ECM2423 AI & Applications
Deep Learning for Predicting Credit Card Defaults

Gregory Korchagin
gk385@exeter.ac.uk

Term 2, 2024

---

1 Abstract

This report presents the development and evaluation of a deep learning model designed to predict credit card defaults. Leveraging a sequential neural network architecture with multiple dense layers, the model undergoes a detailed evaluation process including hyperparameter tuning and performance assessment using accuracy and F1 scores. Despite the challenges inherent in predictive modeling, such as imbalanced datasets and feature selection, the report and studies it is based on, introduce effective strategies for preprocessing and model optimisation. The main findings underscore the importance of careful feature preprocessing and hyperparameter exploration in enhancing model performance.

2 Introduction

Credit card default prediction poses a significant challenge due to the complexity of financial behaviors and the often imbalanced nature of datasets, where defaults may represent a small fraction of all observations. The primary aim is to develop a predictive model capable of accurately identifying potential credit card defaults, thereby aiding financial institutions in mitigating risks associated with credit issuance.

The task of predicting credit card defaults is challenging, given the complexity of financial behaviors and the often imbalanced nature of datasets. Recent advancements in deep learning, particularly with multilayer perceptrons (MLPs) and residual networks (ResNets), have shown promise in addressing these challenges[1]. These models leverage deep architectures and sophisticated training techniques, including the Adam optimiser which I have also used in my model. By incorporating this technique my model aims to enhance the predictive performance, offering a more sophisticated framework for identifying potential credit card defaults, by leveraging a sequential neural network architecture, benefiting from the adaptive learning rate properties of the Adam optimiser.

Recent trends in deep learning for credit card default prediction reflect significant advancements and diversification in methodologies aimed at enhancing the accuracy and reliability of these models. A systematic review by Hayashi (2022)[2] highlights the shift from traditional machine learning algorithms to deep learning (DL) techniques for credit scoring due to the superior accuracy of DL algorithms, particularly noting the effectiveness of deep belief networks (DBNs) over shallower networks.

In evaluating the performance of the developed deep learning model for predicting credit card defaults, the model was assessed using accuracy and the F1 score as primary metrics. These choices were made to address the challenges presented by the imbalanced nature of the dataset, where defaults constitute a smaller portion of the total observations. The F1 score, a harmonic mean of precision and recall, is particularly valuable in this context as it balances the importance of accurately predicting defaults (precision) with the need to capture as many true default instances as possible (recall). This balance is crucial for financial institutions that aim to minimise risk without unnecessarily limiting credit access to customers.

The evaluation also involved generating confusion matrices, which offer a visual representation of the model's predictions across the four possible outcomes (true positives, false positives, true negatives, and false negatives), providing insights into the types of errors made by the model. Although ROC curves and AUC were considered, I put more emphasis on F1 score and accuracy as I deemed most appropriate for the goals of this project and the specific challenges of the dataset.

This approach to evaluation acknowledges the complexity of the prediction task at hand and ensures a comprehensive assessment of the model's ability to perform in a real-world financial setting, where the costs of false positives and false negatives can be significant.

Following this introduction, Section 2 provides an exhaustive review of related work. Section 3 proposes methodology, including data preprocessing and model architecture. Section 4 presents the experimental results, detailing the hyperparameter settings, evaluation process, and obtained outcomes. Finally, Section 5 summarises the key findings, emphasising the implications and potential future directions stemming from this research.

3 Proposed Method

The foundation of my deep learning model's performance lies in the meticulous preprocessing of the dataset. Recognising the pivotal role of data quality and structure, the preprocessing phase is dedicated to optimising the dataset for neural network training. This phase involved two critical steps: normalisation and class imbalance mitigation.

Normalisation was performed using the MinMaxScaler, which adjusts numerical features to a common scale without distorting differences in the ranges of values. This standardisation is crucial for deep learning models, as it ensures that all inputs contribute equally to the model's learning process, preventing features with larger scales from dominating the model's attention, which I have found out thanks to an article by Jason Brownlee PhD[3]

Addressing class imbalance, a prevalent challenge in credit card default prediction, required a nuanced approach. Given the disproportionate representation of default versus non-default cases, there was a significant risk of the model developing a bias towards the majority class. To counter this, I employed data augmentation techniques, strategically synthesising new samples from the underrepresented class. This method not only balanced the dataset but also introduced a diversity of examples, enhancing the model's ability to generalise from its training data.

