Case Study 1
Credit Scoring

Jerzy Korczak
http://www.korczak-leliwa.pl
http://kti.ue.wroc.pl
email:jerzy.korczak@ue.wroc.pl

Summer School on Data Science, Wrocław, 2016

Outline

Problem definition
Data sets
Models: Logistic Regression, IDT, CN2, Neural Nets, etc.
Model building issues
Model evaluation
Research and implementation

Problem definition

Basis for financial institutions to evaluate the likelihood for credit applicants to default

- Goal: Discover a credit scoring model from data (grant or reject credit)
- Given: data set of applications for a bank loan containing:
  - Population repays their loan -> labeled "good"
  - Population goes into some form of default -> labeled "bad"
- Model building
  - Build a model that can discriminate population 2a from population 2b
  - Usually treated as a classification problem
  - Typically want to estimate p(good | features) and rank applications
- Widely used by banks and credit card companies
  - Similar problems occur in direct marketing and other "scoring" applications

Different applications for Credit Scoring

- Other financial applications:
  - Delinquent loans: who is most likely to pay up
    - Uses historical data on who paid in the past
    - Often used to create "portfolios" of delinquent debt
  - Customer revenue
    - How much will each customer generate in revenue over the next K years
- Predicting marketing response
  - Cost of a mailer to a customer is order of $1 dollar
  - Targeted marketing
    - Rank customers in terms of "likelihood to respond"
- "Churn" prediction
  - Predicting which customers are most likely to switch to another brand
  - E.g., wireless phone service
  - Scores used to rank customers and then target most likely with incentives
- Many more….

Data sets

- Data from the loan application
  - Age, address, income, profession, SS#, number of credit cards, savings, etc
  - Easy to obtain
- Internal Performance data
  - How the individual has performed on other loans with the same bank
  - May only be available for a subset of customers
- External Performance data:
  - Credit Reports
    - How the individual has performed historically on all loans and credit cards
    - Relatively expensive to obtain (e.g., $1 per individual)
  - Court Judgements
  - Real Estate records
- Macro-level external data
  - Demographic characteristics for applicant's, zip code or census tract

Data sets

- Issues
  - Data entry errors (e.g., birthday = date of loan application)
  - Deliberate falsifications (e.g., over-reporting of income)
  - Data sources
  - Legal issues
    - US Equal Credit Opportunity Acts, 1975/76
    - Illegal to use race, color, religion, national origin, sex, marital status, or age in the decision to grant credit
    - But what if other variables are highly predictive of some of these variables?

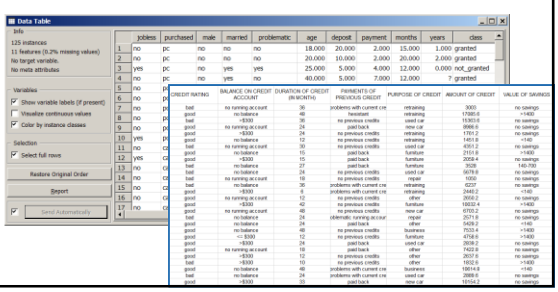
Defining Good and Bad

- Good vs Bad
  - Not necessarily clear how to define 2 classes
    e.g.,
    - bad = ever 3 or more payments in arrears?
    - bad = 2 or more payments in arrears more than once?
  - A "spectrum" of behavior
    - Never any problems in payments
    - Occasional problems
    - Persistent problems
  - Typical to discard the intermediate cases and also those with insufficient experience to reliably classify them
    - Not ideal theoretically, but convenient

Data Set for Model Building

- Sample selection
  - Academic benchmarks: Japaneese Credit Data, German Data, US Data
  - Typical sample sizes ~ 10k to 100k per class
  - Should be representative of customers who will apply in the future
  - Need to be able to get the relevant variables for this set of customers
    - Internal performance data
    - External performance data
- External data sources (e.g., credit reports) can result in a very large number of possible variables
  - e.g., in the 1000's
  - e.g., "number of accounts opened in past 12/24/36/… months"
  - Typically some form of variable selection done before building a model
    - Often based on univariate criteria such as information gain

