Customer Analytics

3 Mar 2013

István Nagy

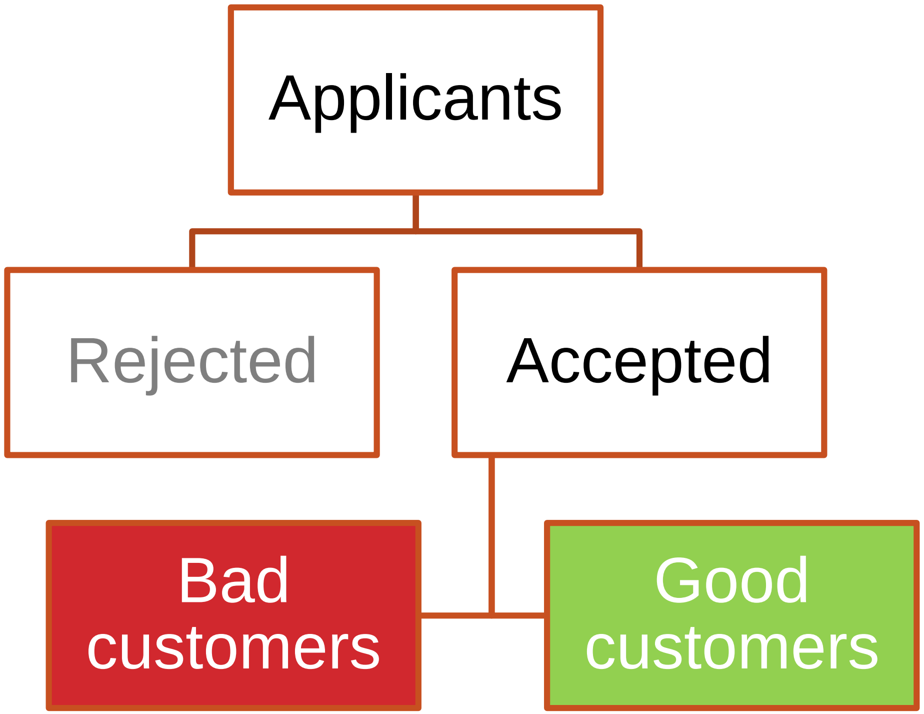
Credit Scoring in Banking

BUSINESS UNDERSTANDING

- **Basic concepts of credit scoring and credit scorecards**
- **Explain why they are used, list reasons for the popularity and success of credit scoring and credit scorecards**
- **Performance and sample window**

CREDIT SCORING

- **Applicants apply for a bank loan**
    - Population 1 is rejected
    - Population 2 is accepted
        - Population 2a repays their loan → labeled good
        - Population 2b goes into some form of default → labeled bad
- **Data mining model building**
    - Build a model that can discriminate population 2a from 2b



Source: Padhraic Smyth, Data mining course, **Credit Scoring**, Department of Information and Computer Science

CREDIT SCORING

Credit scoring is defined as a data mining model that assigns a risk value to prospective or existing credit accounts.

The task is determining the probability that a potential borrower will default on his or her obligations, given the personal characteristics of the borrower.

A credit scorecard is a data mining risk model that was put into a special format designed for ease of interpretation.

After the credit scorecard model is built, it is used to make strategic decisions, such as accepting/rejecting applicants, deciding when to raise a credit line.

CREDIT SCORECARDS

Credit scorecard is a model that was put into a special format, in its simple form, a scorecard is built of a number of characteristics.

Characteristic Name	Attribute	Scorecard Points
AGE	. -> 23	63
AGE	23 -> 25	76
AGE	25 -> 28	79
AGE	28 -> 34	85
AGE	34 -> 46	94
AGE	46 -> 51	103
AGE	51 -> .	105
CARDS	"AMERICAN EXPRESS," "VISA OTHERS," "VISA MYBANK," "NO CREDIT CARDS"	80
CARDS	"CHEQUE CARD," "MASTERCARD/EUROC," "OTHER CREDIT CARD"	99
EC_CARD	0	86
EC_CARD	1	83
INCOME	. -> 500	93
INCOME	500 -> 1,550	81
INCOME	1,550 -> 1,850	75
INCOME	1,850 -> 2,550	80
INCOME	2,550 -> .	88
STATUS	"E," "T," "U"	79

Source: Naeem Siddiqi, **Credit Risk Scorecards**, Developing and Implementing Intelligent Credit Scoring, Wiley, 2006

CREDIT SCORECARDS

Credit scorecard is a model that was put into a special format designed for ease of interpretation.

