Some Statistical Aspects of Credit Scoring

Gilbert Saporta

Conservatoire National des Arts et Métiers, Paris

saporta@cnam.fr

Outline

1. Introduction
2. Linear techniques for scorecard building
3. Categorical predictors
4. Direct scoring
5. Validation and model comparisons
6. Reject inference
7. Survival analysis
8. Conclusion

1.Introduction

Credit scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit.

Credit scoring is one the most successful applications of statistical modeling in finance and banking. Yet because credit scoring does not have the same glamour as the pricing of exotic financial derivatives or portfolio analysis, the literature on the subject is very limited.

Thomas & al. 2002

- **Basel 2**
  - Basel Committee on Banking Supervision from the Bank for International Settlements
  - « banks are expected to provide an estimate of the PD and LGD »
    - PD (probability of default)
    - LGD (loss given default)
  - Impulse on statistical analysis; massive recruitments
  - New Basel Capital Accord will regulate bank's lending from 2007

- Statistical framework of credit scoring:
  - response variable Y with 2 categories (« good » « bad »)
  - $X_1,\ldots,X_p$ predictors
- Belongs to :
  - classification
  - supervised learning
  - discrimination
  - pattern recognition

- Not only a classification problem
  - Risk assessment more than a binary decision
- Some specificities:
  - Reject inference
  - Long term loans

2. Linear techniques and scorecards

- Discriminant analysis
- Logistic regression
- Linear SVM
- Regularized regressions
  - PLS
  - ridge regression
- Others (GLM,linear programming,...)

2.1 Discriminant analysis

- **2.1.1 Fisher's linear discriminant function (1936)**
  - For numerical predictors:

$$\boldsymbol{\beta} = \mathbf{W}^{-1}(\mathbf{g}_1 - \mathbf{g}_2) = \mathbf{W}^{-1}\begin{pmatrix} \bar{x}_1^1 - \bar{x}_2^1 \\ . \\ \bar{x}_1^p - \bar{x}_2^p \end{pmatrix}$$

  - The « best » linear predictor which maximizes Student's T
  - Fisher's score:

$$S(\mathbf{x}) = (\mathbf{g}_1 - \mathbf{g}_2)'\mathbf{W}^{-1}\mathbf{x} - \frac{1}{2}(\mathbf{g}_1 - \mathbf{g}_2)'\mathbf{W}^{-1}(\mathbf{g}_1 + \mathbf{g}_2)$$

$$= \beta_1\mathbf{x}_1 + \ldots + \beta_p\mathbf{x}_p + \beta_0$$

2.1.2 a « non-correct » regression

- $\mathbf{y}$ with 2 values (-1;+1) or (0;1) or (a;b)
- a=n/n_1 b=-n/n_2

$$\boldsymbol{\beta} = \mathbf{V}^{-1}(\mathbf{g}_1 - \mathbf{g}_2)$$

$$D_p^2 = \frac{n(n-2)}{n_1 n_2} \frac{R^2}{1-R^2}$$

- D_p Mahalanobis distance between groups
- A lot of controversies!

2.1.4 Linear discriminant analysis and probabilistic assumptions

- LDA is optimal (Bayes rule) for normal predictors with equal covariance matrices
  - With priors : $S(\mathbf{x}) = (\mathbf{g}_1 - \mathbf{g}_2)'\mathbf{W}^{-1}\mathbf{x} - \ln(\frac{p_2}{p_1}) - \frac{1}{2}(\mathbf{g}_1 - \mathbf{g}_2)'\mathbf{W}^{-1}(\mathbf{g}_1 + \mathbf{g}_2)$
  - posterior probability $P(G_1 / \mathbf{x}) = \frac{\exp(S(\mathbf{x}))}{1 + \exp(S(\mathbf{x}))}$

logistic function

- May be applied even if these assumptions are not fulfilled

2.2 Logistic regression

$$\pi(\mathbf{x}) = P(Y = 1 / \mathbf{X} = \mathbf{x}) = \frac{e^{\beta_0 + \beta_1 \mathbf{x}_1 + \ldots + \beta_p \mathbf{x}_p}}{1 + e^{\beta_0 + \beta_1 \mathbf{x}_1 + \ldots + \beta_p \mathbf{x}_p}}$$

- Berkson (1944), Cox (1958): medical statistics, epidemiology
- Later in econometrics with Nobel prize McFadden (1973)
- Risk factors, not individual prediction

$$score = \beta_0 + \beta_1 \mathbf{x}_1 + \ldots + \beta_p \mathbf{x}_p$$

- Preferred by econometricians. The « industry standard »
  - Looks more « scientific » : prediction of probability, maximum likelihood estimation, standard errors, interpretation of coefficients as odds-ratios
  - Software procedure allows categorical predictors, without manipulating indicator variables

- But:
  - No solution in case of perfect separation
  - Conditional likelihood, asymptotics
  - Standard errors may be computed by bootstrap in LDA
  - In practice:

    « It is generally felt that logistic regression is a safer, more robust bet than the LDA model, relying on fewer assumptions . It is our experience that the models give very similar results , even when LDA is used in inappropriately, such as with qualitative variables. » Hastie and al.(2001)
  - A model should be choosen according to its performance, not to ideology!

2.3 Posterior probabilities and stratified sampling

- Probability estimation requires true priors
- Changing priors modifies only β_0 in LDA and in logistic regression:
  - Important for probabilities, not for score

2.4 Other methods derived from linear regression

Useful in case of multicollinearity. May be viewed as a modification of Fisher's LDA

- **2.3.1 Ridge regression**

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$$

$$\min \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \quad \text{with} \quad \|\boldsymbol{\beta}\|^2 < d^2$$

- Choice of k : cross-validation or test sample

2.3.2 PLS discriminant analysis

- Look for components explaining both Y and X's
- Tucker's criterion:

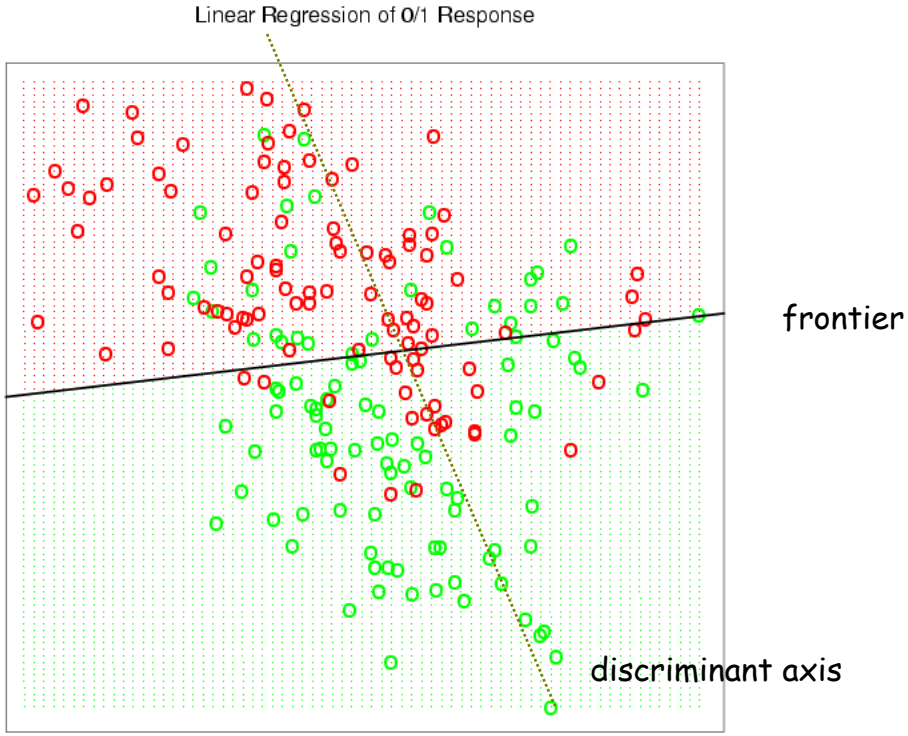
$$\max \quad (\mathrm{cov}(\mathbf{y}; \mathbf{Xw}))^2$$

$$(\mathrm{cov}(\mathbf{y}; \mathbf{Xw}))^2 = r^2(\mathbf{y}; \mathbf{Xw}).V(\mathbf{Xw}).V(\mathbf{y})$$