Examples of datasets


Data Table

Info
125 instances
11 features (0.2% missing values)
No target variable.
No meta attributes

Variables
Show variable labels (if present)
Visualize continuous values
Color by instance classes

Selection
Select full rows

Restore Original Order
Report
Send Automatically

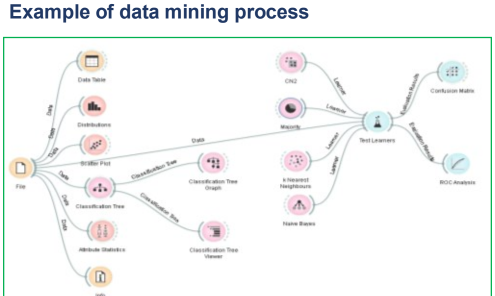
	jobless	purchased	male	married	problematic	age	deposit	payment	months	years	class
1	no	pc	no	no	no	18.000	20.000	2.000	15.000	1.000	granted
2	no	pc	no	no	no	20.000	10.000	2.000	20.000	2.000	granted
3	yes	pc	no	yes	yes	25.000	5.000	4.000	12.000	0.000	not_granted
4	no	pc	no	yes	no	40.000	5.000	7.000	12.000	?	granted
5	no										
6	no										
7	no										
8	no										
9	no										
10	yes										
11	no										
12	yes										
13	no										
14	no										
15	no										
16	no										
17	no										

CREDIT RATING	BALANCE ON CREDIT ACCOUNT	DURATION OF CREDIT (IN MONTH)	PAYMENTS OF PREVIOUS CREDIT	PURPOSE OF CREDIT	AMOUNT OF CREDIT	VALUE OF SAVINGS
bad	no running account	36	problems with current cre	retraining	3003	no savings
good	no balance	48	hesitant	retraining	17085.6	>1400
bad	>$300	36	no previous credits	used car	15363.6	no savings
good	no running account	24	paid back	new car	8986.6	no savings
good	>$300	24	no previous credits	retraining	1761.2	no savings
good	no balance	12	no previous credits	retraining	1451.8	<140
bad	no running account	30	no previous credits	used car	4351.2	no savings
good	no balance	15	paid back	furniture	2151.8	>1400
good	>$300	15	paid back	furniture	2059.4	no savings
bad	no balance	27	paid back	furniture	3528	140-700
bad	no balance	24	no previous credits	used car	5679.8	no savings
bad	no running account	18	no previous credits	repair	1050	no savings
bad	no balance	36	problems with current cre	retraining	6237	no savings
good	>$300	6	problems with current cre	retraining	2440.2	<140
good	no running account	12	no previous credits	other	2650.2	no savings
good	>$300	42	no previous credits	furniture	10032.4	>1400
good	no running account	48	no previous credits	new car	6703.2	no savings
bad	no balance	24	oblematic running accour	repair	2571.8	no savings
good	no balance	24	paid back	other	5429.2	<140
good	no balance	48	no previous credits	business	7533.4	>1400
good	<= $300	12	no previous credits	furniture	4758.6	>1400
good	>$300	24	paid back	used car	2839.2	no savings
good	no running account	18	paid back	other	7422.8	no savings
good	>$300	12	no previous credits	other	2637.6	no savings
bad	>$300	10	no previous credits	other	1832.6	>1400
good	no balance	48	problems with current cre	business	10814.8	<140
bad	no balance	24	problems with current cre	used car	2888.6	no savings
good	>$300	33	paid back	new car	10154.2	no savings

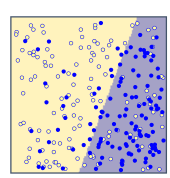
Models used in Credit Scoring

- Regression:
  - Ignore the fact that we are estimating a probability
  - Typically linear regression is used
- Classification (more common approach)
  - Logistic regression (most widely used)
  - Inductive Decision trees (becoming more popular)
  - Neural networks (experimented with, but not used in practice so much)
  - SVMs – recently used
  - Model combining - some work in this area
- General comments
  - Many trade-secrets, companies do not publish details
  - Generally the industry is conservative: prefer well-established methods
  - Classification accuracy is only one part of the overall solution….

