Assignment

by User user

Submission date: 06-Dec-2024 01:26AM (UTC+0400)
Submission ID: 2542153511
File name: userfile (316.38K)
Word count: 628
Character count: 3990

Neural Network Predictive Modeling of Account Balance

Introduction

The project aims to use neural network modeling as an effective tool for making predictions about account balance by considering all possible financial and demographic variables. It would do this by infusing advanced machine learning to find any complicated relationship between many features of a person's description and their account balance.

Dataset Description

Data Source and Selection Rationale

Care was taken in the selection to provide a rich and multidimensional overview of factors that could explain account balance. The selected dataset contains many variables to capture various features in an individual's financial and personal profile.

Variable Definitions

Target Variable

- **Balance (Scaled):** Represents the normalized account balance, serving as the primary predictive target of our analysis.

Explanatory Variables

- **Income (Scaled):** Normalized representation of one's yearly income

- **Credit Limit (Scaled):** Normalized credit limit assigned to the individual

- **Credit Rating (Scaled):** Standardized credit score/rating

- Number of Credit Cards (Scaled): Normalized count of credit cards

- Age (Scaled): Normalized age of the individual

- Education (Scaled): Normalized years of formal education

- Gender (Dummy Coded): Binary representation of gender

- Student Status (Dummy Coded): Indicator of students' enrollment

- Marital Status (Dummy Coded): Marital status indicator

- Ethnicity (Dummy Coded): Categorical representation of ethnic background

Preprocessing Methodology

Data Transformation Techniques

The preprocessing step was very significant in preparing data for analysis by a neural network.

Key transformation steps included:

- Variable Scaling: - All numeric variables were scaled in a 0-1 range. Scaling ensures equal weight and no domination by variables with larger magnitudes. It allows neural network algorithms to learn more effectively.

Categorical Variable Encoding: Categorical variables are converted to dummy variables, which enables numerical representation of categorical data and allows a neural network to process categorical information effectively.

Data Partitioning: The dataset is split into 60% training (model development) and 40% validation (model assessment), which ensures a sound evaluation of model performance and prevents overfitting by using separate datasets for training and validation.

Neural Network Modeling Approach

Model Configurations

Two distinct neural network architectures were developed and compared:

- **Model 1:** Single Hidden Layer - Structure: Input Layer → 3 Nodes Hidden Layer → Output Layer - Designed to capture primary linear and non-linear relationships

- **Model 2:** Two Hidden Layers - Structure: Input Layer → First Hidden Layer (2 Nodes) → Second Hidden Layer (2 Nodes) → Output Layer - Intended to capture more complex, hierarchical patterns in the data

Computational Methodology

Utilized R's neural net package for model implementation, set random seed (123) for reproducibility, and employed Mean Squared Error (MSE) as the primary performance metric.

Model Performance Analysis

Quantitative Results

Model Configuration	Mean Squared Error (MSE)
Single Hidden Layer (3 nodes)	0.0004358617
Two Hidden Layers (2 nodes each)	0.0004027084