My model's architecture was carefully designed to capture the complex, non-linear relationships inherent in financial behavior data. The architecture begins with an input layer that directly corresponds to the dimensionality of the preprocessed dataset, ensuring a seamless transition of data into the model.

Following the input layer are two dense layers, both activated by the Rectified Linear Unit (ReLU) function. I chose ReLU for its efficiency and effectiveness in introducing non-linearity to the model, allowing it to learn more complex patterns in the data as per advice from an article by Jason Brownlee[4]. The first dense layer consists of 64 neurons, while the second comprises 32 neurons. This configuration was determined to offer a balance between model complexity and computational efficiency.

To address the binary classification task of predicting credit card defaults, the model culminates in a softmax output layer. The softmax function provides a probability distribution over the two possible outcomes, default and non-default, facilitating a clear and interpretable prediction.

One of the primary challenges in training deep neural networks is overfitting, where the model performs well on training data but fails to generalise to unseen data. To mitigate this, I incorporated dropout as a regularisation technique as per a scholarly review underlining the importance of proper regularisation[5], which also provided me with valuable insights for future research and development. Dropout randomly "drops" a subset of neurons in each training iteration, preventing the model from becoming overly reliant on any specific set of features. This introduces redundancy in the model's architecture, forcing it to learn more robust features that are generalisable across the dataset.

Overall, the proposed method combines thoughtful data preprocessing with a strategically designed model architecture and regularization techniques to address the challenges of credit card default prediction. Through normalisation, class imbalance correction, and the application of dropout, the model is positioned to learn effectively from the data, providing reliable predictions that can inform risk management strategies in the financial sector.

4 Experimental Results

In the development of our deep learning model for credit card default prediction, hyperparameter tuning emerged as the most important phase which I found out thanks to this article[6]. I focused on three primary hyperparameters: learning rate, batch size, and the number of epochs. These were optimised using a validation set, allowing us to refine the model's learning process without direct exposure to the test data.

The learning rate was initially set to a default value commonly used in the Adam optimiser, 0.001. However, through iterative testing and validation, adjustments were made to find an optimal balance that minimised loss without causing the model to overshoot the minimum of the loss function. Batch size was selected based on computational constraints and the desire for model stability during training; a size of 32 provided a good compromise between computational efficiency and the stability of gradient descent updates. The number of epochs was determined by observing the point of convergence for validation loss, with an early stopping criterion implemented to halt training when the validation loss ceased to decrease, preventing overfitting. This approach allowed me to run training for up to 50 epochs, though in practice, training often concluded earlier thanks to early stopping with a patience of 5 epochs, monitoring the validation loss. This approach ensured that training ceased when the model ceased to make significant learning progress.

I chose the Adam optimiser for its adaptive learning rate properties, which adjust the learning rate dynamically based on estimates of lower-order moments. This feature is particularly advantageous in handling the noise and sparsity encountered in large datasets, making Adam a suitable choice for our predictive modeling task. Next time I would try using HN Adam, which a modified version of the Adam Algorithm proposed by Mohammad Arafa in his research article[7], to improve its accuracy and convergence speed. The HN Adam algorithm is modified by automatically adjusting the step size of the parameter updates over the training epochs.

My model underwent rigorous evaluation, employing training and validation datasets to gauge performance accurately. This strategy ensured that my assessments reflected the model's ability to generalize beyond the data on which it was trained. The F1 score, a harmonic mean of precision and recall, was especially critical as it balanced the importance of correctly predicting defaults against the need to minimise false positives—a vital consideration in financial risk modeling.