- Further components ; stopping rule: crossvalidation
- Only univariate regressions

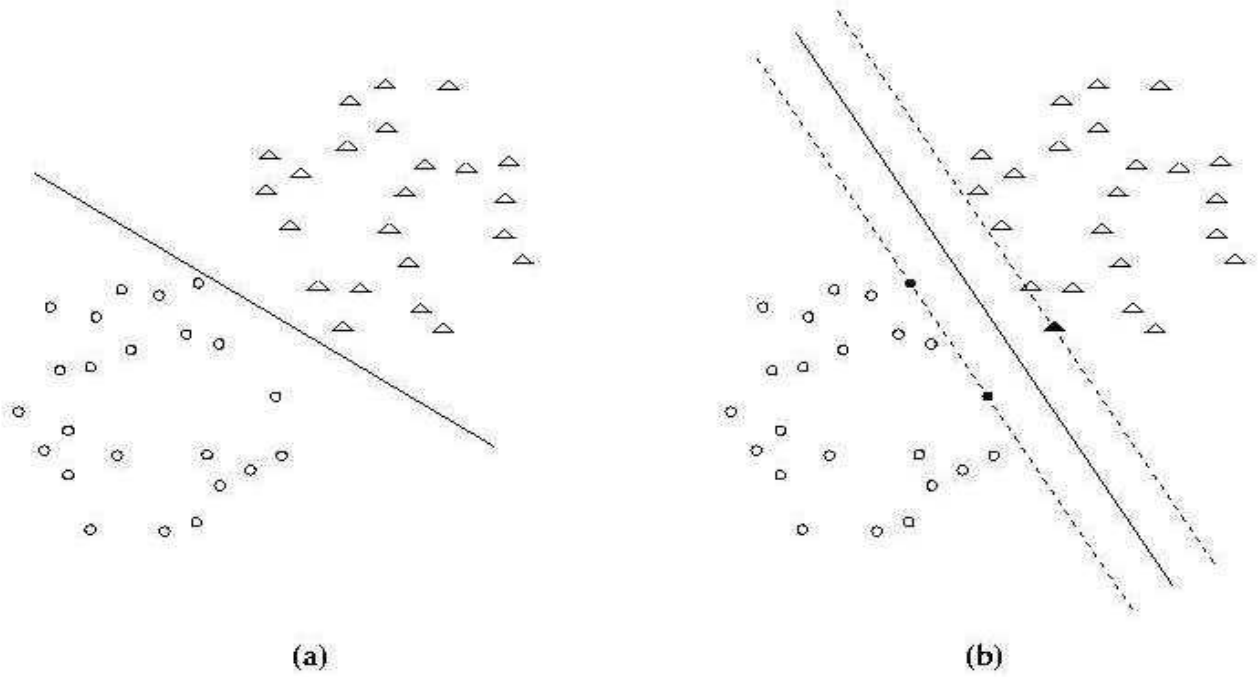
2.5 Linear Support Vector Machines (SVM)

- Linear score= linear frontier

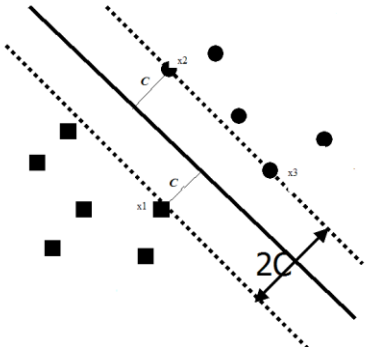


Adapted from Hastie & al. 2001

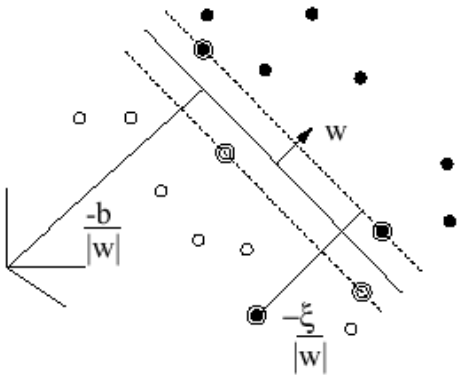
- Vapnik's optimal hyperplane maximizes the margin



- Margin: if perfect separation, distance of the closest point to the hyperplane



- Non separable case: slack variables



- Trade-off between error rate and margin
- Quadratic programming

$$y_i = \pm 1$$

$$\min \|\mathbf{w}\|$$

$$\text{subject to : } \begin{cases} y_i(\mathbf{x_i}'\mathbf{w} + b) \geq 1 - \xi_i \\ \sum \xi_i < \gamma \end{cases}$$

- **Classifier or score function**

$$f(\mathbf{x}) = \mathbf{w}'\mathbf{x} + b = \sum_{\alpha_i > 0} \alpha_i y_i \mathbf{x}_i' \mathbf{x} + b$$

- $f(\mathbf{x})$ depends only on support vectors
- is a linear combination of the variables
- Decision rule according to the sign of $f(\mathbf{x})$
- Less sensitive to outliers than LDA

3. Categorical predictors

- Frequent in consumers' credit, but not in publications..
  - Profession
  - Employment status
  - Marital status
  - Etc.

