Cost-Sensitive Loss Function Design for Credit Card Fraud Detection

Kevin Zong
kzong@umd.edu

Michael Obajemu
mobajemu@umd.edu

Emerald San
esan1@umd.edu

1. Introduction

The widespread adoption of credit cards has led to a significant increase in fraud, costing financial institutions billions of dollars annually [1]. Effective fraud detection systems are essential for minimizing these losses while maintaining customer trust. This paper addresses the problem of credit card fraud detection using machine learning techniques.

Fraud detection is fundamentally a classification problem, where the objective is to distinguish fraudulent transactions from legitimate ones. However, this task presents several unique challenges. First, datasets are extremely imbalanced, with fraudulent transactions typically representing less than 1 percent of all transactions. Second, the cost of misclassification is highly asymmetric. Missing a fraudulent transaction results in significant financial losses, while incorrectly flagging a legitimate transaction primarily causes customer inconvenience and operational overhead. Third, not all errors are equal. Missing a $10,000 fraudulent transaction has far greater business impact than missing a $10 transaction, yet standard machine learning approaches treat these errors identically.

We propose a novel loss function called Amount-Weighted Cost-Sensitive Focal Loss (AWCS-FL), specifically designed for credit card fraud detection. Our approach makes three key contributions: (1) transaction amount weighting that incorporates transaction amounts directly into the loss function, ensuring that errors on high-value transactions receive proportionally higher penalties, (2) asymmetric business costs reflecting the real cost difference between false negatives and false positives, and (3) a temporal decay mechanism that gives higher importance to recent misclassifications, allowing the model to adapt more quickly to evolving fraud patterns.

2. Problem statement

2.1. Dataset

We use the Credit Card Fraud Detection dataset from Kaggle, which contains 284,807 European credit card transactions from September 2013. This dataset is widely used as a benchmark for fraud detection research.

2.2. Input/Output

Input: A transaction represented as a feature vector $x \in \mathbb{R}^{30}$ containing PCA-transformed features V_1–V_{28}, time elapsed, and the transaction amount.

Output: A binary prediction $\hat{y} \in \{0, 1\}$, where $\hat{y} = 0$ indicates a legitimate transaction and $\hat{y} = 1$ indicates a fraudulent transaction.

2.3. Evaluation Metrics

To test how well the model works, we use several measures:

- Precision, Recall, F1 Score, and AUC-PR to check how accurate and balanced the model is.
- Total Financial Loss and Saved Loss Rate to measure how much money the model helps save by catching frauds.
- Cost-Weighted F1 to measure how well the model performs while considering the cost of mistakes.

2.4. Expected Results

Based on published results on this dataset, we establish the following expectations:

- Binary Cross-Entropy Baseline: Precision: 0.85-0.90, Recall: 0.75-0.85, F1: 0.80-0.87
- Focal Loss Baseline: Precision: 0.87-0.92, Recall: 0.80-0.90, F1: 0.83-0.91

Expected performance of proposed AWCS-FL:

- Hypothesis 1: 15-25% reduction in total financial losses.
- Hypothesis 2: 10-20% improvement in cost-weighted F1 compared to focal loss.
- Hypothesis 3: Statistical metrics competitive with focal loss (within 5%).
- Hypothesis 4: The approach will demonstrate better performance on high-value transactions.

2.5. Difference from baseline implementation

Our approach differs from existing work in several ways. Compared to standard neural networks, we incorporate transaction amounts directly into the loss function rather than treating all errors equally. Additionally, unlike focal loss, we add amount weighting and asymmetric business costs. Lastly, compared to existing cost-sensitive methods, we introduce temporal adaptation and provide a unified loss function combining all three components.

3. Technical approach

This project uses the Kaggle Credit Card Fraud Detection Dataset, which includes 284,807 transactions made by European cardholders in 2013. Only 492 of these are fraud cases, showing how rare fraudulent transactions are. Each record has 30 features that include the transaction time, amount, and other hidden (PCA-transformed) values that keep the data private.

3.1. Data preprocessing

Before training the model, we prepare and clean the data: We normalize the time feature using min-max scaling to [0,1] and apply a log transformation to the transaction amount followed by standardization so that very large numbers do not affect the model too much.

The data is split into 60% for training, 20% for validation, and 20% for testing, keeping the order of time to match how real transactions happen. Because the dataset is unbalanced (very few frauds compared to normal cases), we use class weights or oversampling methods like SMOTE to help the model learn from both classes fairly.

3.2. Model design

We create a neural network that uses a special loss function called the Amount Weighted Cost Sensitive Focal Loss (AWCSFL). This function improves older methods in three main ways: It gives more importance to high-value transactions, since losing more money should matter more. It gives recent transactions more weight, helping the model adjust to new types of fraud. It takes into account real business costs, where missing a fraud is worse than checking a real transaction. We also compare our model with other methods such as Binary Cross-Entropy, Weighted Binary Cross-Entropy, and Focal Loss to see how much improvement it brings.

3.3. Network setup

The neural network has: An input layer with 30 features Hidden layers with 64, 32, and 16 neurons, each using dropout to reduce overfitting.

An output layer with a sigmoid function that predicts whether a transaction is fraud or not. The model is trained with the Adam optimizer using a learning rate of 0.0001 and a batch size of 256.

3.4. Implementation

We implement the approach in PyTorch 2.0.

4. Intermediate/Preliminary results

The preprocessing, model setup, and baseline implementations have been completed. Training and initial testing are underway. Early baseline results show the following:

- Binary Cross-Entropy: F1 = 0.82, Precision = 0.88, Recall = 0.76
- Weighted BCE: F1 = 0.84, Precision = 0.89, Recall = 0.79
- Focal Loss: F1 = 0.86, Precision = 0.91, Recall = 0.81

The AWCS-FL model is expected to further improve Recall and cost-weighted F1 by emphasizing high-value frauds and newer transaction patterns.

References

[1] D. Robertson. Card fraud losses worldwide in 2023. The Nilson Report, 2025.