Financial Machine Learning

Homework 9

Due at 07:00 pm (Korea Standard Time) on Sunday, November 13.

Submit one file: written solutions with executable Python code

Problem 1. Textbook: Hands-on Machine Learning. Submit .ipynb file.

(a) Practice and review all the codes in the Textbook Chapter 12. Show that they work well.

(b) Practice and review all the codes in the Textbook Chapter 13. Show that they work well.

Problem 2. Credit Card Fraud Detection

Context

It is important that credit card companies can recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

Content

The dataset contains transactions made by credit cards in September 2013 by European cardholders.
This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable, and it takes value 1 in case of fraud and 0 otherwise.

Given the class imbalance ratio, we recommend measuring the accuracy using the Area Under the Precision-Recall Curve (AUPRC). Confusion matrix accuracy is not meaningful for unbalanced classification.

Acknowledgements