In its simples form, a scorecard is built of a number of characteristics.

Each characteristic is comprised of a number of attributes.

Each attribute is associated with a number of scorecard points.

Characteristic Name	Attribute	Scorecard Points
AGE	. -> 23	63
AGE	23 -> 25	76
AGE	25 -> 28	79
AGE	28 -> 34	85
AGE	34 -> 46	94
AGE	46 -> 51	103
AGE	51 -> .	105
CARDS	"AMERICAN EXPRESS," "VISA OTHERS," "VISA MYBANK," "NO CREDIT CARDS"	80
CARDS	"CHEQUE CARD," "MASTERCARD/EUROC," "OTHER CREDIT CARD"	99
EC_CARD	0	86
EC_CARD	1	83
INCOME	. -> 500	93
INCOME	500 -> 1,550	81
INCOME	1,550 -> 1,850	75
INCOME	1,850 -> 2,550	80
INCOME	2,550 -> .	88
STATUS	"E," "T," "U"	79

Source: Naeem Siddiqi, **Credit Risk Scorecards**, Developing and Implementing Intelligent Credit Scoring, Wiley, 2006

CREDIT SCORECARDS

Credit scorecard is a model that was put into a special format designed for ease of interpretation.

In its simples form, a scorecard is built of a number of characteristics.

Each characteristic is comprised of a number of attributes.

Each attribute is associated with a number of scorecard points.

These scorecard points are statistically assigned to differentiate risk, based on the predictive power of the variables, correlation between variables and business considerations.

The total score of an applicant is the sum of the scores for each attribute present in the scorecard. Smaller scores imply a higher risk of default, and vice versa, higher scores imply a lower risk of default. After the score card model is built, risk manager should define a cut-off value for sum of the scores: under the cut-off value applicant would be refused.

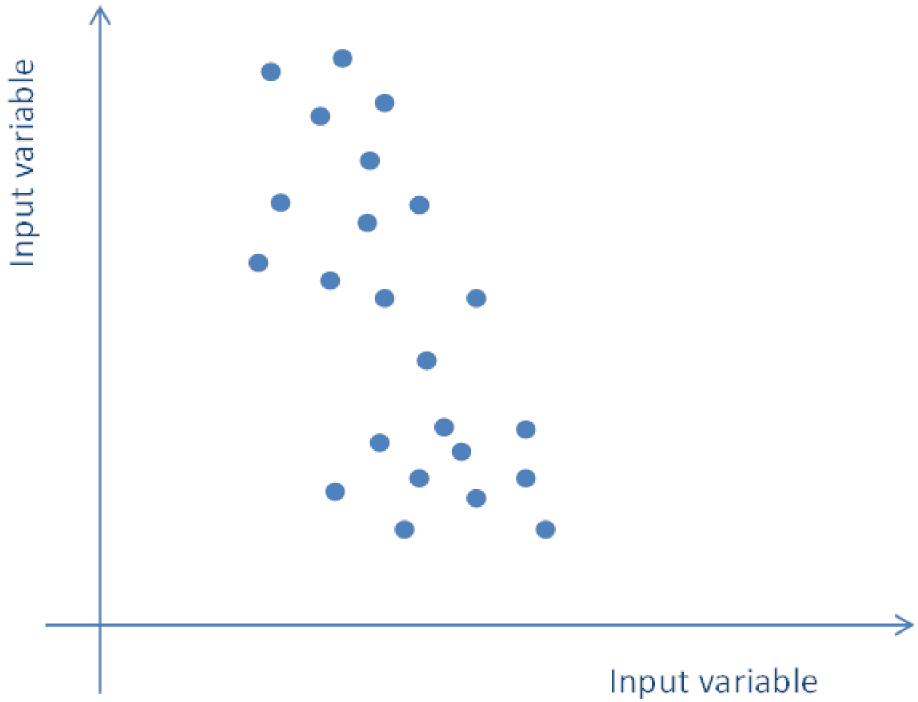
CREDIT SCORECARDS

The credit scorecard format is very popular and sucessful in the consumer credit world for the following reasons:

- **People up and down within the organization generally find it easy to understand and use. This format is the easiest to interpret, and it appeals to a broad range of risk managers and analysts who do not have advanced knowledge of statistics or data mining.**
- **Reasons for declines, low scores, or high scores can be explained to customers, auditors, regulators, senior management, and other staff, in simple business terms.**
- **The development process for these scorecards is not a black box, and is widely understood. As such, it can easily meet any regulatory requirement on method transparency.**
- **The scorecard is very easy to diagnose and monitor, using standard reports.**

