CS360, Machine Learning Lab

PROJECT ON :

NEURAL APPROACH FOR CREDIT CARD FRAUD DETECTION

Professor : Dr. Moumita Roy

-By Soumyajit Karmakar
1901200
5th Sem, IIIT G
Monsoon Semester 2021

PROBLEM DEFINITION

Problem statement : Neural approach for credit card Fraud Detection

Description : For this project I have to tried to create a model which is capable of estimating the probability of default in credit card payment. The given dataset contains various features such as age, gender, marital status, etc. from this we need to predict if there will be a default in credit card payment for that person. The dataset contains labelled data from the year 2005 in the country of Taiwan.

In this project I will first briefly compare the performance of some classification models such as Logistic Regression and Single Layer Perceptron with the Multi Layer Perceptron to show why MLP is the most suitable model for this dataset.

The following slide contains description of the various features given in the dataset.

This dataset employed a binary variable, default payment (Yes = 1, No = 0), as the response variable. This project uses the following 23 variables as explanatory variables:

- X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
- X2: Gender (1 = male; 2 = female).
- X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
- X4: Marital status (1 = married; 2 = single; 3 = others).
- X5: Age (year).
- X6 - X11: History of past payment. The past monthly payment records (from April to September, 2005) are as follows: X6 = the repayment status in September, 2005; X7 = the repayment status in August, 2005; . . .;X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.
- X12-X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September, 2005; X13 = amount of bill statement in August, 2005; . . .; X17 = amount of bill statement in April, 2005.
- X18-X23: Amount of previous payment (NT dollar). X18 = amount paid in September, 2005; X19 = amount paid in August, 2005; . . .;X23 = amount paid in April, 2005.
- X24: default payment (Yes = 1, No = 0) ----------> Target

LITERATURE SURVEY

Presently, the use of a credit card has become an integral part of current banking and financial system. Predicting potential credit card defaulters or debtors is a crucial business opportunity for financial institutions. For this reason many kinds of datamining techniques using machine learning have been developed. 2 prominent research papers, after 2019, in this domain are listed ahead.

RESEARCH PAPER 1

-01/01/2019

Forecasting Crashes, Credit Card Default, and Imputation Analysis on Missing Values by the use of Neural Networks

-Jazmin Quezada

University of Texas at El Paso

Generalization of Backpropagation with application to Recurrent model by Werbos , P.J. shows how Artificial Neural Networks (ANNs) outperform the classical statistical methods such as linear regression and Box-Jenkins approaches. In this study the author has tried to use machine learning techniques to predict Credit Default payment and Stock Market evolution and crashes

Apart from these the author also tried to impute missing values in a data set with imputations methods and along with neural network methodologies.

In this work, the theoretical description of the neural network methodology and some practical applications which are based on real world data are presented. The author used the Multilayer perceptron to identify financial market crashes and also compute the credit card default payments of customers of a financial institution. The problem of detecting market crashes and credit card default payments were modeled as a special class of classification problem. The neural network technique is very efficient and robust compared to other classification techniques since it correctly discriminates with good accuracy

Link for the paper.

RESEARCH PAPER 2

-26/3/2021

Artificial neural network technique for improving prediction of credit card default: A stacked sparse autoencoder approach

-Sarah. A. Ebiaredoh-Mienye, E. Esenogho, Theo G. Swart

University of Johannesburg, Johannesburg, South Africa

In this paper the authors discussed how a new approach which can improve the effectiveness of the learning model. They used a stacked sparse autoencoder to achieve optimal feature learning. In the proposed autoencoder network, they introduced a batch normalization technique to enhance the performance, speed, and stability of the model and to prevent overfitting further.

In this paper, an approach is proposed to improve the classification performance of classifiers by using the unsupervised feature learning capability of autoencoders. During the training of the autoencoder, sparsity is encouraged, and the model is optimized using the AdaMax algorithm instead of the conventional stochastic gradient descent.