Categorisation of numerical predictors

- Age groups instead of age
- A loss of precision?
- A way towards non-linearity

$$S = \sum_{j=1}^{p} \varphi_j(X_j) \qquad \varphi_j \text{ step-functions}$$

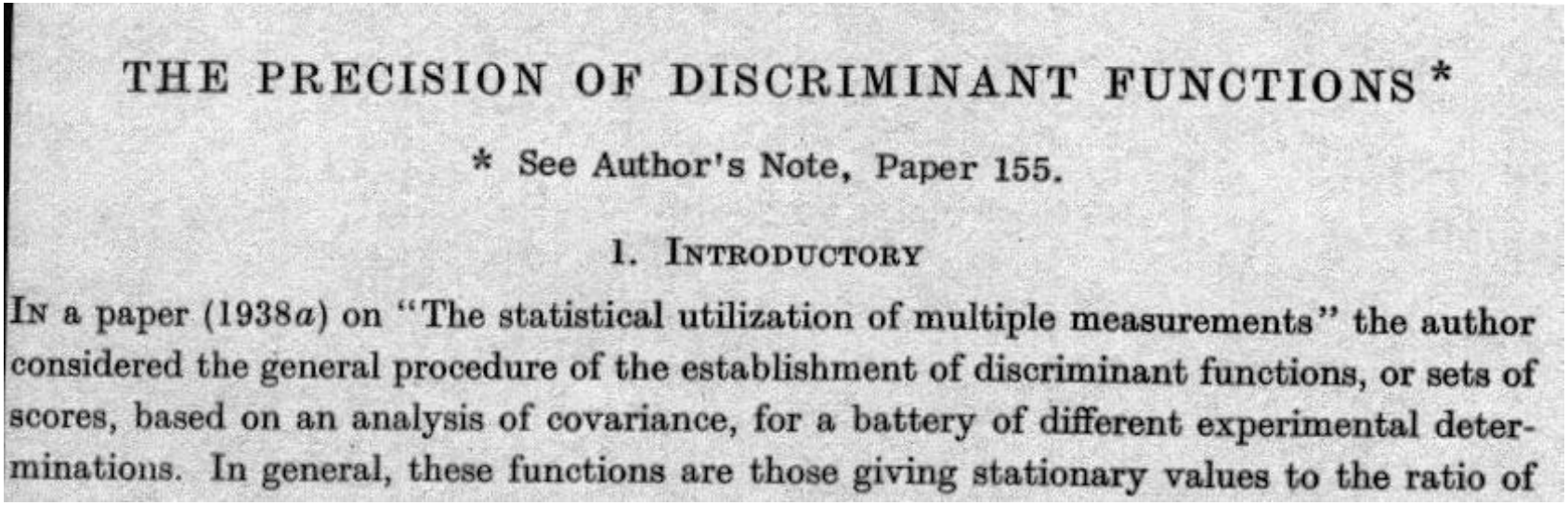
- Resistant to outliers: robustness
- Missing value category

Preprocessing

- Variable selection, discretisation, detection of interactions $X_j * X_k$ need a lot of time
- New automatic tools :
  - K2C, Khiops, Datalab..

3.1 LDA for categorical predictors: a bit of (pre)history

- Fisher (1940)
  - Only one predictor
  - Identical to correspondence analysis
  - « Scores » were introduced



THE PRECISION OF DISCRIMINANT FUNCTIONS *

* See Author's Note, Paper 155.

1. INTRODUCTORY

IN a paper (1938*a*) on "The statistical utilization of multiple measurements" the author considered the general procedure of the establishment of discriminant functions, or sets of scores, based on an analysis of covariance, for a battery of different experimental determinations. In general, these functions are those giving stationary values to the ratio of

For example, in a contingency table individuals are cross classified in two categories, such as eye colour and hair colour, as in the following example (Tocher's data for Caithness compiled by K. Maung of the Galton Laboratory).

Eye colour	Hair colour					
	Fair	Red	Medium	Dark	Black	Total
Blue	326	38	241	110	3	718
Light	688	116	584	188	4	1580
Medium	343	84	909	412	26	1774
Dark	98	48	403	681	85	1315
Total	1455	286	2137	1391	118	5387

Variation among the four eye colours may be regarded as due to variations in three variates defined conveniently in some such way as the following:

Eye colour	x_1	x_2	x_3
Blue	0	0	0
Light	1	0	0
Medium	0	1	0
Dark	0	0	1

We may then ask for what eye colour scores, i.e. for what linear function of x_1, x_2, x_3, are the five hair colour classes most distinct. The answer may be found in a variety of ways. For example, by starting with arbitrarily chosen scores for eye colour, determining from these average scores for hair colour, and using these latter to find new scores for eye colour.

Apart from a contraction of scale by a factor R^2 for each completed cycle, this form tends to a limit, and yields scores such as the following:

Eye colour	x	Hair colour	y
Light	−0·9873	Fair	−1·2187
Blue	−0·8968	Red	−0·5226
Medium	0·0753	Medium	−0·0941
Dark	1·5743	Dark	1·3189
		Black	2·4518

The particular values given above have been standardized so as to have mean values zero, and mean square deviations unity. In the sample from which they are derived each score has a linear regression on the other, the regression coefficient being 0·44627; this is, of course, equal to the correlation coefficient between the two scores regarded as variates. Hotelling has called pairs of functions of this kind canonical components. It may be noticed that no assumption is introducéd as to the order of the classes of each category. In Tocher's schedule Light eyes come between Blue and Medium, but the discriminant function puts Blue between Medium and Light, though near the latter.

3.2 General case: p predictors

- Optimal scaling (quantification) approach:
  - Allot partial scores to predictor categories in order to maximize Mahalanobis distance in $\mathbb{R}^p$
- A discriminant analysis where categorical variables are replaced by indicator variables

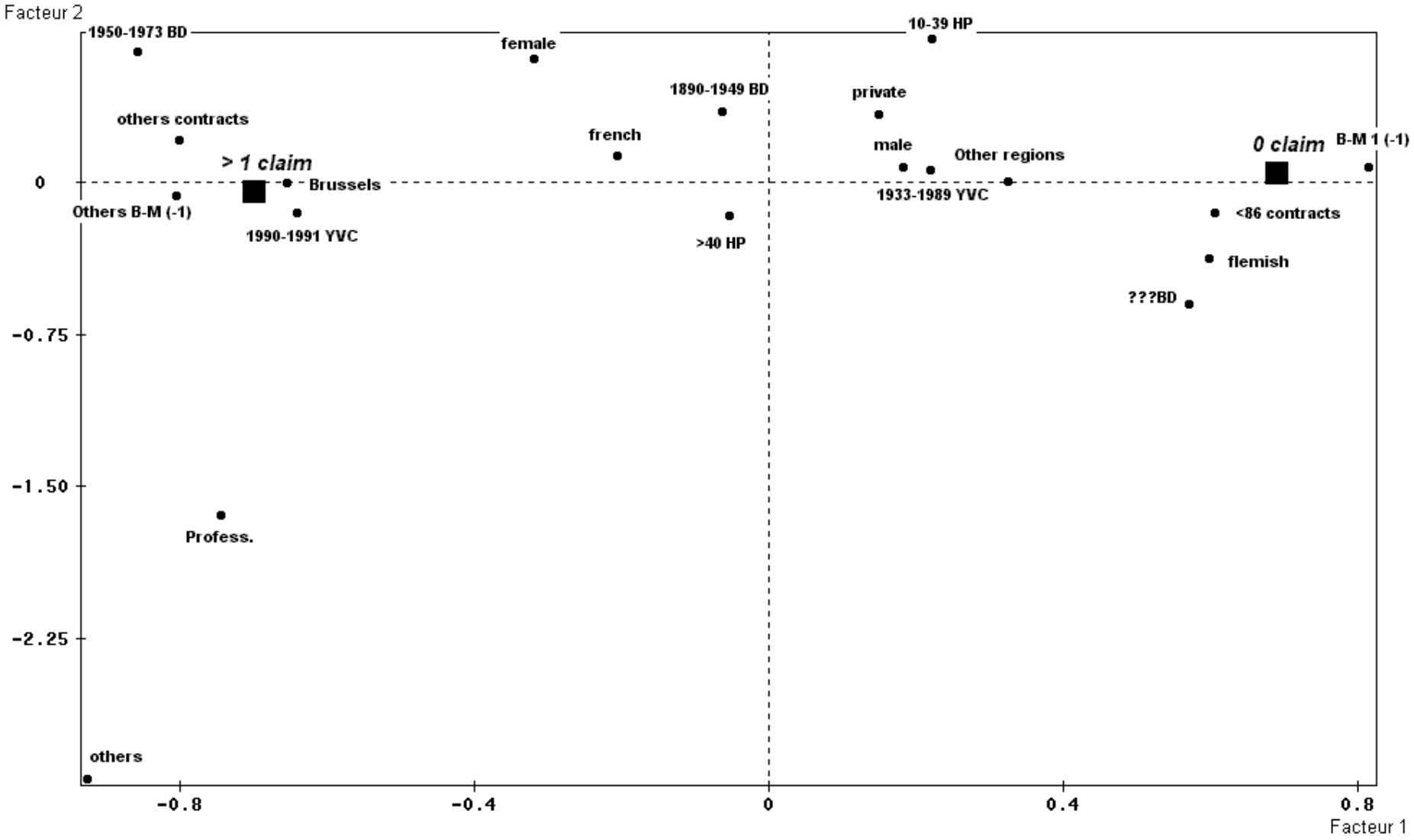
$$X = \left(\begin{array}{ccc|cc|} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{array}\right)$$