CREDIT SCORING AND DATA MINING

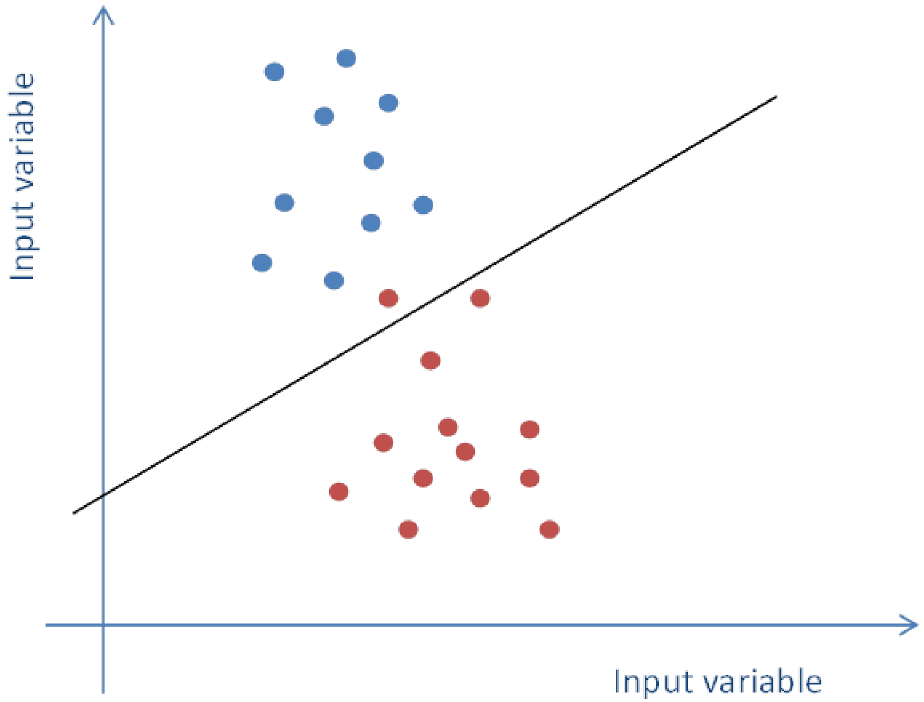
Type of data mining problems



CREDIT SCORING AND DATA MINING

Type of data mining problems

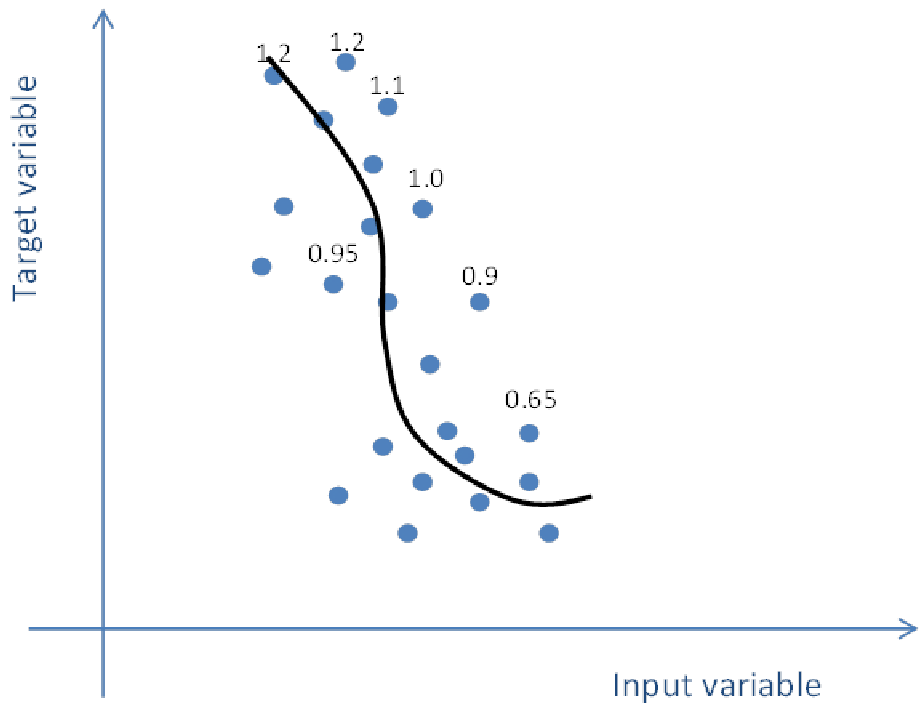
- **Classification**



CREDIT SCORING AND DATA MINING

Type of data mining problems

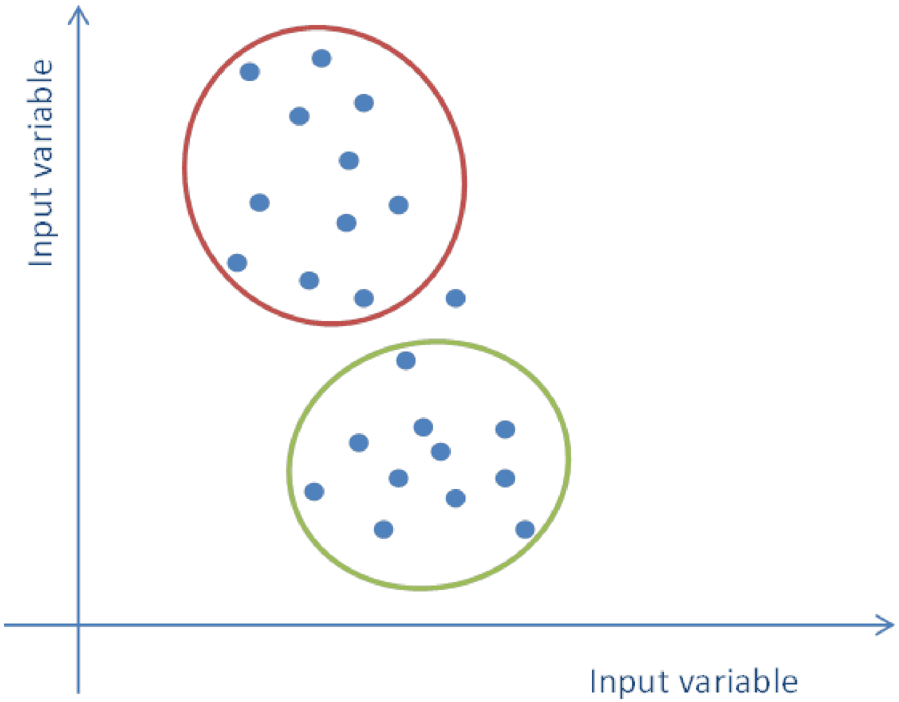
- **Regression**



CREDIT SCORING AND DATA MINING

Type of data mining problems

- **Segmentation**



CREDIT SCORING AND DATA MINING

Type of data mining problems

- **Classification**
  - Target: whether an applicant default
- **Regression**
