FIT5186 Intelligent Systems

Week 6 Tutorial

In this tutorial, you will be considering a realistic classification problem from the banking sector. This example is commonly considered part of *data mining*, since we are trying to extract information from a large data set.

The problem:
Suppose we have 1000 previous applications for credit at a bank, together with those applicants' subsequent classification as "good" or "bad" customers. Can we learn to classify new applicants as "good" or "bad" based only on their applications? In other words, can we learn which combinations of responses indicate a "good" credit risk and which indicate a "bad" credit risk?

The data:
The questions used in the application form can be seen on the following pages. The file credit.txt contains the original responses to these questions (1000 rows = 700 good & 300 bad). This data has been pre-processed to replace categorical data with numerical data. The pre-processed file is called credit.xls. An example showing the transformation of the first 5 rows of the data is shown below:

SAMPLE RESPONSES TO ORIGINAL QUESTIONS (in credit.txt)

A11 6 A34 A43 1169 A65 A75 4 A93 A101 4 A121 67 A143 A152 2 A173 1 A192 A201 1
A12 48 A32 A43 5951 A61 A73 2 A92 A101 2 A121 22 A143 A152 1 A173 1 A191 A201 2
A14 12 A34 A46 2096 A61 A74 2 A93 A101 3 A121 49 A143 A152 1 A172 2 A191 A201 1
A11 42 A32 A42 7882 A61 A74 2 A93 A103 4 A122 45 A143 A153 1 A173 2 A191 A201 1
A11 24 A33 A40 4870 A61 A73 3 A93 A101 4 A124 53 A143 A153 2 A173 2 A191 A201 2

SAMPLE RESPONSES AFTER PRE-PROCESSING OF DATA (in credit.xls)

Q1	Q2	Q3	Q5	Q6	Q7	Q9	Q11	Q12	Q13	Q14	Q16	Q18	Q19	Q20	Q4	Q10	Q15	Q17	CLASS
1	6	4	12	5	5	3	4	1	67	3	2	1	2	1	0 0	1 0	0 1	0 0 1	1
2	48	2	60	1	3	2	2	1	22	3	1	1	1	1	0 0	1 0	0 1	0 0 1	2
4	12	4	21	1	4	3	3	1	49	3	1	2	1	1	0 0	1 0	0 1	0 1 0	1
1	42	2	79	1	4	3	4	2	45	3	1	2	1	1	0 0	0 0	0 0	0 0 1	1
1	24	3	49	1	3	3	4	4	53	3	2	2	1	1	1 0	1 0	0 0	0 0 1	2

The known classification of each customer in the data set as "good" or "bad" can be coded as either:
credit class - 1 or 2 (a single column) or good - 1 0 and bad 0 1 (2 columns)

The Task:
Use NeuroShell 2 to learn the classification. Experiment with using a single output (predicting a 1 or 2) and two outputs (predicting 1 0 or 0 1). For this example, the final predictions would be rounded to generate a classification (i.e. 0.8 0.4 would be classified as "good"), so R-squared is not very useful. Instead, count the mismatches.

QUESTIONS FOR CREDIT APPLICATION EXAMPLE (GERMAN DATA)

QUESTION 1: (qualitative)
Status of existing checking account
A11 : ... < 0 DM
A12 : 0 <= ... < 200 DM
A13 : ... >= 200 DM
A14 : no checking account

QUESTION 2: (numerical)
Duration of checking account in months

QUESTION 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)

QUESTION 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

QUESTION 5: (numerical)
Credit amount

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

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

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

QUESTION 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

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

QUESTION 11: (numerical)
Number of years in current residence

QUESTION 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 QUESTION 6
A124 : unknown / no property

QUESTION 13: (numerical)
Age in years

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

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

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

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

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

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

QUESTION 20: (qualitative)
foreign worker
A201 : yes
A202 : no