Example of data mining process



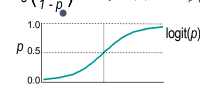
Logistic Regression Models



Training Data

logit(p)

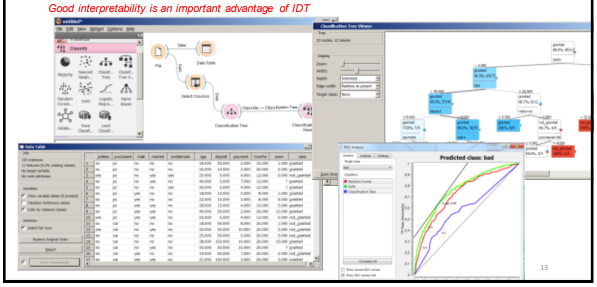
$$\log\left(\frac{p}{1-p}\right) = w_0 + w_1x_1 + \ldots + w_px_p$$



Note that near 0, logit(p) is almost linear, so linear and logistic regression will be similar

Inductive Decision Trees – Algorithm C4.5

Good interpretability is an important advantage of IDT



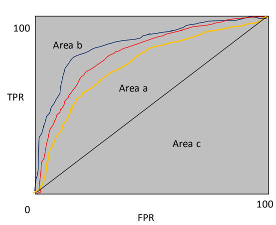
Evaluation Methods

- Decile/Centile reporting:
  - Rank customers by predicted scores
  - Report "lift" rate in each decile (and cumulatively) compared to accepting eve
- Receiver Operation Characteristics
  - Vary classification threshold
  - Plot proportion of good risks accepted vs. bad risks accepted
- Bad Risk rate = bad risk among those accepted
  Let p = proportion of good risks
  Let a = proportion accepted

  e.g., can show that, with a > p, the bad risk rate among those accepted is lower bounded by 1 – p/a

  e.g., p = 0.45, a =0.70 => bad risk rate must be between 0.35 and 0.78

Measure of performance



ROC-curve, AUC and Gini coefficient
- Prediction accuracy
- AUC Area under the curve
  AUC=area a + area c
- Gini coefficient
  Gini=area a/(area a + area b)
- Gini=2*AUC-1

Research and Implementation Issues

- Threshold selection
  - Above what threshold should loans be granted
  - Depends on goals of the project
    e.g., focusing on a small set of high-scoring customers versus "widening the net" to include a larger number (but still minimizing risk)
- Time-dependent classification
  - What really matters is what the customer will do at time t+T
  - Can we model the "state" of a customer (rather than statically)?
- Overrides
  - Loans are still manually "signed-off". The bank may sometimes override the system's recommendation

Research and Implementation Issues

- Implementation
  - Depends on whether the model is replacing an existing automated model or is the first time modeling is being applied to the problem
  - Many software issues in terms of databases, security, etc
- Monitoring and tracking
  - Important to see how the scorecard works in practice
  - Generating monthly/quarterly reports on scorecard performance
  - Analyzing in detail at performance on segments, by attribute, etc
- Time for a new model?
  - e.g., population has changed significantly
  - e.g., new (cheap and useful) data available
  - e.g., new modeling technology available

References

B. Baesens, B. Setiono, R. Mues, A.J. Vanthienen, *Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation*. Management Science, vol. 49, 2003, pp. 312-329.

D. J. Hand and W. E. Henley, *Statistical Classification Methods in Consumer Credit Scoring: a Review*, J. of the Royal Statistical Society: Series A, Vol. 160: Issue 3, Nov 1997

A. Kraus, H. Kuchenhoff, *Credit Scoring and Credit Control*, Univ. Munchen, 2013

E. Mays (ed.), *Credit Risk Modeling*, American Management Association, 1998.

O. Renault, A. De Servigne, *The Standard & Poor's Guide to Measuring and Managing Credit Risk*, McGraw-Hill, 2004.

L. C. Thomas, D. B. Edelman, J. N. Crook, *Credit Scoring and its Applications*: SIAM, 2002