  - Target: probability of default for an applicant

PERFORMANCE AND SAMPLING WINDOW

Scorecards are developed using the assumption that “future performance will reflect past performance.” Based on this assumption, the performance of previously opened accounts is analyzed in order to predict the performance of future accounts.

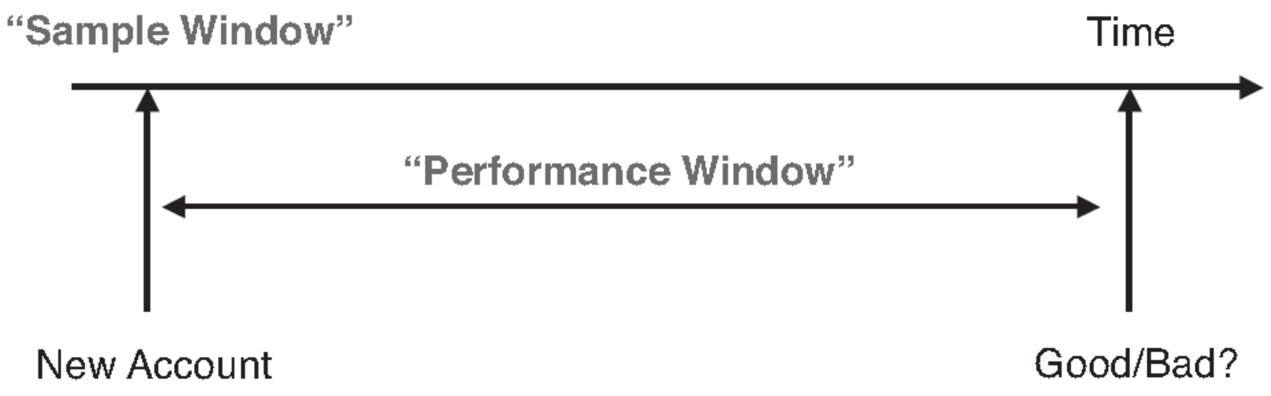
In order to perform this analysis, we need to gather data for accounts opened during a specific time frame, and then monitor their performance for another specific length of time to determine if they were good or bad.

