Detecting Fraud in Credit Card Transactions

Fatimah Alshaikh, Moustafa Makhlouf, Muhammad Haider Asif

Problem Statement

The goal of this project is to develop a Machine Learning model that is able to detect fraud in credit card transactions.

Hypotheses

1. *Discriminative Supervised Approach:* Can we accurately classify transactions as fraudulent/non-fraudulent using a regularized logistic regression model?
2. *Generative Supervised Approach:* Can we accurately classify fraudulent/non-fraudulent transactions using a Naive Bayesian model?
3. *Unsupervised Approach:* Can we cluster fraudulent/non-fraudulent transactions in their respective clusters using K-means clustering algorithm?
4. *Deep-Learning Approach:* Can we use convolutional neural network to accurately classify transactions as fraudulent/non-fraudulent?
5. *Optimization:* In the most accurate model, can we drop some non-relevant features to improve accuracy?

Data

We obtained the dataset from Kaggle, The data contains transactions made by credit cards over two days in September 2013 by european cardholders. Each example has 28 anonymous features, "Delta-T", "Amount", and "Class". The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. We pre-processed the dataset by checking all values are not null, appending an id, and then inserting all valid examples into a transactions table in sqlite3 database for faster data loading.

Supervised Learning Approach

Regularized Logistic Regression:

This model is best suited for this task because it uses a logistic function to model a binary dependent variable "Class". This model is L2 - regularized, and yields best results with 165 iterations.

Figure 1: accuracy vs number of iterations

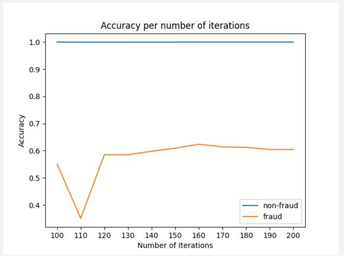
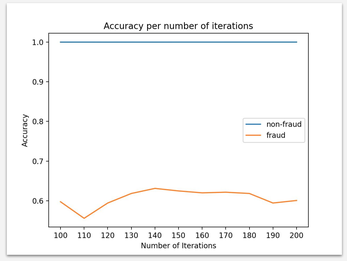


Figure 2: accuracy vs number of iterations without delta-T feature



- 99.91% overall accuracy
- 63.14% accuracy for fraudulent
- 99.96% accuracy for non-fraudulent

Gaussian Naive Bayes

Assuming gaussian likelihood and the independence of features from each other, this model gave us the best accuracy for the fraudulent transactions after removing "Delta_T" field from the data. This shows us that there is some form of independence in between the features of the dataset

- 97.77% overall acc.
- 82.05% - fraudulent acc
- 97.79% - non-fraudulent

Unsupervised Learning Approach

K-means Clustering:

Used the elbow point test to determine optimal k. Optimal K = 2, which makes sense as the transactions are to be classified as fraudulent or otherwise.

Figure 3: elbow point plot

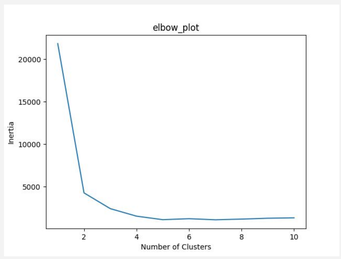
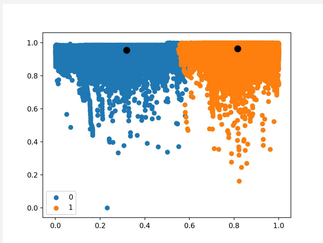


Figure 4: centroids (black) and respective clusters



Deep Learning Approach

Convolutional Neural Networks (CNN)

Our optimal model contained one 1D convolution layer to capture local feature patterns with Leaky RELU activation followed by a max-pooling layer and flattening layer before a fully connected layer with a softmax function to calculate class probabilities. We used an Adam optimizer with learning rate of 1e-5, and trained in batches of size 512.

- 99.91% overall accuracy
- 55.2% accuracy for fraudulent
- 99.99% accuracy for non-fraudulent

Figure 5: accuracy vs number of iterations

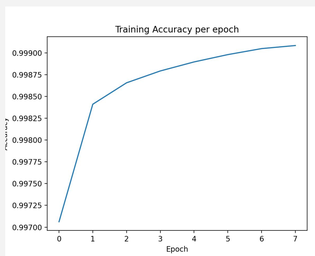
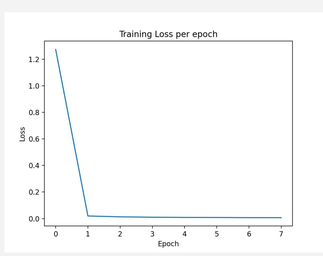


Figure 6: loss vs number of iterations



Conclusions

1. *Discriminative Supervised Approach:* Given the binary nature of the classification problem, a logistic regression model performs well on this dataset, second to Naive Bayes, with a 63.5% accuracy for fraudulent transactions.
2. *Generative Supervised Approach:* The model was able to classify the data with higher accuracy than the logistic regression model with a 82.05% accuracy for fraudulent transactions.
3. *Unsupervised Approach:* Using KMeans clusters did not provide any significant analysis of the dataset, due to the extreme imbalance of the dataset. Moreover, for a labeled dataset, unsupervised learning is not ideal.
4. *Deep-Learning Approach:* CNN's although theoretically could do well did not do as well on the fraudulent transactions as expected even after hyper parameter and architecture tuning
5. *Optimization:* Due to the anonymity of the fields, we had less flexibility with dropping features. Since we know 'Delta-T' is relative to the first entry, we decided to drop it, which improved the accuracy of our NB model by ~19%.

Limitations

The analysis made on this dataset doesn't provide any broad statements about credit card fraud due to the short time frame of the data (2 days), its specific location (Europe) and its skewed nature. The anonymity of the fields gave us very little room to manipulate the data before training. To learn to detect a very small number of fraud transactions, we think the model needs more data.