# 

# I526 Applied Machine Learning

## Homework 0 – Dataset Selection

### Created By:

### Himanshu Goyal (hgoyal)

### Rituraj Singh (singhri)

### Problem: A 1 paragraph description of the data set, the list of features, the number of examples and the class label distribution.

#### Description of Data Set:

This dataset is for determining the solidness of borrowers via credit scoring. The given score for the categorical covariates is based on the assessment of experienced bank specialists dealing with credits. The source of this dataset is following link. <http://www.statistik.lmu.de/service/datenarchiv/kredit/kreditvar_e.html>

#### Features:

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **Description** | **Categories** | **Score** |
|
|
|
| CurAcctBal | Balance of current account | no balance or debit | 2 |
| 0 <= ... < 200 DM | 3 |
| ... >= 200 DM or checking account for at least 1 year | 4 |
| no running account | 1 |
| DurationMonth | Duration in months (categorized) | <=6 | 10 |
| 6 < ... <= 12 | 9 |
| 12 < ... <= 18 | 8 |
| 18 < ... <= 24 | 7 |
| 24 < ... <= 30 | 6 |
| 30 < ... <= 36 | 5 |
| 36 < ... <= 42 | 4 |
| 42 < ... <= 48 | 3 |
| 48 < ... <= 54 | 2 |
| > 54 | 1 |
| PrevPayments | Payment of previous credits | no previous credits / paid back all previous credits | 2 |
| paid back previous credits at this bank | 4 |
| no problems with current credits at this bank | 3 |
| hesitant payment of previous credits | 0 |
| problematic running account / there are further credits running but at other banks | 1 |
| CreditPurpose | Purpose of credit | new car | 1 |
| used car | 2 |
| items of furniture | 3 |
| radio / television | 4 |
| household appliances | 5 |
| repair | 6 |
| education | 7 |
| vacation | 8 |
| retraining | 9 |
| business | 10 |
| other | 0 |
| DMCreditAmt | Amount of credit in DM (categorized) | <=500 | 10 |
| 500 < ... <= 1000 | 9 |
| 1000 < ... <= 1500 | 8 |
| 1500 < ... <= 2500 | 7 |
| 2500 < ... <= 5000 | 6 |
| 5000 < ... <= 7500 | 5 |
| 7500 < ... <= 10000 | 4 |
| 10000 < ... <= 15000 | 3 |
| 15000 < ... <= 20000 | 2 |
| > 20000 | 1 |
| SavingsVal | Value of savings or stocks | < 100,- DM | 2 |
| 100,- <= ... < 500,- DM | 3 |
| 500,- <= ... < 1000,- DM | 4 |
| >= 1000,- DM | 5 |
| not available / no savings | 1 |
| EmpDuration | Has been employed by current employer for | unemployed | 1 |
| <= 1 year | 2 |
| 1 <= ... < 4 years | 3 |
| 4 <= ... < 7 years | 4 |
| >= 7 years | 5 |
| rate | Installment in % of available income | >= 35 | 1 |
| 25 <= ... < 35 | 2 |
| 20 <= ... < 25 | 3 |
| < 20 | 4 |
| MaritalStatus | Marital Status / Sex | male: divorced / living apart | 1 |
| female: divorced / living apart / married | 2 |
| male: single | 2 |
| male: married / widowed | 3 |
| female: single | 4 |
| Gaurantors | Further debtors / Guarantors | none | 1 |
| Co-Applicant | 2 |
| Guarantor | 3 |
| StayDurantion | Living in current household for | < 1 year | 1 |
| 1 <= ... < 4 years | 2 |
| 4 <= ... < 7 years | 3 |
| >= 7 years | 4 |
| MostValAsset | Most valuable available assets | Ownership of house or land | 4 |
| Savings contract with a building society / Life insurance | 3 |
| Car / Other | 2 |
| not available / no assets | 1 |
| Age | Age in years (categorized) | 0 <= ... <= 25 | 1 |
| 26 <= ... <= 39 | 2 |
| 40 <= ... <= 59 | 3 |
| 60 <= ... <= 64 | 5 |
| >= 65 | 4 |
| FurtherCredits | Further running credits | at other banks | 1 |
| at department store or mail order house | 2 |
| no further running credits | 3 |
| ApptType | Type of apartment | rented flat | 2 |
| owner-occupied flat | 3 |
| free apartment | 1 |
| PrevCreditCount | Number of previous credits at this bank (including the running one) | one | 1 |
| two or three | 2 |
| four or five | 3 |
| six or more | 4 |
| Occupation | Occupation | unemployed / unskilled with no permanent residence | 1 |
| unskilled with permanent residence | 2 |
| skilled worker / skilled employee / minor civil servant | 3 |
| executive / self-employed / higher civil servant | 4 |
| PersonMaintCount | Number of persons entitled to maintenance | 0 to 2 | 2 |
| 3 and more | 1 |
| Telephone | Telephone | no | 1 |
| yes | 2 |
| Foreignworker | Foreign worker | yes | 1 |
| no | 2 |
| Creditabilty | Creditabilty | credit-worthy | 1 |
| not-credit-worthy | 0 |

#### Number of Examples:

1000

#### Class Label Distribution:

Not Credit Worthy – 300

Credit Worthy- 700