Assume a new account is approved and granted credit at a particular time. At some point in time in the future, you need to determine if this account had been good or bad.

“Performance Window” is the time window where the performance of accounts opened during a particular time frame is monitored to assign class (target). “Sample Window” refers to the time frame from which known good and bad cases will be selected for the development sample.

PERFORMANCE AND SAMPLING WINDOW

“Performance Window” is the time window where the performance of accounts opened during a particular time frame is monitored to assign class (target).

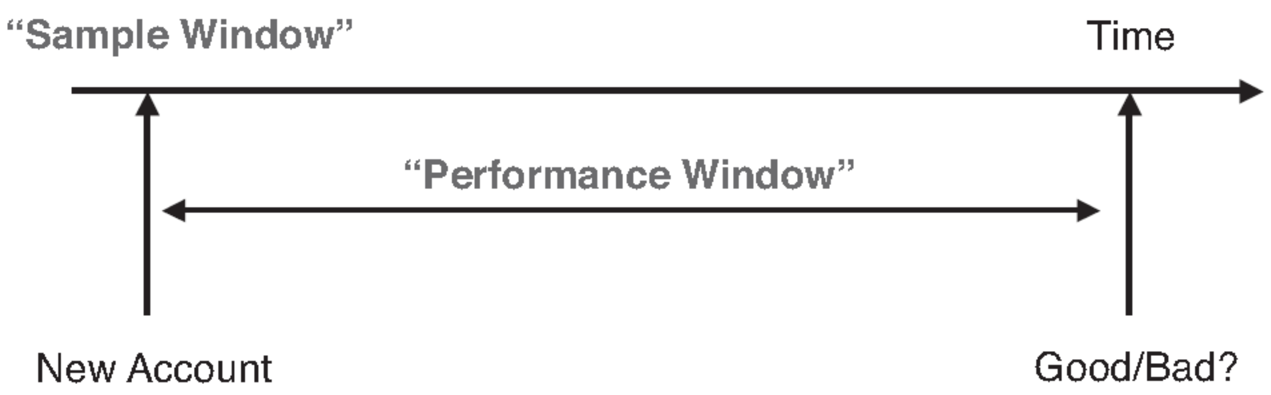


Source: Naeem Siddiqi, **Credit Risk Scorecards**, Developing and Implementing Intelligent Credit Scoring, Wiley, 2006

PERFORMANCE AND SAMPLING WINDOW

“Performance Window” is the time window where the performance of accounts opened during a particular time frame is monitored to assign class (target).

“Sample Window” refers to the time frame from which known good and bad cases will be selected for the development sample.

