

Descrição do conjunto de dados de crédito alemão.

1. Título: Dados de crédito alemães

2. Informações da fonte

Professor Dr. Hans Hofmann  
Institut für Statistik und "Okonometrie  
Universität Hamburg  
FB Wirtschaftswissenschaften  
Von-Melle-Park 5  
2000 Hamburg 13

3. Número de instâncias: 1000

Dois conjuntos de dados são fornecidos. o conjunto de dados original, no formulário fornecido pelo Prof. Hofmann, contém atributos categóricos/simbólicos e está no arquivo "german.data".

Para algoritmos que precisam de atributos numéricos, a Universidade de Strathclyde produziu o arquivo "german.data-numeric". Este arquivo foi editado e várias variáveis indicadoras adicionadas para torná-lo adequado para algoritmos que não conseguem lidar com variáveis categóricas. Vários atributos que são ordenados categóricos (como o atributo 17) têm sido codificados como inteiro. Este foi o formulário usado pelo StatLog.

6. Número de atributos alemão: 20 (7 numéricos, 13 categóricos)  
Número de atributos german.numer: 24 (24 numéricos)

7. Descrição do atributo para alemão

Atributo 1: (qualitativo)  
Status da conta corrente existente  
A11 : ... < 0 DM  
A12 : 0 <= ... < 200 DM  
A13 : ... >= 200 DM /  
atribuições salariais por pelo menos 1 ano  
A14 : sem conta corrente

Atributo 2: (numérico)  
Duração em mês

Atributo 3: (qualitativo)  
Histórico de crédito  
A30 : sem créditos gozados/  
todos os créditos pagos de volta devidamente  
A31 : todos os créditos neste banco devidamente reembolsados  
A32: créditos existentes devidamente reembolsados até agora  
A33 : atraso no pagamento no passado  
A34 : relato crítico/  
outros créditos existentes (não neste banco)

Atributo 4: (qualitativo)  
Propósito  
A40 : carro (novo)  
A41 : carro (usado)  
A42 : móveis/equipamentos

A43 : rádio/televisão  
 A44 : eletrodomésticos  
 A45 : reparações  
 A46 : educação  
 A47 : (férias - não existe?)  
 A48: reciclagem  
 A49 : negócios  
 A410 : outros

Atributo 5: (numérico)  
 Valor do crédito

Atributo 6: (qualitativo)  
 Conta poupança/obrigações  
 A61 : ... < 100 DM  
 A62 : 100 <= ... < 500 DM  
 A63 : 500 <= ... < 1000 DM  
 A64 : .. >= 1000 DM  
 A65: desconhecido/sem conta poupança

Atributo 7: (qualitativo)  
 Emprego atual desde  
 A71 : desempregados  
 A72 : ... < 1 ano  
 A73 : 1 <= ... < 4 anos  
 A74 : 4 <= ... < 7 anos  
 A75 : .. >= 7 anos

Atributo 8: (numérico)  
 Taxa de parcelamento em percentual da renda disponível

Atributo 9: (qualitativo)  
 Status pessoal e sexo  
 A91 : masculino : divorciado / separado  
 A92 : feminino : divorciado / separado / casado  
 A93 : masculino : solteiro  
 A94 : masculino : casado/viúvo  
 A95 : fêmea : solteiro

Atributo 10: (qualitativo)  
 Outros devedores / fiadores  
 A101 : nenhum  
 A102 : correquerente  
 A103 : fiador

Atributo 11: (numérico)  
 Residência atual desde

Atributo 12: (qualitativo)  
 Propriedade  
 A121 : Imobiliário  
 A122 : se não A121 : acordo de poupança da sociedade de construção/  
 seguro de vida  
 A123 : se não A121/A122 : carro ou outro, não no atributo 6  
 A124 : desconhecido / sem propriedade

Atributo 13: (numérico)  
 Idade em anos

Atributo 14: (qualitativo)  
 Outros planos de parcelamento

A141 : banco  
 A142 : lojas  
 A143 : nenhum

Atributo 15: (qualitativo)

Habitação

A151 : arrendamento

A152 : próprio

A153 : de graça

Atributo 16: (numérico)

Número de créditos existentes neste banco

Atributo 17: (qualitativo)

Trabalho

A171 : desempregados/não qualificados - não residentes

A172 : não qualificado - residente

A173 : empregado qualificado / funcionário

A174 : Gestão/ Trabalhadores por conta própria/

Funcionário/Oficial Altamente Qualificado

Atributo 18: (numérico)

Número de pessoas responsáveis por prestar alimentos

Atributo 19: (qualitativo)

Telefone

A191 : nenhum

A192 : sim, registado em nome do cliente

Atributo 20: (qualitativo)

trabalhador estrangeiro

A201 : sim

A202 : não

## 8. Matriz de Custos

Este conjunto de dados requer o uso de uma matriz de custos (veja abaixo)

1 2

-----

1 0 1

-----

2 5 0

(1 = Bom, 2 = Ruim)

As linhas representam a classificação real e as colunas  
 a classificação prevista.

É pior classificar um cliente como bom quando ele é ruim (5),  
 do que classificar um cliente como ruim quando ele é bom (1).