
HOMework ASSIGNMENT I

- Work individually.
- Submit a code and a short explanation along the code
- Any type of plagiarism will not be tolerated and will lead to disciplinary actions.
- **Due Date: 13th of August 2023, 23:59. Please submit your report through LMS. Only ONE submission per person.**

1 Introduction

In this assignment, you are expected to predict whether the given credit will be default or not. There are two datafiles `creditTrain.csv` and `creditTest.csv`. Please use `creditTrain.csv` to construct your model and `creditTest.csv` to validate your model. Our performance indicator will be “ROC AUC SCORE” over the test data. The data is explain in detail below. As we discussed in class, you are expected to use a “scoring” method and test the performance of that method in comparison with the one that uses all the features (remember the Housing example we have seen in class). If you have any questions, please send an email to: orsan.ozener@ozyegin.edu.tr

The binary variable, default payment (Yes = 1, No = 0), is the response variable. There are 23 variables as explanatory variables:

X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.

X2: Gender (1 = male; 2 = female).

X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).

X4: Marital status (1 = married; 2 = single; 3 = others).

X5: Age (year).

X6 - X11: History of past payment. The past monthly payment records (from April to September, 2005) as follows:

X6 = the repayment status in September, 2005;

X7 = the repayment status in August, 2005; . . .;

X11 = the repayment status in April, 2005.

The measurement scale for the repayment status is:

-1 = pay duly;
1 = payment delay for one month;
2 = payment delay for two months; . . .;
8 = payment delay for eight months;
9 = payment delay for nine months and above.

X12-X17: Amount of bill statement (NT dollar).
X12 = amount of bill statement in September, 2005;
X13 = amount of bill statement in August, 2005; . . .;
X17 = amount of bill statement in April, 2005.

X18-X23: Amount of previous payment (NT dollar).
X18 = amount paid in September, 2005;
X19 = amount paid in August, 2005; . . .;
X23 = amount paid in April, 2005.