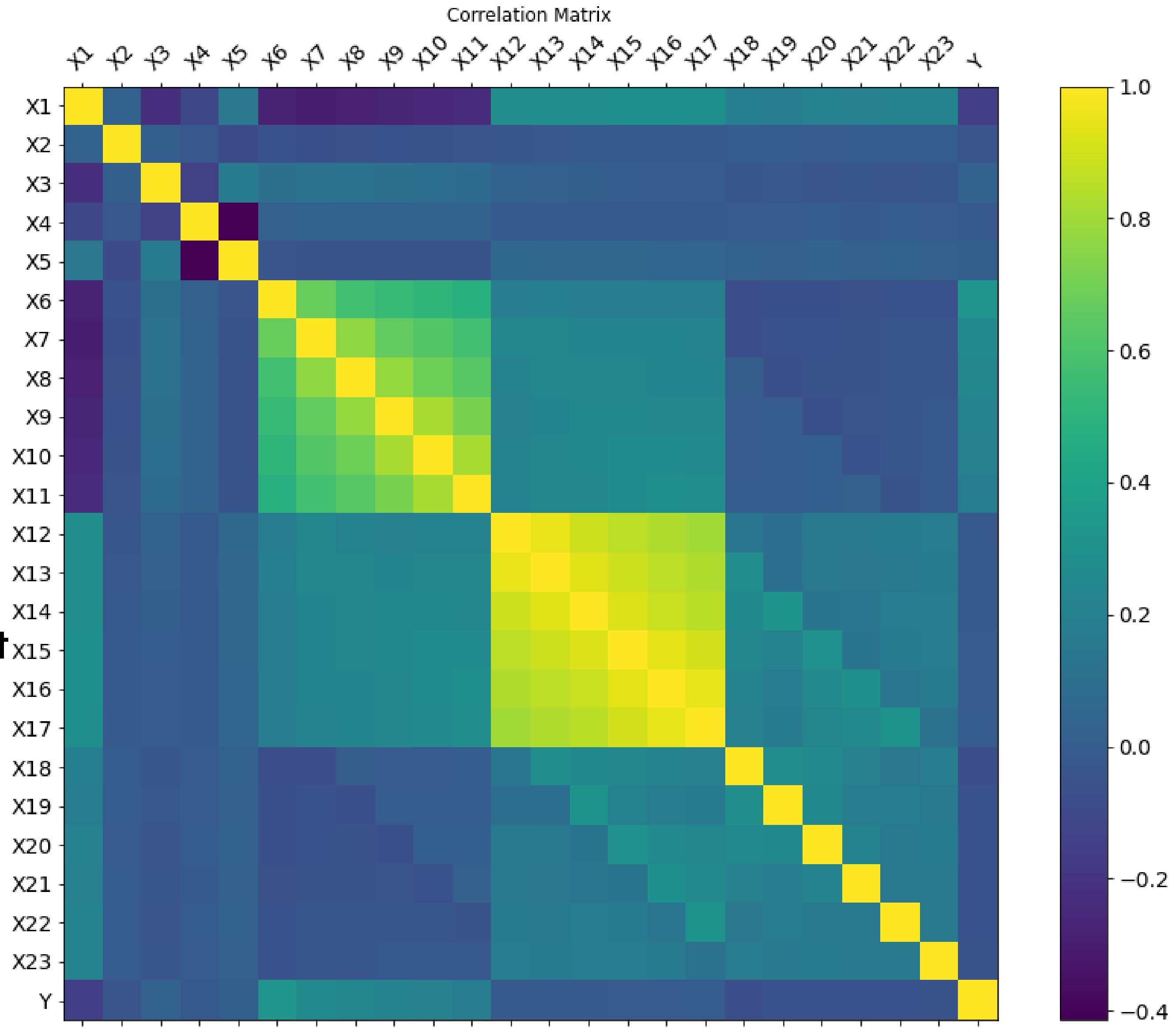
Also, to further prevent overfitting and enhance the performance, speed, and stability of the network, we introduced the batch normalization technique to the network. The low-dimensional features are then used to train various classifiers, including logistic regression (LR), classification and regression tree (CART), k-nearest neighbor (KNN), support vector machine (SVM), and linear discriminant analysis (LDA). The performance of the proposed method is compared with an instance where the classifiers were trained with the raw data.

Link for the paper.

RESULTS AND ANALYSIS OF RESULTS

First of all this heat map gives us the inter relation between the features. Larger positive value means stronger direct relation, while larger negative value means stronger inverse relation, zero means no relation.

Here we see that X6 to X11(history of past payments) seem to have a strong impact on the target variable. Also X3 (education) seems to have a strong impact as well. Among the various features X1(amount of the given credit) and X18 to X23 (amount of previous payment) seem to be less important.



