Variable Selection and Optimization in Default Prediction

Dedy Dwi Prastyo
Wolfgang Karl Härdle

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

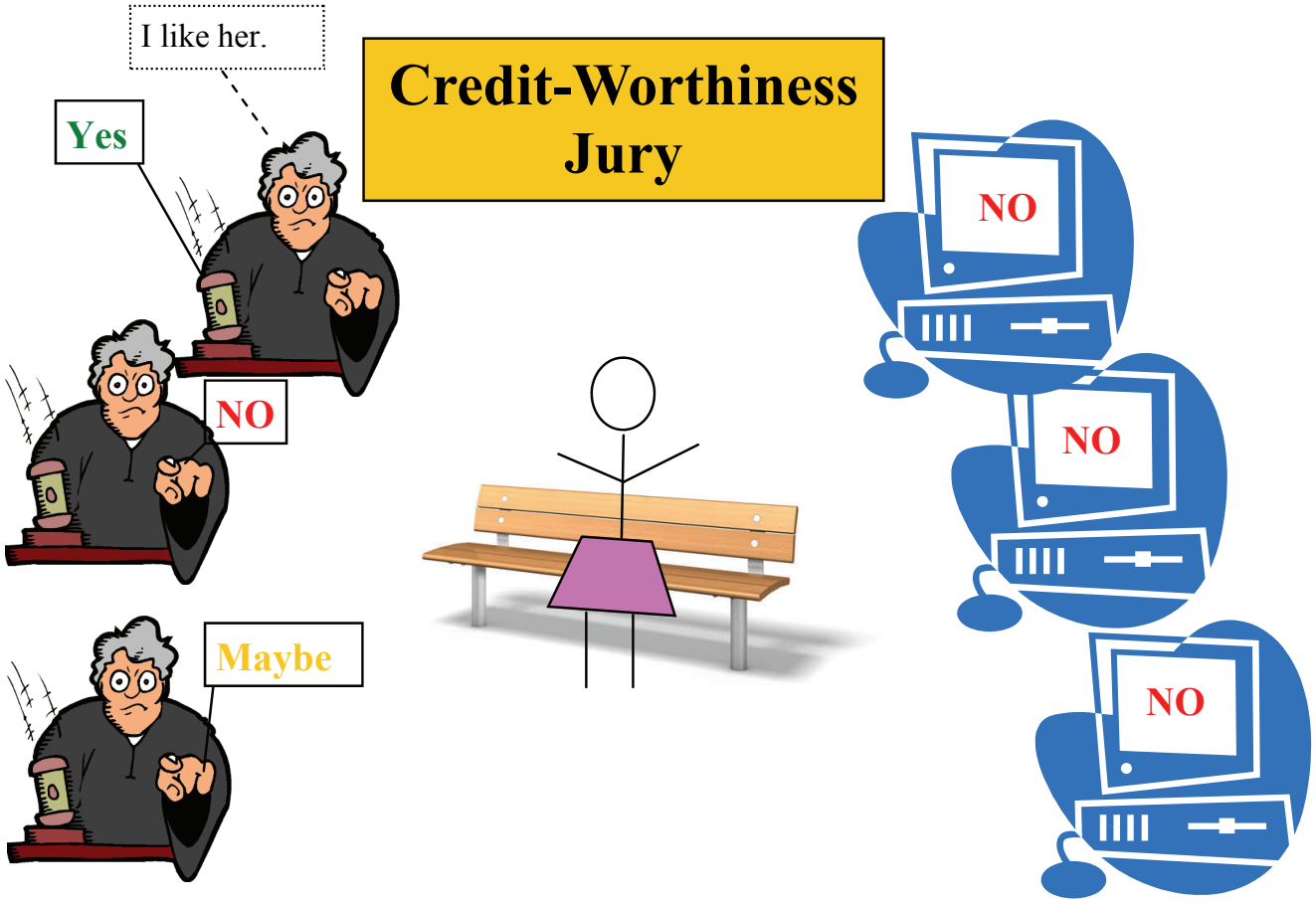


Figure 1: Who is more precise?

Credit rating

- **Score** (S)

  Quantitative indicator for customers w.r.t. their individual default risk

- **Probability of Default** (PD)

  One-to-one mapping of the score, $S \to \mathrm{PD}(S)$

- **Rating**

  Classification of customers (private, corporate, sovereign) into groups of equivalent default risk

Default prediction

- **Time-series data (market data)**
  - Merton approach (stock price as estimate for the market value), S = *distance to default*
- **Cross-sectional data (i.e. balance sheet)**
  - Discriminant analysis
  - Categorical regression (logit, probit)
  - Support Vector Machines (SVM)

Research questions

- What are the variables (i.e. accounting) significantly contribute to default ?
- How to optimize the default prediction (classification) ?