The results of my experimental evaluation underscore the efficiency of my deep learning model in predicting credit card defaults. My model achieved high accuracy, consistently outperforming baseline machine learning models such as logistic regression and decision trees. These traditional models, while valuable in many contexts, often struggle with the complex patterns and imbalanced data characteristic of default prediction tasks, so next time I could use tools like NVIDIA's NVTabular[8] for efficient data processing, which show promise in matching or even exceeding traditional models' performance on complex datasets, which is critical for financial predictions.

The model's performance in terms of the F1 score (train F1 averaging 0.6771, val F1 averaging 0.6598) suggests a moderate ability to balance precision and recall. This balance is essential in credit card default prediction to equitably manage the costs associated with false negatives and false positives. While the scores indicate that the model has learned to a certain extent, they also highlight opportunities for enhancement. Achieving a higher F1 score would signify a model with a stronger capacity to accurately identify defaults, minimising both types of error in a manner critical for financial risk management. This could be done by utilising data augmentation and resampling, and feature engineering.

The confusion matrix and ROC curve further illustrated the model's strengths, with a high true positive rate and an acceptable false positive rate. The Area Under the ROC Curve (AUC) metric confirmed the model's ability to discriminate between default and non-default cases compared to baseline models.

Overall, my deep learning approach to credit card default prediction not only achieved high accuracy but also excelled in handling the challenges posed by an imbalanced dataset. The nuanced understanding of financial behaviors captured by the model's architecture, combined with optimal hyperparameter settings, gave the way for its success. These results provide a compelling case for the adoption of deep learning techniques in financial risk assessment, offering a more nuanced and effective tool for predicting defaults.

5 Summary

This report has successfully demonstrated the development and application of a deep learning model to predict credit card defaults. The model, built on a sequential neural network architecture, addresses significant challenges like imbalanced datasets and the intricacies of feature selection. Our methodological approach, emphasizing thorough data preprocessing and the strategic tuning of hyperparameters, has proven critical in optimizing the model's predictive performance.

The preprocessing phase, focusing on normalization and class balance, laid a solid foundation for model training. This step was essential for ensuring that the model could effectively learn from the data without bias toward the majority class. The architecture of the deep learning model, featuring layers with ReLU activation and a softmax output, alongside dropout for regularization, was designed to capture complex patterns in the data relevant to default prediction.

Experimental results showcased the model's strength in predicting credit card defaults, achieving high accuracy but F1 moderate scores. These metrics indicate the model's capability to accurately classify potential defaults but also indicate room for improvement, which needs to be done in a manner that balances precision and recall effectively.

Looking ahead, there's potential to enhance model performance further by exploring more sophisticated architectures and incorporating a wider array of data sources. Such advancements could refine the model's predictive accuracy and its applicability to a broader spectrum of financial risk assessment tasks.

In conclusion, the findings from this report affirm the value of deep learning in financial default prediction, particularly in overcoming common data challenges. The insights I gained here laid a foundation for future research and development in this area, with the goal of achieving more accurate, reliable, and efficient predictive models for credit card defaults.

References

[1] MathWorks. Compare deep learning networks for credit default prediction. *MATLAB & Simulink*, 2023. https://www.mathworks.com/.

[2] Yoichi Hayashi. Emerging trends in deep learning for credit scoring: A review. *Electronics*, 11(19):3181, 2022.

[3] Jason Brownlee. How to use data scaling improve deep learning model stability and performance. https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/, 2021. Accessed: 2023-09-24.

[4] Jason Brownlee. A gentle introduction to the rectified linear unit (relu). *Machine Learning Mastery*, 2021.

[5] Imrus Salehin and Dae-Ki Kang. A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics*, 12:3106, 2023.

[6] Analytics Vidhya. Fine-tuning neural network hyperparameters: Best practices. *Analytics Vidhya*, 2024.

[7] Mohamed Reyad, Amany Sarhan, and Mohammad Arafa. A modified adam algorithm for deep neural network optimization. *Neural Computing and Applications*, 35:1–18, 04 2023.

[8] NVIDIA Technical Blog. Deep learning vs machine learning challenger models for default risk with explainability. https://developer.nvidia.com/blog/deep-learning-vs-machine-learning-challenger-models-for-default-risk-with-explainabi 2023. Accessed: 2024-03-19.