Logistic Regression and Single Layer Perceptron

With logistic regression we are able to get around 75% accuracy. After hyperparameter tuning we could go up to 78-80% accuracy. The best results are as,

Class	Precision	Recall	F1-score	Patterns
0	0.81	0.97	0.89	7014
1	0.69	0.29	0.32	1986

With single layer perceptron we can achieve around 76% accuracy. After hyperparameter tuning we can reach up to 77% accuracy. The best results are as,

Class	Precision	Recall	F1-score	Patterns
0	0.79	1	0.88	7051
1	0.07	0.01	0.04	1949

Now we move to Multi Layer Perceptron model. First we will do the hyperparameter tuning and then the k-fold cross validation

Hyperparameter Tuning (Accuracy Table in %)

No. of hidden nodes

Learning Rate

	5	10	15	20	25	30
0.5	82.5,81.6,82.1	82.3,82.1,81.6	81.5,81.8,82.5	81.3,82,82.5	81.8,82,81.5	81.9,82.2,82.2
0.1	81.5,81.7,81.3	82.2,81.7,81.4	81.3,81.5,81.6	81.9,81.8,82.5	81.7,81.8,81.6	81.6,81.8,81.1
0.05	82.3,81.3,81.5	81,81.8,81.9	82,82.2,**82.7**	81.8,81.7,82.3	81.3,81.8,81.3	81.3,81.2,82.1
0.01	81.8,80.9,82	81.1,81.5,80.9	80.5,81.5,81.5	80.9,80.9,80.3	81.1,81.3,81.4	80.5,81,81.2
0.005	80.7,81.5,80.8	79.8,80.4,80.5	80.3,80.7,80.6	81.1,80.7,80	80.7,80.5,80.7	80,80.8,80.9
0.001	77.9,77.3,78.2	78.1,77.7,78	78.2,77.9,78.6	77.9,77.8,76.9	78.4,77.6,79.5	77.8,78.8,79.7

The tuple (x1,x2,x3) denotes the overall performance of the model when the training set to validation set percentage is 60:40, 70:30, 80:20 respectively.

We observe that we get the best results for the model with 15 hidden nodes, learning rate equal to 0.05 and train set to validation set ratio set to 80:20.

To ensure that the project is able to give accurate predictions, I have used a very low tolerance value (ρ) of 0.0001 throughout the training phase.

K-Fold Cross Validation

For testing purpose we use k-fold cross validation method, with the value of k = 5, to evaluate the performance of the model.

Using the optimal values of the hyperparameter we get the following average accuracy precision and recall (the results of all the individual folds are in the results sheet) :

Class	Average Precision	Average Recall	Average F1-score	Average Patterns
0	0.836	0.952	0.89	~4673
1	0.668	0.348	0.46	~1327

Macro Average	0.752	0.65	0.674
Weighted Average	0.8	0.82	0.798

Overall accuracy = 82%

ANALYSIS

From the results of the hyperparameter training we can see that the performance of the model depends heavily on the value of learning rate. Using very small learning rate was unfeasible as the improvement of the model was very slow and sometimes it went below the tolerance level and the model assumed it had converged prematurely.

Having larger learning rate is not advisable as sometimes they may converge and sometimes they may not, depends on the dataset.

In such cases we need to find the middle value, for this dataset I got 0.05 as the best learning rate.

To prevent the problem of overfitting I kept a track of the error as it was training and also tested the performance for various ratios of training set and validation set. I got the best results for 80:20 ratio.

Interestingly for this dataset the number of hidden nodes did not affect the performance too much. The best result was at 15.

FURTHER ANALYSIS

Any kind of dataset containing information about abnormal activities suffer a common problem of class imbalance, since the occurrence of such events is very rare as compared to the normal event.

This dataset which is about the default payment of credit cards comes under this category. The occurrence of default in credit card payment is very rare when we take into account the huge number of legitimate credit card payments, which is the normal event.

Out of the total 30,000 data points, this dataset contains 23,364 (77.88 %) normal payments and 6,636 (22.12 %) credit card defaults. Due to this the model will be biased to some extent. I have tried my best to reduce the bias by using the best possible hyperparameters.

NOVELTY

Due to the heavy class imbalance found in these kinds of datasets we cannot use the traditional machine learning models. We must modify them in such a way that even if one class is in minority the model is still able to predict it properly.

For this we may use a more stringent error function or change the performance metrics. Furthermore we may as well introduce error function such that it penalises strongly if the model makes a wrong prediction on minority class as compared to the majority class. We may employ tricks like resampling the data, for example, we can duplicate some of the data points of the minority to increase their number or we can create synthetic samples which would be similar to the existing minority data points to help train the model better.