Description of the German credit dataset.

1. Title: German Credit data

2. Source Information

Professor Dr. Hans Hofmann
Institut f"ur Statistik und "Okonometrie
Universit"at Hamburg
FB Wirtschaftswissenschaften
Von-Melle-Park 5
2000 Hamburg 13

3. Number of Instances: 1000

Two datasets are provided. the original dataset, in the form provided
by Prof. Hofmann, contains categorical/symbolic attributes and
is in the file "german.data".

For algorithms that need numerical attributes, Strathclyde University
produced the file "german.data-numeric". This file has been edited
and several indicator variables added to make it suitable for
algorithms which cannot cope with categorical variables.   Several
attributes that are ordered categorical (such as attribute 17) have
been coded as integer.   This was the form used by StatLog.

6. Number of Attributes german: 20 (7 numerical, 13 categorical)
   Number of Attributes german.numer: 24 (24 numerical)

7. Attribute description for german

Attribute 1: (qualitative)
            Status of existing checking account
              A11 :   ... <   0 DM
            A12 : 0 <= ... < 200 DM
            A13 :   ... >= 200 DM /
               salary assignments for at least 1 year
              A14 : no checking account

Attribute 2: (numerical)
            Duration in month

Attribute 3: (qualitative)
            Credit history
            A30 : no credits taken/
               all credits paid back duly
              A31 : all credits at this bank paid back duly
            A32 : existing credits paid back duly till now
              A33 : delay in paying off in the past
            A34 : critical account/
               other credits existing (not at this bank)

Attribute 4: (qualitative)

```
            Purpose
            A40 : car (new)
            A41 : car (used)
            A42 : furniture/equipment
            A43 : radio/television
            A44 : domestic appliances
            A45 : repairs
            A46 : education
            A47 : (vacation - does not exist?)
            A48 : retraining
            A49 : business
            A410 : others

Attribute 5:  (numerical)
            Credit amount

Attibute 6:  (qualitative)
            Savings account/bonds
            A61 :           ... <   100 DM
            A62 :    100 <= ... <   500 DM
            A63 :    500 <= ... < 1000 DM
            A64 :           .. >= 1000 DM
              A65 :   unknown/ no savings account

Attribute 7:  (qualitative)
            Present employment since
            A71 : unemployed
            A72 :         ... < 1 year
            A73 : 1  <= ... < 4 years
            A74 : 4  <= ... < 7 years
            A75 :         .. >= 7 years

Attribute 8:  (numerical)
            Installment rate in percentage of disposable income

Attribute 9:  (qualitative)
            Personal status and sex
            A91 : male   : divorced/separated
            A92 : female : divorced/separated/married
              A93 : male   : single
            A94 : male   : married/widowed
            A95 : female : single

Attribute 10: (qualitative)
            Other debtors / guarantors
            A101 : none
            A102 : co-applicant
            A103 : guarantor

Attribute 11: (numerical)
            Present residence since

Attribute 12: (qualitative)
            Property
            A121 : real estate
```

```
            A122 : if not A121 : building society savings agreement/
                            life insurance
               A123 : if not A121/A122 : car or other, not in attribute 6
            A124 : unknown / no property

Attribute 13: (numerical)
            Age in years

Attribute 14: (qualitative)
            Other installment plans
            A141 : bank
            A142 : stores
            A143 : none

Attribute 15: (qualitative)
            Housing
            A151 : rent
            A152 : own
            A153 : for free

Attribute 16: (numerical)
              Number of existing credits at this bank

Attribute 17: (qualitative)
            Job
            A171 : unemployed/ unskilled  - non-resident
            A172 : unskilled - resident
            A173 : skilled employee / official
            A174 : management/ self-employed/
                 highly qualified employee/ officer

Attribute 18: (numerical)
            Number of people being liable to provide maintenance for

Attribute 19: (qualitative)
            Telephone
            A191 : none
            A192 : yes, registered under the customers name

Attribute 20: (qualitative)
            foreign worker
            A201 : yes
            A202 : no
```

8. Cost Matrix

This dataset requires use of a cost matrix (see below)

```
        1        2
----------------------------
   1    0        1
----------------------
```

```
   2    5          0
```

(1 = Good, 2 = Bad)

the rows represent the actual classification and the columns
the predicted classification.

It is worse to class a customer as good when they are bad (5),
than it is to class a customer as bad when they are good (1).