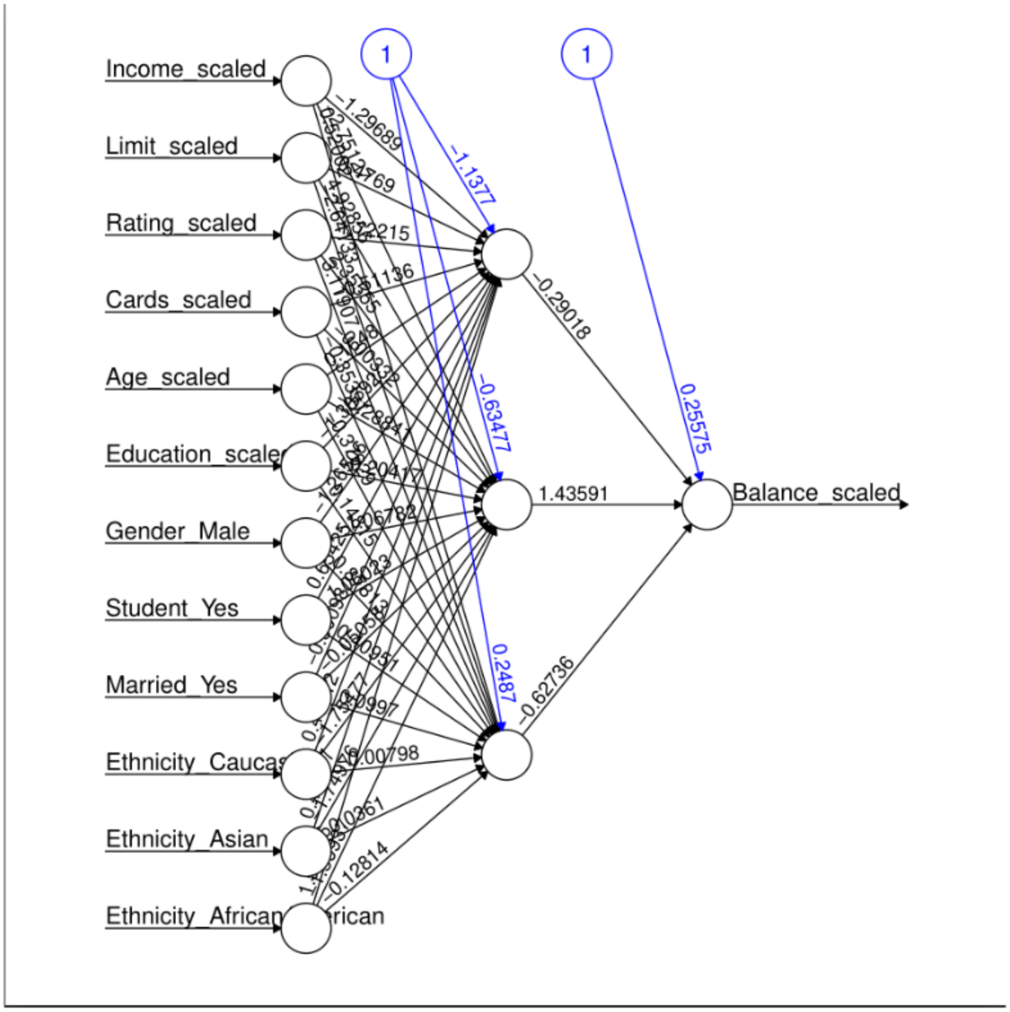
Performance Interpretation

The two-hidden-layer model demonstrated slightly superior performance, with a lower Mean Squared Error (MSE) compared to the single-hidden-layer model. The marginal difference suggests that both models effectively capture the underlying data patterns. The additional complexity of the two-layer model provides a minimal but noticeable improvement in predictive accuracy, enabling more nuanced feature interaction and representation.

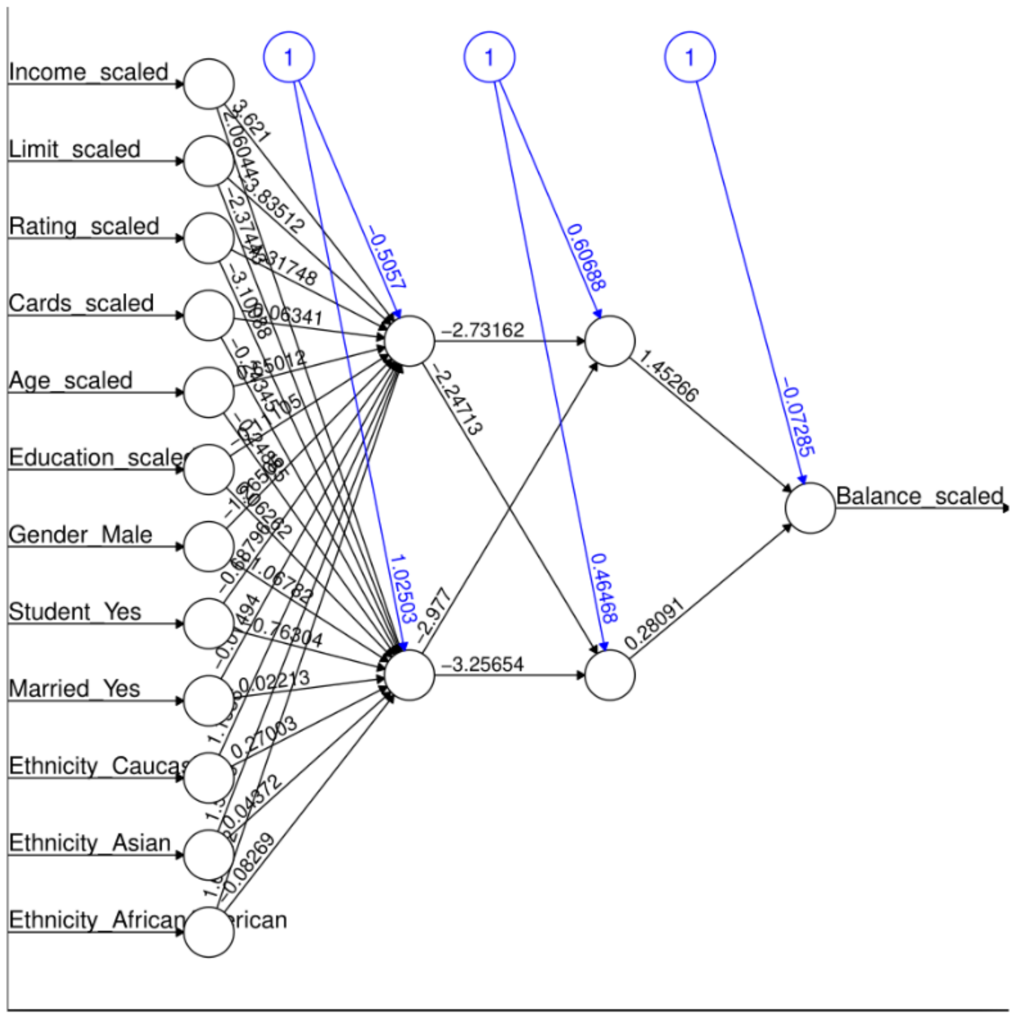
Visualization Insights

The neural network plots visually represent the network structures, illustrating the weights and connections between layers. These visualizations offer insights into how different features contribute to predicting account balance.

1. **Single Hidden Layer**



2. Two Hidden Layers



Limitations and Future Directions

Potential Improvements

- Explore alternative neural network architectures

- Conduct more extensive hyperparameter tuning

- Investigate additional feature engineering techniques

- Consider ensemble modeling approaches

Recommendations

- Validate findings with additional datasets

- Incorporate domain expertise to refine feature selection

- Continuously monitor and update the model with new data

Conclusion

The neural network models successfully demonstrated the potential of machine learning in predicting account balance. The two-layer model's marginally better performance highlights the value of exploring more complex network structures while maintaining interpretability.

Assignment

ORIGINALITY REPORT

2%	2%	0%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Bath Student Paper	2%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		