Tbrain X 玉山銀行
信用卡盜刷偵測

隊伍名稱: 阿罵我要吃糖果

Teams

香港中文大學統計系
金融業-資料科學家

工研院量測中心
友達光電韌體/資料工程師

香港中文大學工程系
軟件工程師

成功大學製造資訊與系統研究所
博士研究生

Agenda

- Define eval_metric
- Model Explanation
- Feature Engineering
- Summary
- Model Specification

Define eval_metric

Aim

1. Criteria for early stopping (Boosting, NN)

2. Assessment for model performance (Feature Selection)

Define eval_metric

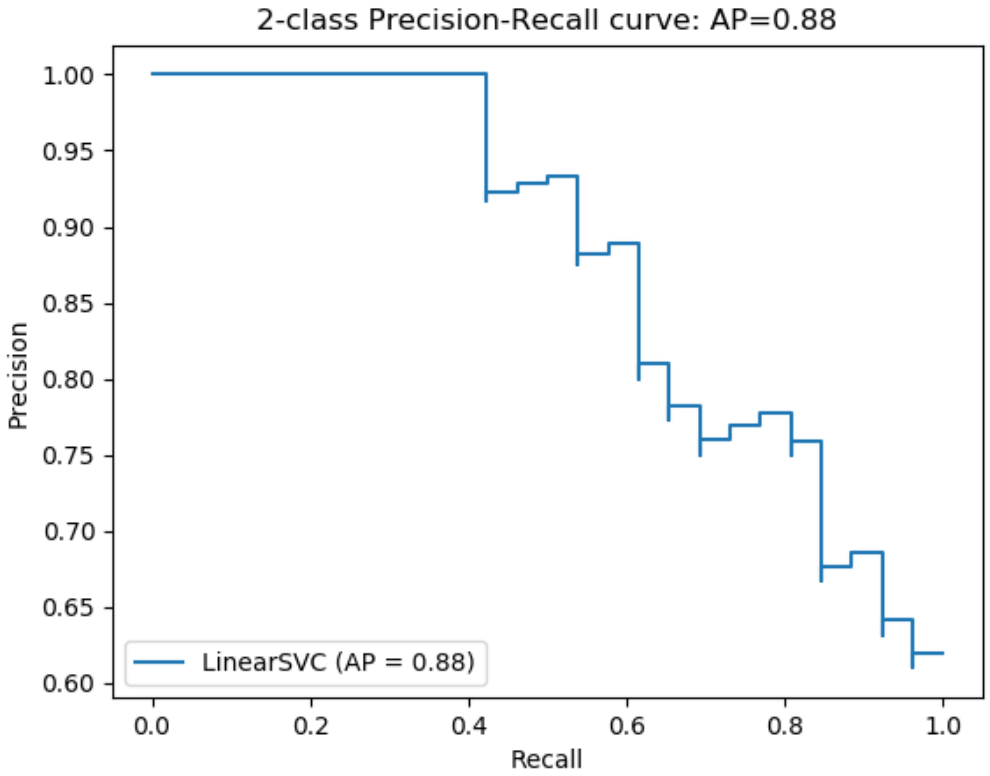
Problem of F1-Score

- Good performance in training but bad result in validation
- Easy to overfit

Define eval_metric

Average Precision Score

- Summarize a precision-recall curve



Define eval_metric

Problem of Average Precision Score

- The average precision score increases but maximum F1 Score decrease

Define eval_metric

Solution

- Using average precision score for early stopping
- Using maximum F1 Score for assessment of model performance

Model Explanation

- Total: 5 models(1 Base, 4 Special)
- 20 raw features, 13 preprocessed features, 4 special features
- Base model: raw features and preprocessed features.
- Special model : Base features plus one special feature.