- X not of full rank: rank(X)=Σm_i-p
  - Classical solution : discard one indicator variable for each predictor
  - Disqual (Saporta, 1975):
    - LDA performed on a selection of components of Multiple Correspondence Analysis of X. Similar to Principal Components Regression
    - Components selected in an expert way according to 2 criteria: inertia and correlation with the response

An insurance example (SPAD data set)

- 1106 belgian automobile insurance contracts :
- 2 groups: « 1 good », « 2 bad »
- 9 predictors: 20 categories
  - Use type(2), gender(3), language (2), agegroup (3), region (2), bonus-malus (2), horsepower (2), duration (2), age of vehicle (2)

Principal plane MCA



Fisher's LDA

FACTORS	CORRELATIONS	LOADINGS
1 F 1	0.719	6.9064
2 F 2	0.055	0.7149
3 F 3	-0.078	-0.8211
4 F 4	-0.030	-0.4615
5 F 5	0.083	1.2581
6 F 6	0.064	1.0274
7 F 7	-0.001	0.2169
8 F 8	0.090	1.3133
9 F 9	-0.074	-1.1383
10 F 10	-0.150	-3.3193
11 F 11	-0.056	-1.4830
INTERCEPT		0.093575

R2 = 0.57923 F = 91.35686
D2 = 5.49176 T2 = 1018.69159

Score= 6.90 F1 - 0.82 F3 + 1.25 F5 + 1.31 F8 - 1.13 F9 - 3.31 F10

- **3.3 Transforming scores**
  - Standardisation between 0 and 1000 is often convenient
  - Linear transformation of score implies the same transformation for the cut-off

Scorecard

CATEGORIES	COEFFICIENTS DISCRIMINANT FUNCTION	TRANSFORMED COEFFICIENTS (SCORE)
2 . Use type		
USE1 - Profess.	-4.577	0.00
USE2 - private	0.919	53.93
4 . Gender		
MALE - male	0.220	24.10
FEMA - female	-0.065	21.30
OTHE - companies	-2.236	0.00
5 . Language		
FREN - French	-0.955	0.00
FLEM - flemish	2.789	36.73
24 . Birth date		
BD1 - 1890-1949 BD	0.285	116.78
BD2 - 1950-1973 BD	-11.616	0.00
BD? - ???BD	7.064	183.30
25 . Region		
REG1 - Brussels	-6.785	0.00
REG2 - Other regions	3.369	99.64
26 . Level of bonus-malus		
BM01 - B-M 1 (-1)	17.522	341.41
BM02 - Others B-M (-1)	-17.271	0.00
27 . Duration of contract		
C<86 - <86 contracts	2.209	50.27
C>87 - others contracts	-2.913	0.00
28 . Horsepower		
HP1 - 10-39 HP	6.211	75.83
HP2 - >40 HP	-1.516	0.00
29 . year of vehicle construction		
YVC1 - 1933-1989 YVC	3.515	134.80
YVC2 - 1990-1991 YVC	-10.222	0.00

3.4 PLS and barycentric discrimination

- First PLS component: univariate regression onto all indicator variables
- Getting the first PLS component comes down to p PLS regressions performed separately
- Each PLS of Y against indicators of Xj is equivalent to OLS regression (Y should be standardised, not X, and no intercept)

- PLS with one component is equivalent to CA of the concatenation of the contingency tables crossing Y with the X_j

		good	bad
1	cusag1	29	96
2	cusag2	344	272
3	sexe1	288	253
4	sexe2	76	78
5	sexe3	9	37
6	clang1	250	295
7	clang2	123	73
8	age3m1	118	99
9	age3m2	40	163
10	age3m3	215	106
11	cpost2m1	75	172
12	cpost2m2	298	196
13	bm2m_11	298	59
14	bm2m_12	75	309
15	puis2m1	91	47
16	puis2m2	282	321
17	dpoli2m1	277	137
18	dpoli2m2	96	231

Previous technique known as barycentric discrimination :

- The score of a unit: the sum of the *p* conditional probabilities of being a member of group 2 for each categories.

- Barycentric discrimination, similar to the "naive Bayes classifier" : multiplicative score equal to the product of the conditional probabilities.

- Barycentric discrimination is equivalent to *Disqual* only if the predictors are pairwise independent.

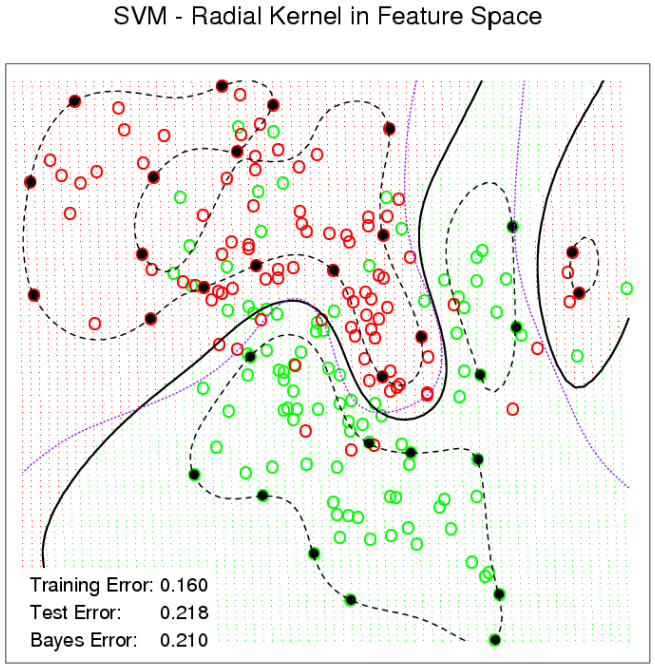
4. Direct scoring

- Non linear methods give directly a score, or a probability of being « good » or « bad » for each unit
- Remark: a probability is a score between 0 and 1. Just multiply it by 1000…

- Density estimation : posterior probabilities
- Neural networks: posterior probabilities
- Non linear SVM: score function
- k-nearest neighbours

Non-linear frontiers

$$f(\mathbf{x}) = \sum_{i \in \text{supports}} \alpha_i y_i K(\mathbf{x}_i; \mathbf{x}) + b = 0$$



Hastie & al. 2001

Black-boxes:

- Lack of interpretability
  - Cannot be used for consumer's credit: legal obligations to explain rejection
- Should be adapted to categorical predictors
  - Principal components from MCA, or pre-scores