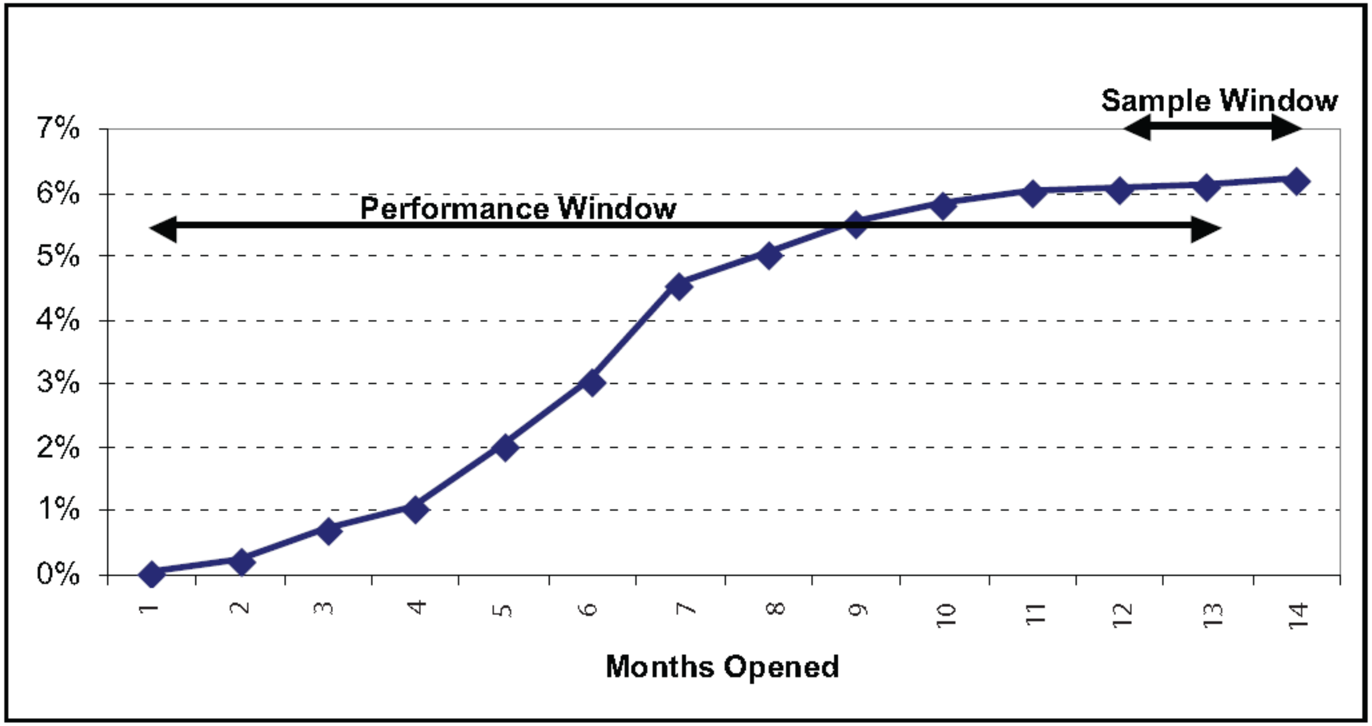


Source: Naeem Siddiqi, **Credit Risk Scorecards**, Developing and Implementing Intelligent Credit Scoring, Wiley, 2006

PERFORMANCE AND SAMPLING WINDOW

A simple way to establish performance and sample windows is to analyze payment or delinquency performance of the portfolio, and plot the development of defined “bad” cases over time.

PERFORMANCE AND SAMPLING WINDOW



This exhibit shows an example from a typical credit portfolio where the bad rate has been plotted for accounts opened in a 14-month period. It shows the bad rate developing rapidly in the first few months and then stabilizing as the account age nears 12 months.

Source: Naeem Siddiqi, **Credit Risk Scorecards**, Developing and Implementing Intelligent Credit Scoring, Wiley, 2006

PERFORMANCE AND SAMPLING WINDOW

The development sample is then chosen from a time period where the bad rate is deemed to be stable. In the preceding example, a good sample window would be anywhere between 12 and 14 months in the past, which would have a 13-month average performance window.

Selecting development samples from a mature portfolio is done to minimize the chances of misclassifying performance (i.e., all accounts have been given enough time to go bad), and to ensure that the “bad” definition resulting from an immature sample will not understate the final expected bad rates.

For example, if the development sample were chosen from accounts opened seven months ago, about 4.5% of the sample would be classified as bad. A mature sample for this portfolio should have a bad rate of about 6%. Therefore some accounts from this period that are bad would be erroneously labeled as good if the development sample were to be taken from that period.

The time taken for accounts to mature varies by product and by “bad” definition selected.

DATA UNDERSTANDING

Type of credit scoring

- **application scoring**
  - demographics (e.g., age, time at residence, time at job, postal code, income, profession)
- **behaviour scoring**
  - existing relationship (e.g., time at bank, number of products, payment performance, previous claims)

DATA PREPARATION

- **Good versus Bad**
  - Not necessarily clear how to define 2 classes
    - Bad = ever 3 or more payments in arrears
  - Spectrum of behaviour
    - Never any problems in payments
    - Occasional problems
    - Persistent problems
- **Developmental (training) and validation data sets**
  - Typical sample sizes ~ 10k or 100k per class
  - Should be representative of customers who will apply in the future
  - Need to be able to get the relevant variables sor this set of customers

Source: Padhraic Smyth, Data mining course, **Credit Scoring**, Department of Information and Computer Science

DATA PREPARATION

- **Initial characteristic analysis**
  - Grouping the attributes of characteristics
  - For categorical variables, attributes with similar WOE values are grouped together in accordance with business logic
  - For interval variable, groups are created that exhibit increasing or decreasing trends of WOE values
- **WOE, Weights of Evidence**
  - WOE measures the strength of the attributes of a characteristic in separating good and bad accounts
  - WOE is based on comparing the proportion of goods to bads at each attribute level
  - It is defined as ln(Distribution of good / Distribution of bad) for each attribute i of a characteristic

