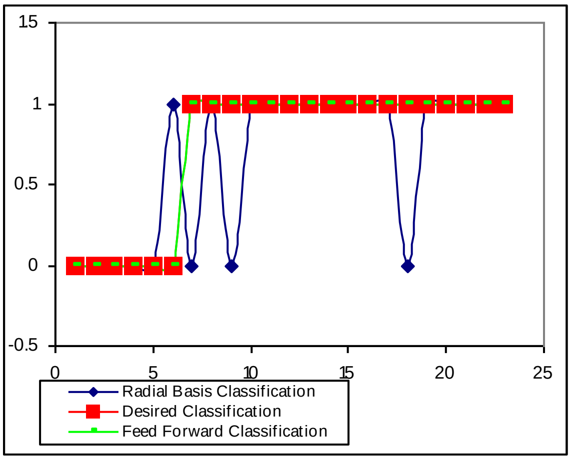
Graph 2: Classification rate by RBNN

Table 4: Classification rate with RBNN

Spread	779	780	781
Classification rate	83%	83%	74%

3.6 Comparison

Comparison between desired and actual output obtained by FFNN and RBNN is as shown in graph 3.



Graph 3: Comparison Of Classification rate by FFNN & RBNN.

All material on each page should fit within a rectangle of 18 x 23.5 cm (7" x 9.25"), centered on the page, beginning 2.54 cm (1") from the top of the page and ending with 2.54 cm (1") from the bottom. The right and left margins should be 1.9 cm (.75”). The text should be in two 8.45 cm (3.33") columns with a .83 cm (.33") gutter.

4. INTERPRETATION & COMPARATIVE ANALYSIS

First, it greatly reduces the time needed in the loan approval process. A study by the Business Banking Board found that the traditional loan approval process averages about 12-1/2 hours per small-business loan, and in the past, lenders have taken up to two weeks to process a loan (Allen). Credit scoring can reduce this time to well under an hour, although the time savings will vary depending on whether the bank adheres strictly to the credit score cutoff or whether it reevaluates applications with scores near the cutoff. This time savings means cost savings to the bank and benefits the customer as well. Customers need to provide only the information used in the scoring system, so applications can be shorter.

5. LIMITATIONS

The accuracy of the systems for underrepresented groups is still an open question. Accuracy is a very important consideration. Even if the lender can lower its costs of evaluating loan applications by using scoring, if the models are not accurate, these cost savings would be eaten away by poorly performing loans. The accuracy of a system will depend on the care with which it is developed. The data on which the system is based need to be a rich sample of both well-performing and poorly performing loans. The data should be up to date, and the models should be estimated frequently to ensure that changes in the relationships between potential factors and loan performance are captured. If the bank using scoring increases its applicant pool by mass marketing, it must ensure that the new pool of applicants behaves similarly to the pool on which the model was built; otherwise, the model may not accurately predict the behavior of these new applicants. The use of credit scoring itself may change a bank’s applicant pool in unpredictable ways, since it changes the cost of lending to certain types of borrowers. Again, this change in applicant pool may hurt the accuracy of a model that was built using information from the past pool of applicants. Account should be taken not only of the characteristics of borrowers whom are granted credit but also of those who were denied. Otherwise, a "selection bias" in the loan approval process could lead to bias in the estimated weights in the scoring model. A model's accuracy should be tested. A good model needs to make accurate predictions in good economic times and bad, so the data on which the model is based should cover both expansions and recessions. And the testing should be done using loan samples that were not used to develop the model in the first place. It is probably too soon to determine the accuracy [2].

6. IMPLICATIONS FOR THE BANKING INDUSTRY

Credit scoring is changing the way banks make small-business loans, and large banks are entering the market using credit scoring and processing applications using automated and centralized systems. These banks are able to generate large volumes of small business loans even in areas where they do not have extensive branch networks. Applications are being accepted over the phone, and some banks are soliciting customers via direct mail, as credit card lenders do. For example, Wells Fargo uses centralized processing for loans under $100,000, soliciting these loans nationwide, and uses credit scores not only in the approval process but also for loan pricing. For loans over $100,000, it still uses traditional underwriting, soliciting in areas where it has branches.

7. CONCLUSIONS

FFNN is a multilayer Neural Network, which uses back propagation for learning. From the obtained results, it can be said that, Four layers are sufficient for classification. As number of layers increased more than four, it does not improve the classification rate, but instead of that, if number of epochs are increased upto some limit and varying the learning rate, it improves the recognition rate . For the above dataset we get 100 percent output for 400 number of epochs and Learning rate is 0.08.
Experiment is done on the same dataset of 500 records by two methods: FFNN and RBNN. FFNN gives 100 percent classification rate and RBNN gives 82 percent classification rate. On the other side RBNN gives the output very fast (Few Seconds) as compare to FFNN (4 minutes).

8. REFERENCES

[1] Historical Dataset, Alyuda Research, Inc.

[2] Loretta J. Mester, "What's the Point of Credit Scoring?, Business Review, (Oct 1997), pp. 1-14.

[3] Clarence N. W. Tan and Gerhard E. Wittig, "A Study of the Parameters of a Backpropagation Stock Price Prediction Model" 1993.,pp.288 – 291.

[4] Ying-Hua Lu; Chun-Guo Wu; Yan-Chun Liang, "Center selection for RBF NN in prediction of nonlinear time series" 2003, pp. 1355 - 1359 .

[5] Website:http://www.virtualventures.ca/~neil/neural/neuron-a.html

[6] Yegya Narayana, "Artificial Neural Network", PHI.

[7] Gose, "Pattern Recognition and Image Analysis"

[8] Duda Hart, "Pattern Classification", 2 e,Wiley

[9] Fabrizio De Nittis "Consumer Loan Classification Using Artificial Neural Networks" ICSC EIS'98 conference.

[10] A Guide to MATLAB: Brian R. Hunt.

[11] Fundamentals of Artificial Neural Networks Mohamad Hassoun.

[12] Back Propagation: Theory, Architectures, and Applications, David E. Rumelhart.