5 Validation and comparison

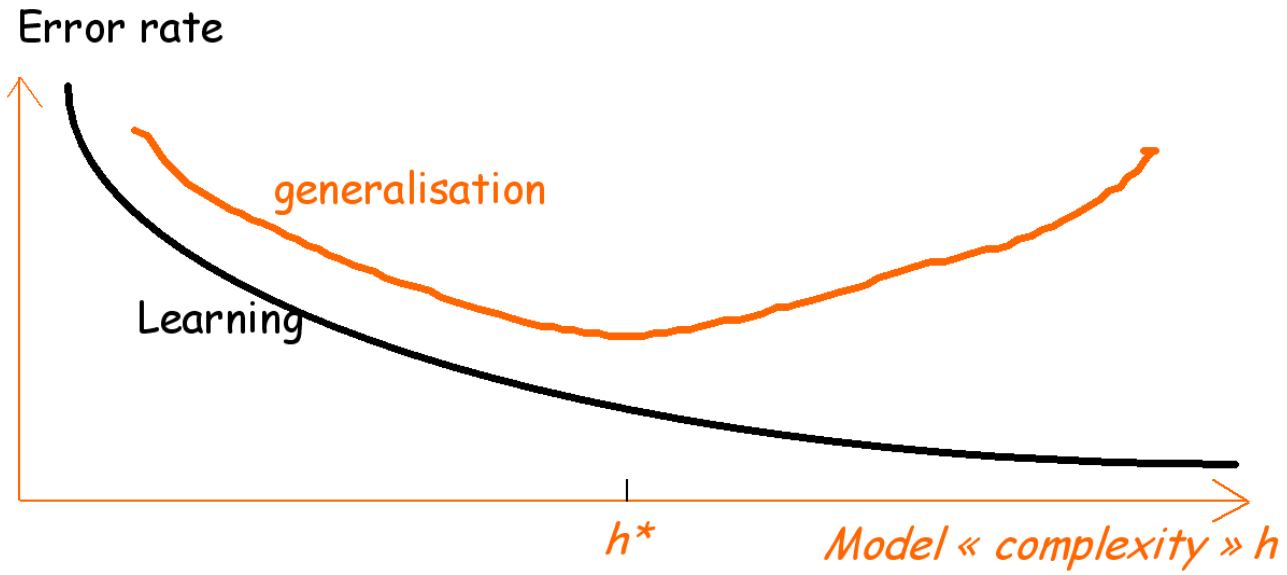
- **5.1 Statistical criteria: are they relevant**?
  - D^2, log-likelihood measure the adequacy of a model to learning data
  - Not related to predictive inference but easy to optimize
  - Penalized likelihood (AIC,BIC): too restrictive
    - Difficult to apply : Neural nets, ridge regression?

- Credit scoring is not science but business

- No need for the « true » model but for efficient rules

5.2 Misclassification rate and Statistical Learning Theory

- Error rate and model complexity



Empirical risk and VC dimension

- Empirical risk= learning error R_{emp}
- generalisation error= R
- Both are expected values

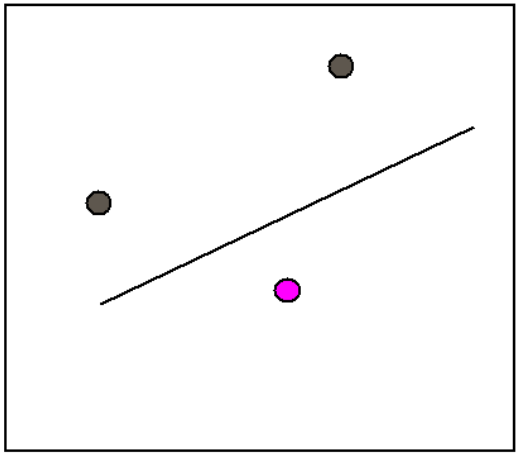
- Vapnik's inequality
  - With probability 1-q

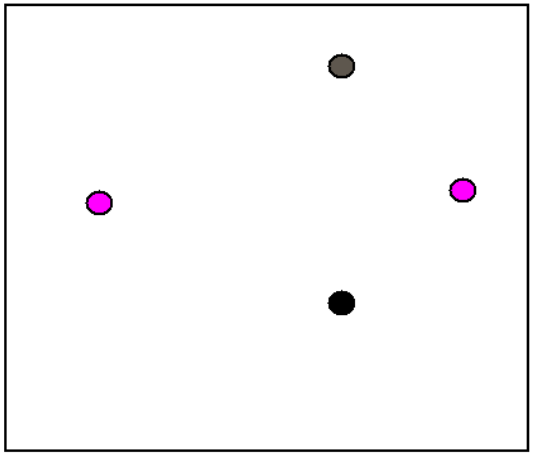
$$R < R_{\text{emp}} + \sqrt{\frac{h\left(\ln\left(2n/h\right)+1\right)-\ln q/4}{n}}$$

Confidence interval

VC dimension h

- A measure of model complexity
- Related to the splitting capacity
- h must be finite (consistent learning)

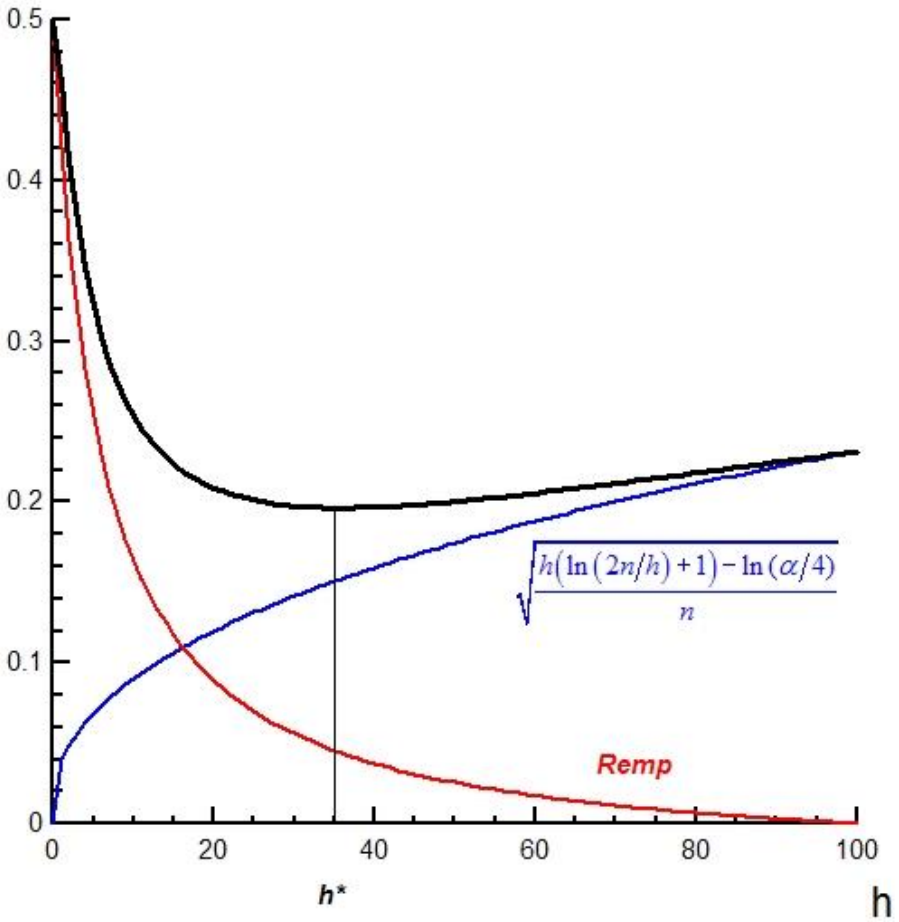




h= maximum number of points always perfectly classified by a model

h=3 for linear frontiers in 2 dimensions

Looking for optimal *h*



Controlling h

- h/n should be small: as n increases, one may choose more complex models
- h decreases with:
  - Dimension reduction (see Disqual)
  - Large margin in SVM
  - Large k in Ridge Regression
- Exact h difficult to find