DATA PREPARATION

Age	Count	Total Distribution	Goods	Distribution Goods	Bads	Distribution Bads	WOE
Missing	1 000	2.50%	860	2.38%	140	3.65%	-0.43
18-22	4 000	10.00%	3 040	8.41%	960	25.00%	-1.09
23-26	6 000	15.00%	4 920	13.60%	1 080	38.13%	-0.73
27-29	9 000	22.50%	8 100	22.40%	900	23.44%	-0.05
30-35	10 000	25%	9 500	26.27%	500	13.02%	0.70
36-44	10 000	17.50%	6 800	18.81%	200	5.21%	1.28
44+	3 000	7.50%	2 940	8.13%	60	1.56%	1.65
Total	**40 000**	**100%**	**36 160**	**100%**	**3 840**	**100%**	

The following table shows a typical grouping for the characteristic variable Age. In the table, the columns Total Distribution, Distribution Goods, and Distribution Bads are the percentage of the total, good and bad cases.

Negative numbers imply that the particular attribute is isolating a higher proportion of baads than goods. Negative numbers are worse in the sense that applicants in an attribute group with a higher negative number are worse credit risks.

MODELING

- **Linear regression**
- **Logistic regression**

LINEAR REGRESSION

input measurement

$$\hat{y} = \hat{w}_0 + \hat{w}_1 \cdot x_1 + \hat{w}_2 \cdot x_2$$

prediction estimate

intercept estimate *parameter estimate*

In standard linear regression, a prediction estimate for the target variable is formed from a simple linear combination of the inputs. The intercept centers the range of predictions, and the remaining parameter estimates determine the trend strength (or slope) between each input and the target. The simple structure of the model forces changes in predicted values to occur in only a single direction (a vector in the space of inputs with elements equal to the parameter estimates).

Intercept and parameter estimates are chosen to minimize the squared error between the predicted and observed target values (least squares estimation). The prediction estimates can be viewed as a linear approximation to the expected (average) value of a target conditioned on observed input values.

Linear regressions are usually deployed for targets with an interval measurement scale.

LOGISTIC REGRESSION

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_0 + \hat{w}_1 \cdot x_1 + \hat{w}_2 \cdot x_2 \quad \textit{logit scores}$$

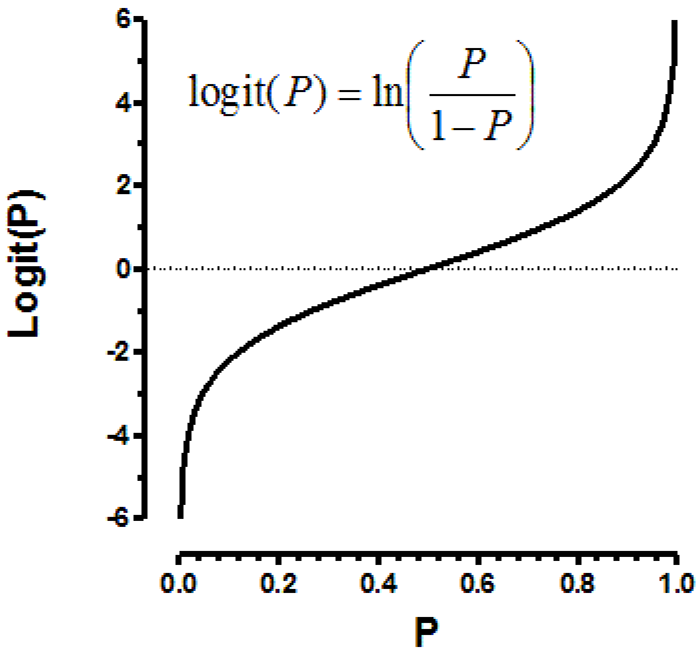
In logistic regression, the expected value of the target is transformed by a link function to restrict its value to the unit interval. In this way, model predictions can be viewed as primary outcome probabilities. A linear combination of the inputs generates a logit score, the log of the odds of primary outcome, in contrast to the linear regression's direct prediction of the target.