Feature Engineering

Transaction Frequency Features

- The number of times of transaction during the period(30 days)
- The number of times of transaction during the day(locdt)
- The number of times of transaction for the merchant(mchno) during the period(30 days)

Feature Engineering

Time Features

- The time(loctm) difference between the transaction and previous/next transaction

- The standard deviation of the transaction time(loctm) during the day(locdt)

Feature Engineering

Transaction Amount Features

- The minimum/maximum amount(conam) of transaction of the card(cano) during the day(locdt)
- The time(loctm) difference between the transaction and the transaction with zero amount(conam)(if exist)

Feature Engineering

Merchant(mchno) Features

- The days(locdt) difference between the first and last transaction with same card(cano) and merchant(mchno)
- The n-th transaction with same card(cano) and merchant(mchno)

Feature Engineering

Changing Card Features

- The days(locdt) difference between the transaction and the last transaction with same card(cano)

- The days(locdt) difference between cardA(canoA) and cardB(canoB) with same user(bacno)

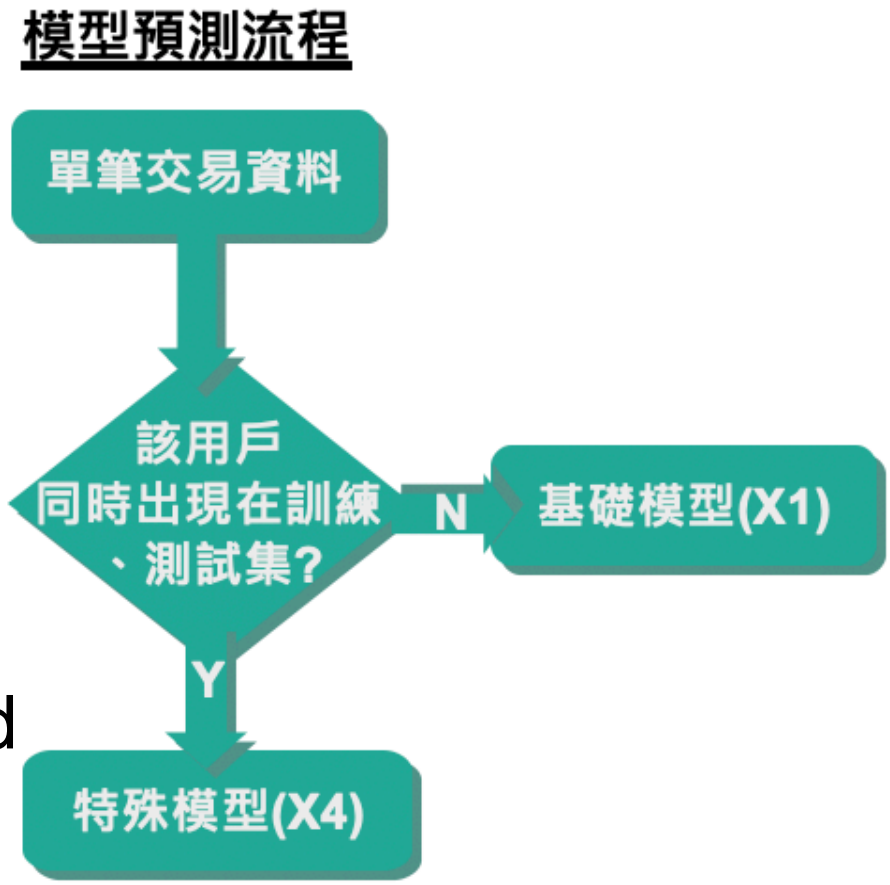
Feature Engineering

Special Features

- Whitelist/Blacklist of merchant(mchno) with same user(bacno)
- Blacklist of transaction amount(cano) with same card(cano)
- The days(locdt) difference between the transaction and the first fraudulent transaction(if exist)

Summary

- Some users(bacno) in both training and testing set.
- Capture the information of the label
- Note: May not work in real world



Model Specification

Simplified Model

- Apply limited number of new features
- Prevent overfitting
- Implement quickly
- Low Maintenance cost

Model Specification

Focus Learning

- Replace the value of categorical features of training set with NA if the value is not in testing set
- Reduce to learn something useless when apply in testing set

Model Specification

Using GroupKfold

- Use early stopping and split the training set by GroupKFold(bacno)
- Stop training once the model performance stops improving on a hold out validation dataset
- Model stop earlier (prevent overfitting)

Model Specification

Drop extreme cases

- Some of the predictions of testing set are very extreme between folds
- Drop if it is out of 1 standard deviation boundary
- Exclude the extreme cases
- Details: **https://github.com/aarontong95/TBrain_Credit_Card**

Thank You!