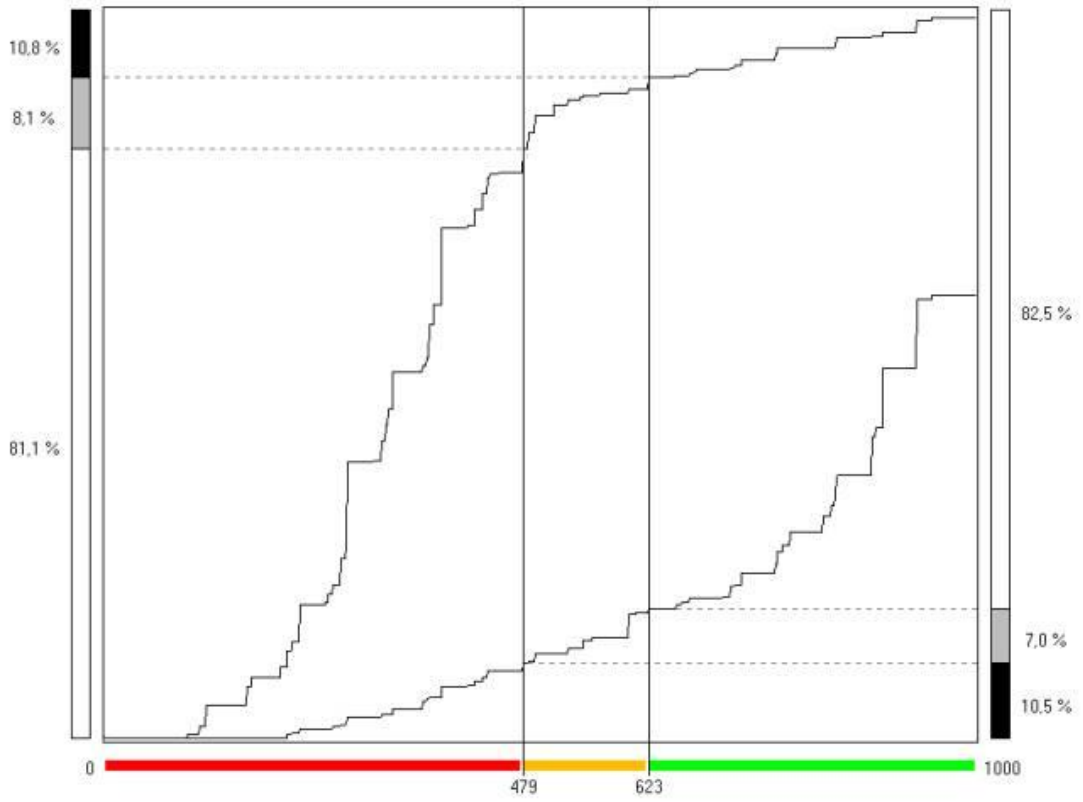
5.3 The 3 samples

- Learning sample: estimating models parameters
- Test sample: selecting the best model
- Validation sample: estimating performance for new data

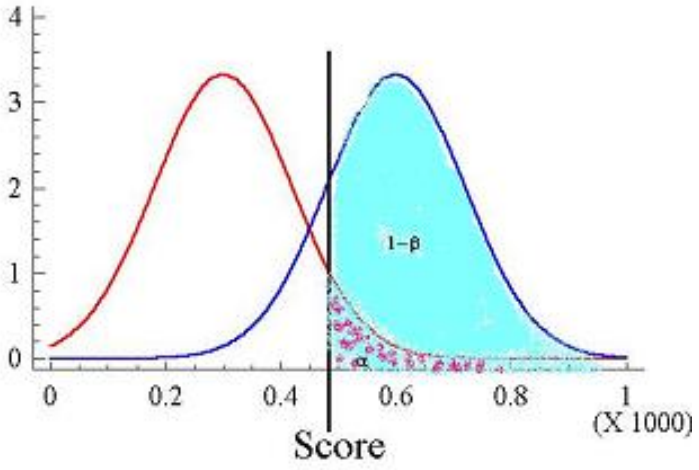
5.4 ROC, lift and related measures

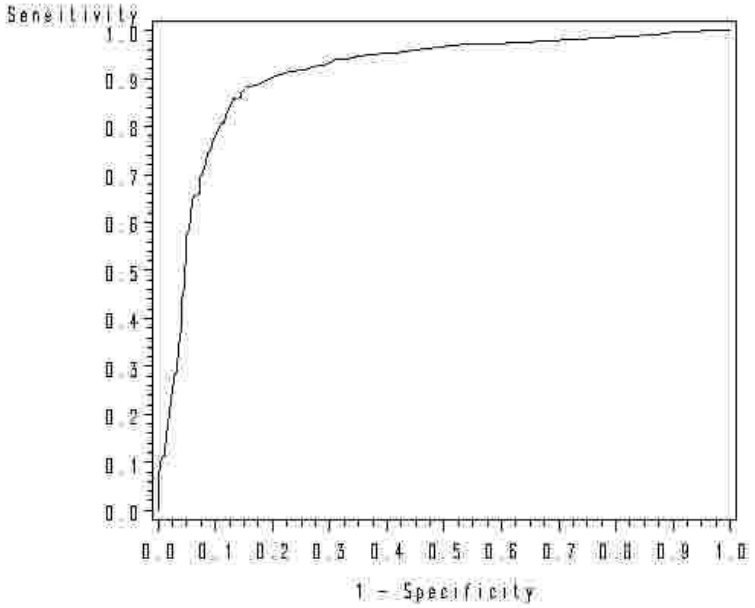
- Misclassification rate: often not the right measure
  - Needs a specific cut-off
    - Posterior probability >0.5
    - Minimizing a cost. But costs often unknown
- Performance of the score function when the cut-off varies

Traffic light zones



When cutoff moves : ROC analysis





% of true « goods » (1-β)
against % of false « goods » (α)