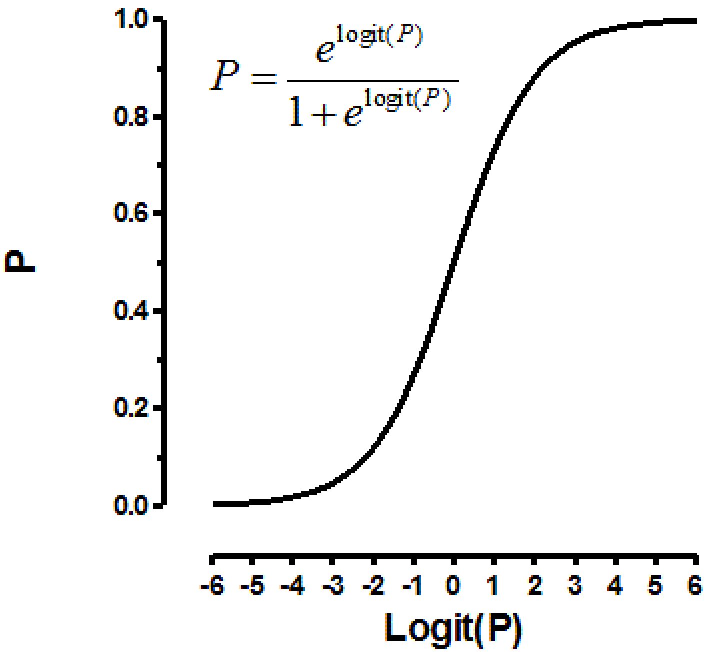
The presence of the link function complicates parameter estimation. Least squares estimation is abandoned in favor of maximum likelihood estimation. The likelihood function is the joint probability density of the data treated as a function of the parameters. The maximum likelihood estimates the values of the parameters that maximize the probability of obtaining the training sample.

LOGISTIC REGRESSION

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_0 + \hat{w}_1 \cdot x_1 + \hat{w}_2 \cdot x_2$$ logit scores

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





PROBLEM OF REJECT INFERENCE

- **Typically the population available for training consists only of past applicants who were accepted**
  - Application data is available for rejects, but no performance data
- **Question**
  - Is there a way to use the data from rejected applicants?
- **Approaches**
  - Reclassification: The worst cases of rejects are selected and reclassified as accepts, a bad status is then assigned
  - Re-weighting: The accounts (rejected and approved) are grouped by similar score, then behavior of the approved accounts in a score interval can be used to infer what the likely behavior of the corresponding rejects would be, had they been approved.

DEPLOYMENT

- **Monitoring and tracking**
  - Important to see how the scorecards model works in practice
  - Generating monthly/quarterly reports on scorecard performance, naturally there will be some delay in this → interpretable models
  - Analyzing in detail at performance on segments, by attributes
- **Time for a new model**
  - Population has changes significantly
  - New (outside) data is available
  - New modeling technique is available

DEPLOYMENT

Scaling the scorecard

In general, the relationship between odds and scores in represented by a linear function:

Score = Offset + Factor * ln(odds)

If the scorecard is developed using "odds at a certain score" and "points to double the odds" (pdo), Factor and Offset can be calculated using the simultaneous equations:

- Score = Offset + Factor * ln(odds)
- Score + pdo = Offset + Factor * ln(2 * odds)

Solving the equations for pdo, you get the following results:

- pdo = Factor * ln(2)

Therefore

- Factor = pdo / ln(2)
- Offset = Score - Factor * ln(odds)

DEPLOYMENT

If a scorecard were scaled where the developer wanted odds of 50:1 at 600 points and wanted the odds to double every 20 points (that is, pdo = 20), Factor and Offset would be as follows:

- Factor = 20 / ln(2) = 28.8539
- Offset = 600 - 28.8539 * ln(50) = 487.123

So, each score corresponding to each set of odds can be calculated as follows:

- Score = 487.123 + 28.8539 * ln(odds)

DEPLOYMENT

For each attribute, the WOE and the regression coefficient are multiplied together. Then, a fraction of the regression intercept is added. Multiply this by -1 and by Factor and finally add a fraction of Offset. An applicant's total score is then proportional to the logarithm of the predicted good/bad odds of that applicant.

The points allocated to attribute i of characteristic j are computed by this formula:

- -(WOE_i * beta_j + a / n) * Factor + Offset / n, where
  - WOE_i is the weight of evidence for attribute i of characteristic j
  - beta_j is the regression coefficient for characteristic j
  - a is the intercept term
  - n is the total number of characteristics
- Typically, points are rounded to the nearest integer.

THANK YOU FOR YOUR ATTENTION

ISTVÁN NAGY