Title: The Evolution of Credit Score Cards

From: Logistic Regression to Machine Learning

Introduction:

Credit score cards have become a prevalent risk control method in the financial industry. By utilizing personal information and data submitted by credit card applicants, these cards aim to predict the probability of future defaults and credit card borrowings, allowing banks to make informed decisions regarding credit card issuance. Credit scores provide an objective measure of risk magnitude, quantifying the creditworthiness of individuals. Historically, credit score cards have relied on logistic regression models. However, as economic fluctuations occur, these models may lose their predictive power. With the advent of machine learning algorithms, alternative predictive methods like Boosting, Random Forest, and Support Vector Machines have been introduced. While these methods offer improved accuracy, they may lack transparency, making it challenging to provide customers and regulators with clear reasons for acceptance or rejection.

Thesis Statement:

This essay will discuss the evolution of credit score cards, from their traditional reliance on logistic regression models to the introduction of machine learning algorithms. It will also explore the construction of a machine learning model to predict the creditworthiness of applicants and address the challenges posed by unbalanced data.

Body:

I. The foundation of credit score cards: Logistic regression

A. Historical reliance on logistic regression models

B. Quantifying risk through coefficient calculation

C. The multiplication and rounding of coefficients in score cards

II. Limitations of logistic regression models during economic fluctuations

A. Loss of predictive power during large economic fluctuations

B. The need for alternative predictive methods

III. Introduction of machine learning algorithms in credit card scoring

A. Boosting: Enhancing accuracy through ensemble learning

B. Random Forest: Leveraging decision trees for improved predictions

C. Support Vector Machines: Effective classification through hyperplanes

IV. The construction of a machine learning model to predict creditworthiness

A. Defining 'good' or 'bad' clients using vintage analysis

B. Addressing the challenge of unbalanced data

Sampling techniques to balance the dataset

Utilizing appropriate evaluation metrics for unbalanced data

V. Transparency challenges with machine learning approaches

A. Lack of clear reasons for acceptance or rejection

B. Difficulty in providing explanations to customers and regulators

Conclusion:

As credit score cards have evolved, they have transitioned from relying solely on logistic regression models to embracing machine learning algorithms. While these algorithms offer improved accuracy in predicting creditworthiness, they often lack transparency, making it challenging to provide clear explanations for acceptance or rejection. Additionally, the construction of machine learning models to predict creditworthiness requires addressing the challenges posed by unbalanced data. By understanding the advantages and challenges associated with logistic regression and machine learning approaches, stakeholders can make informed decisions to mitigate risks and serve customers effectively.

Task

Build a machine learning model to predict if an applicant is 'good' or 'bad' client, different from other tasks, the definition of 'good' or 'bad' is not given. You should use some techique, such as vintage analysis to construct you label. Also, unbalance data problem is a big problem in this task.

Attribute and Description

|  |  |  |
| --- | --- | --- |
| **application\_record** | | |
| Attribute | Description |  |
| Feature name | Explanation | Remarks |
| ID | Client number |  |
| CODE\_GENDER | Gender |  |
| FLAG\_OWN\_CAR | Is there a car |  |
| FLAG\_OWN\_REALTY | Is there a property |  |
| CNT\_CHILDREN | Number of children |  |
| AMT\_INCOME\_TOTAL | Annual income |  |
| NAME\_INCOME\_TYPE | Income category |  |
| NAME\_EDUCATION\_TYPE | Education level |  |
| NAME\_FAMILY\_STATUS | Marital status |  |
| NAME\_HOUSING\_TYPE | Way of living |  |
| DAYS\_BIRTH | Birthday | Count backwards from current day (0), -1 means yesterday |
| DAYS\_EMPLOYED | Start date of employment | Count backwards from current day(0). If positive, it means the person currently unemployed. |
| FLAG\_MOBIL | Is there a mobile phone |  |
| FLAG\_WORK\_PHONE | Is there a work phone |  |
| FLAG\_PHONE | Is there a phone |  |
| FLAG\_EMAIL | Is there an email |  |
| OCCUPATION\_TYPE | Occupation |  |
| CNT\_FAM\_MEMBERS | Family size |  |

|  |  |  |
| --- | --- | --- |
| **credit\_record** | | |
| Attribute | Description |  |
| Feature name | Explanation | Remarks |
| ID | Client number |  |
| MONTHS\_BALANCE | Record month | The month of the extracted data is the starting point, backwards, 0 is the current month, -1 is the previous month, and so on |
| STATUS | Status | 0: 1-29 days past due 1: 30-59 days past due 2: 60-89 days overdue 3: 90-119 days overdue 4: 120-149 days overdue 5: Overdue or bad debts, write-offs for more than 150 days C: paid off that month X: No loan for |