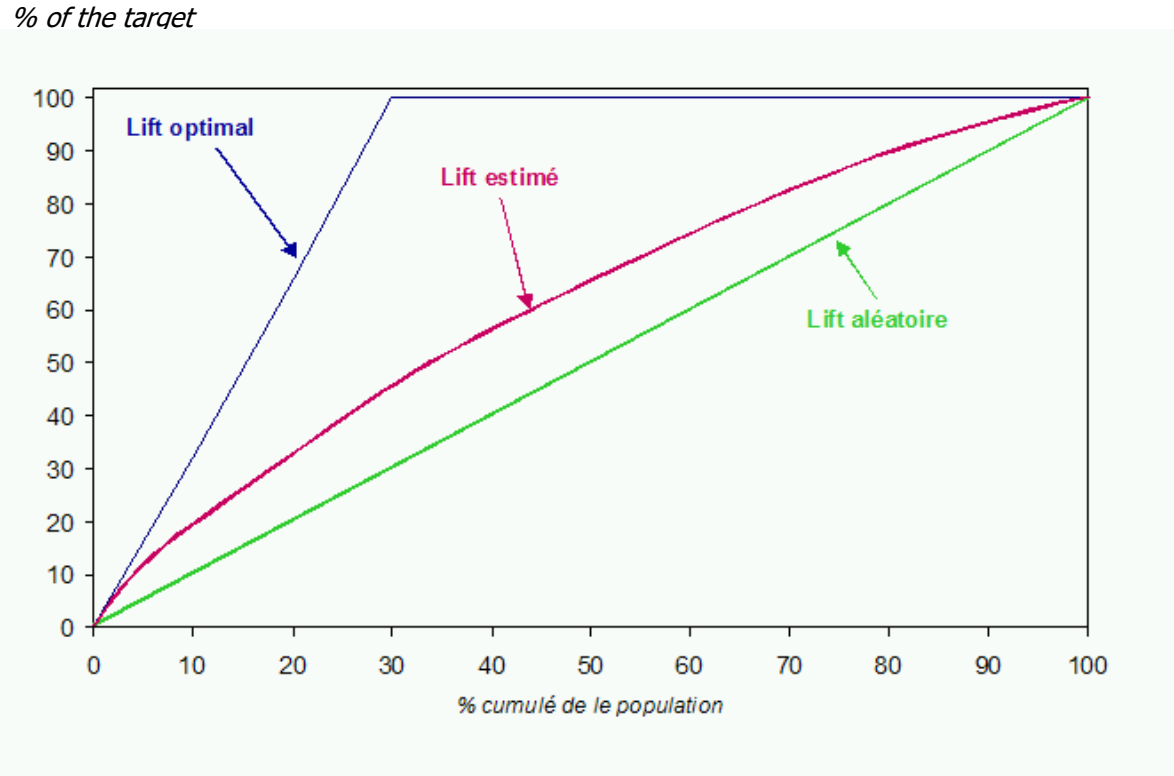
- ROC curve is invariant under any monotonous transformation
- Area under Roc curve is a measure of performance allowing model comparisons
- $$AUC = \int_{s=+\infty}^{s=-\infty} (1-\beta(s))d\alpha(s) = P(X_1 > X_2)$$

  If one takes at random one obs from G_1 and one from G_2
  - AUC estimated by the proportion of concordant pairs

$$c = n_c / n_1 n_2$$

- n_c identical to Wilcoxon-Mann-Whitney statistic

Lift chart


% of the target
100
90
80
70
60
50
40
30
20
10
0
Lift optimal
Lift estimé
Lift aléatoire
0
10
20
30
40
50
60
70
80
90
100
% cumulé de le population

Area under lift

- Proportion of units with score>s

$$p_1(1-\beta)+(1-p_1)\alpha$$

- Area:

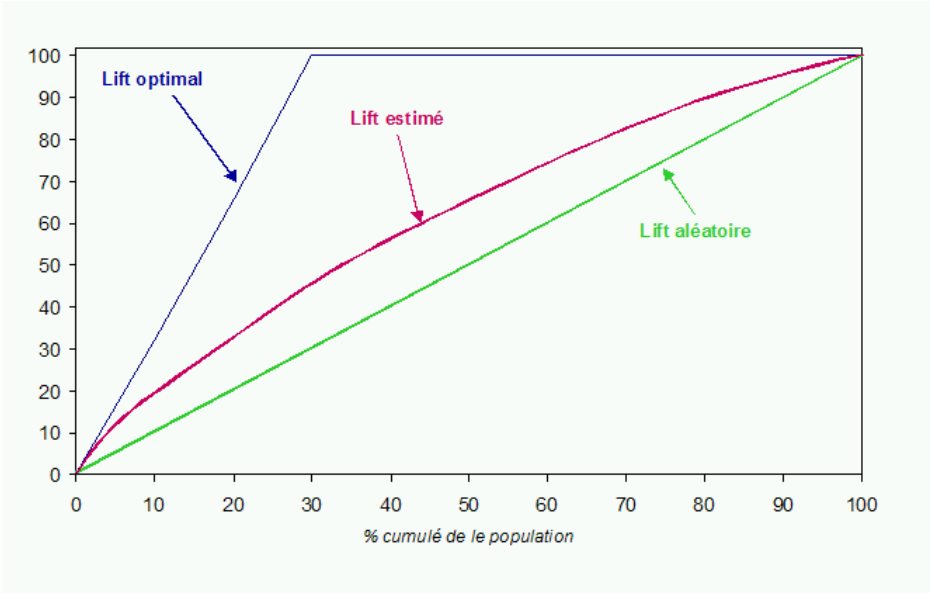
$$L=\int(1-\beta)d\left\{p_1(1-\beta)+(1-p_1)\alpha\right\}=$$

$$\left[p_1\int(1-\beta)d(1-\beta)\right]+\left[(1-p_1)\int(1-\beta)d\alpha\right]$$

$$=\frac{p_1}{2}+(1-p_1)AUC$$

Ki Coefficient (Kxen)

- Ki=(area between estimated lift and random lift) / (area between ideal lift and random lift)

