

PREDICTING DEFAULT CREDIT CARD CLIENT

Group project

UBC

Master of data science

OVERVIEW

- Problem: Predicting whether a credit card client will default or not
- Problem type: Classification (Default Or Not)
- Data set:
 - Credit transaction of credit card clients in Taiwan from Apr 2005 to September 2005
 - 30,000 examples
 - 23 explanatory variables



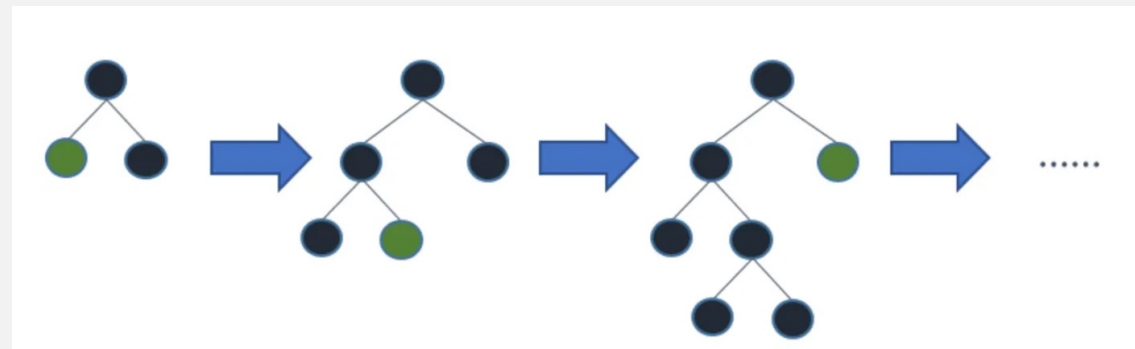
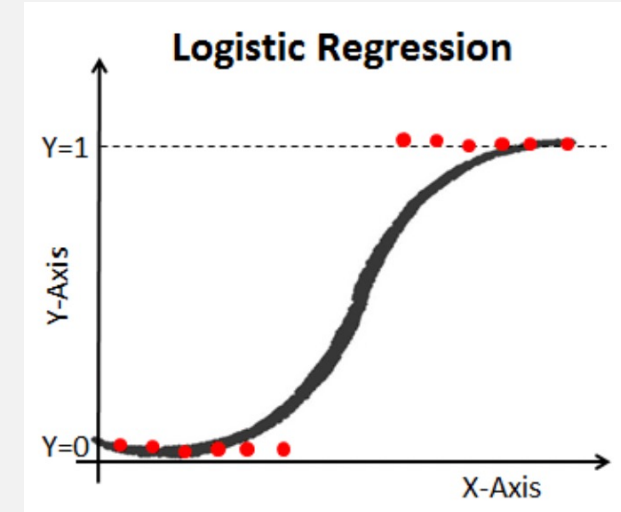
DATA PROCESSING

- Preliminary preprocessing: change feature names, NaN values
- Split the data into train (70%) and test (30%) portions randomly
- Perform exploratory data analysis on the train set
- Perform preprocessing and transformations: Drop features, identify feature types then transform the features accordingly



MODELS

- Choose metric: F1
- Apply Baseline model, linear model (logistic regression), and tuned logistic regression
- Try other model types, compare with above three models to find the best model:
 - LightGBM model
- Find out the most important features: PAY_1 (Repayment status in Sep 2005, value = -1, 1, 2, 3, ..., 9)
- Apply the best performing model on test set and get the score of selected metric



ROLE PLAYED

- The group includes 3 members
- I was the one who pick the data, kick off the project, draft all answers
- Other team members: check and revise the draft answers
- Lessons learned:
 - Only use related and important features
 - Select suitable metric
 - Not all machine learning models can beat linear model.