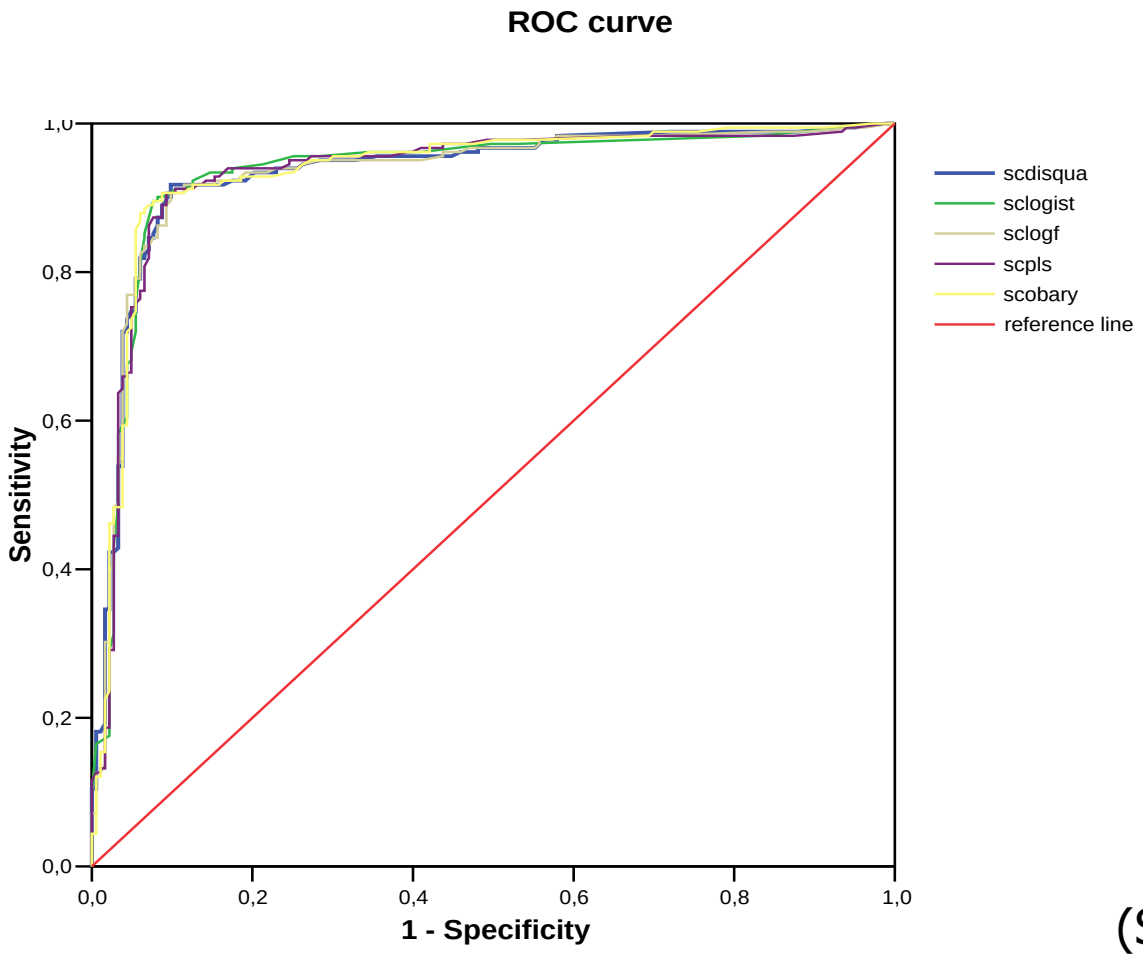


$$Ki = \frac{L - \frac{1}{2}}{\frac{1 - p_1}{2}} = \frac{p_1 + 2(1 - p_1)AUC - 1}{1 - p_1} = 2AUC - 1$$

Ki=Somers' *D* or Accuracy Ratio *AR*

- Optimizing AUC or Ki are equivalent.
- But do not depend on costs: assume that the two kinds of errors have the same importance...
- Comparisons should be done on validation samples

5.5 Experimental results



AUC

Score	AUC
scdisqua	.934
sclogist	.933
sclogf	.932
scpls	.933
scobary	.935

(Saporta, Niang,2003)

Baesens (2003) 17 techniques on 8 data sets

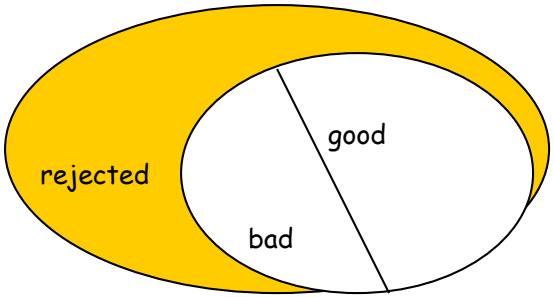
Table 4 Test set AUC on credit scoring data sets

Technique	*Bene1*	*Bene2*	*Germ*	*Austr*	*UK1*	*UK2*	*UK3*	*UK4*	*AR*
LDA	**77.1**	*77.1*	**78.4**	**92.8**	*64.1*	*73.6*	**74.4**	**72.3**	**5.38**
QDA	*73.4*	*72.4*	*71.8*	**91.5**	*63.3*	*72.1*	*68.1*	*68.3*	*10.8*
LOG	**77.0**	*78.0*	**77.7**	**93.2**	*63.9*	*73.0*	**74.6**	**72.7**	**4.38**
LP	*76.1*	*77.5*	**76.3**	**92.6**	*56.4*	*62.3*	*62.0*	*62.2*	*11.9*
RBF LS-SVM	**77.6**	77.8	**77.4**	**93.2**	65.0	74.7	**72.9**	**73.1**	**3.38**
Lin LS-SVM	**76.9**	*77.1*	**78.4**	**92.9**	*64.4*	*73.7*	**73.8**	**72.5**	**5.50**
RBF SVM	*76.7*	*77.1*	**77.2**	**92.6**	*59.3*	*65.4*	*67.3*	*63.4*	9.13
Lin SVM	*75.9*	*77.5*	**76.6**	**93.6**	*56.4*	*63.9*	*62.9*	*62.9*	10.1
NN	**76.9**	**79.1**	**78.7**	**91.7**	**66.4**	**75.8**	**74.6**	**72.9**	**3.25**
NB	**76.5**	*70.6*	**77.2**	**93.1**	**65.8**	*73.7*	*66.9*	67.9	7.88
TAN	75.5	**78.2**	**78.3**	**93.4**	**66.8**	74.5	*64.0*	*66.6*	**5.63**
C4.5	*72.2*	*71.1*	**74.7**	91.6	*56.1*	*65.7*	*50.0*	*49.9*	*14.7*
C4.5rules	*71.6*	*74.2*	*62.0*	*85.3*	*61.7*	*70.4*	*60.3*	68.4	*13.0*
C4.5dis	*73.0*	*73.2*	**74.6**	**93.1**	*50.0*	*50.0*	*50.4*	*49.9*	*13.7*
C4.5rules dis	*73.0*	*71.5*	*64.4*	93.1	65.2	*71.5*	*66.7*	*64.9*	10.8
KNN10	*71.7*	*69.6*	*70.2*	91.4	*58.9*	*65.4*	*63.0*	67.0	*14.1*
KNN100	*74.9*	*71.5*	**76.1**	**93.0**	*62.8*	*69.9*	*70.0*	**70.4**	9.5

« However, it has to be noted that simple, linear classifiers such as LDA and LOG also gave very good performances, which clearly indicate that most credit scoring data sets are only weakly non-linear ».

6. Reject inference

- Analysis done on approved loans :Biased sample



- Empirical techniques:
  - Define rejected as bad
  - Extrapolation
  - Augmentation or reweighting
- Probabilistic models
  - Missing data estimation (EM)
  - Bivariate probit
  - Tobit

- If reject variables X_1 are a subset of scorecard X variables: an unbiased model can be built in some cases
  - If $X_1 \not\subset X$: no unbiased model is possible

- Müller & al, 2005: Non parametric bounds for misclassifications rate and AUC

Few published evaluations

The scope for improved predictive performance by any form of reject inference is modest . Reject inference in the form of re-weighting applicants within a training sample of accepted cases and adopting a cut-off point based on those accepted cases appears to perform no better than unweighted estimation. In fact where the rejection rate is high, results appear to be quite noticeably worse. Reject inference in the form of extrapolation appears to be both useless and harmless. (Crook, Banasik 2002)

Many methods have been used for tackling this problem. Most of those used in practice are demonstrably ineffective. The best strategies are to build a formal sample selection model to supplement the classification model, and to obtain data about the rejected applicants. This can come from a small sample of people who would normally be rejected (this is done in mail order) or from other sources, such as other supplier (Hand 2005)

7. New frontier: survival analysis

- Not « if » but « when » default occurs
  - Integrates censored data: may solve the problem of incomplete data for long term loans (definition of default)
  - Useful for lifetime value and LGD computations (Basel II)
  - Stepanova, Thomas, 2001: Cox proportional hazard model

Conclusions and perspectives

- Credit scoring: an attractive and active field for statisticians
- Still place for further research (strong interest from firms)
- LDA and LR perform well, compared to new methods
- But: the precision of refined models could be an illusion
  - If data quality is not present
  - If there are changes in population

References

- Baesens: « Developing intelligent systems for credit scoring using machine learning techniques » Ph.D, Leuven, 2003
- Bardos: « Analyse discriminante », Dunod, 2001
- Hastie, Tibshirani, Friedman : « The Elements of Statistical Learning», Springer-Verlag, 2001
- Mays ed. « Handbook of credit scoring » Glenlake, 2001
- Thomas, Edelman,Crook: « Credit scoring and its applications », SIAM, 2002
- Credit Research Center http://www.crc.man.ed.ac.uk
- http://www.defaultrisk.com/
- Basel Committee publications: http://www.bis.org/bcbs/publ.htm