Outline

1. Introduction ✓
2. Variable Selection – Regularized GLM, elastic net
3. Evolutionary optimization
4. Genetic Algorithm SVM

Some problems

- Number of predictors is greater than number of observation, $p \gg n$
- There are correlated variables
- Sparsity (elements of predictor matrix $X \approx 0$)
- VLDS (very large data set)

Model with convex penalty

Apply fast algorithm to estimate

- Model
  - Linear regression
  - Two-class logistic regression
- Penalties
  - Lasso (ℓ_1)
  - Ridge regression (ℓ_2)
  - Elastic net (mixture of ℓ_1 and ℓ_2)

Linear regression

Suppose $Y \in \mathbb{R}$ and $X \in \mathbb{R}^p$

$$\mathrm{E}\left(Y|X = x\right) = \beta_0 + x^\top \beta,$$

If $\theta = (\beta_0, \beta)$, a penalty $P_\alpha(\beta)$ and multiplier λ, then

$$\hat{\theta} \quad = \quad \text{argmin} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \beta_0 - x_i^\top \beta \right)^2 + \lambda P_\alpha(\beta) \right\} \qquad (1)$$

Elastic net

The penalty is a compromise between ridge and lasso

$$\begin{aligned} P_\alpha(\beta) &= \frac{1}{2}(1-\alpha)\,\|\beta\|_{\ell_2}^2 + \alpha\,\|\beta\|_{\ell_1} \\ &= \sum_{j=1}^{p}\left\{\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha|\beta_j|\right\} \end{aligned} \tag{2}$$

Be a ridge regression, if $(\alpha = 0)$, and lasso, if $(\alpha = 1)$.
Useful when $p \gg n$ or there are many correlated variables.

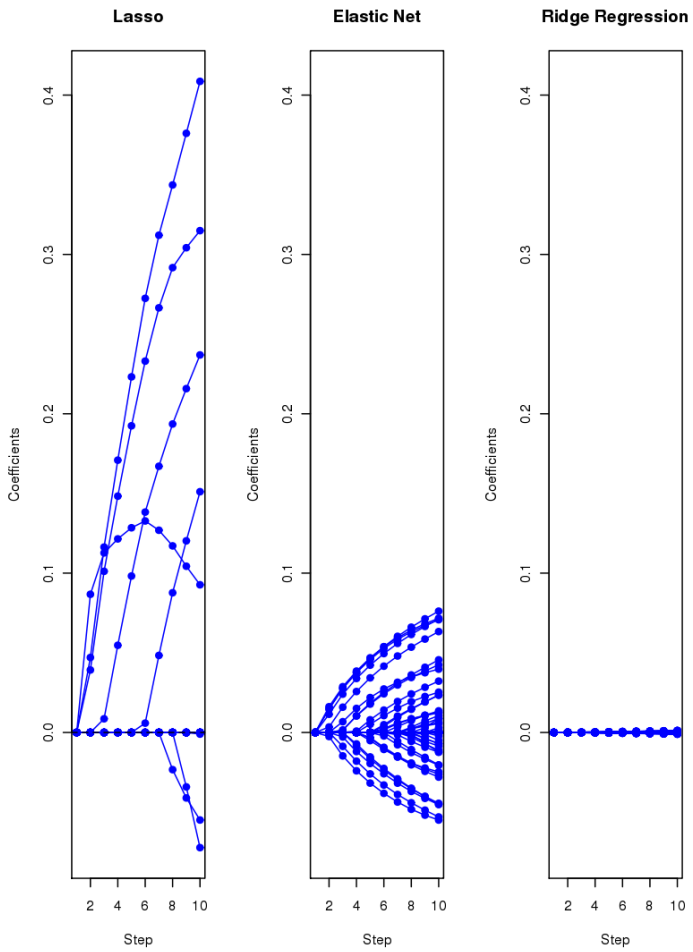


Figure 2: Profile of estimated coefficients, $\lambda = 1, \ldots, 10$, elastic net ($\alpha = 0.2$), on `Leukimia` data, $p = 72$ and $n = 3571$ (Friedman *et al.*, 2010)

Binary logit

Suppose $p(x_i) = \mathrm{P}(Y = 1|x_i)$, with $Y \in \{0, 1\}$,

$$\begin{aligned} \mathrm{P}(Y = 1|x) &= \left\{1 + e^{-(\beta_0 + x^\top \beta)}\right\}^{-1}, \qquad (3) \\ \mathrm{P}(Y = 0|x) &= \left\{1 + e^{(\beta_0 + x^\top \beta)}\right\}^{-1}, \\ \log\left\{\frac{\mathrm{P}(Y = 1|x)}{\mathrm{P}(Y = 0|x)}\right\} &= \beta_0 + x^\top \beta \end{aligned}$$

Penalized log likelihood

$$\max_{\beta_0,\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}\ell(\beta_0,\beta)-\lambda P_\alpha(\beta)\right\} \tag{4}$$

where

$$\ell(\beta_0,\beta)=\frac{1}{n}\sum_{i=1}^{n}y_i(\beta_0+x_i^\top\beta)-\log(1+e^{\beta_0+x_i^\top\beta}), \tag{5}$$

is a concave function of the parameter.
Maximizing $\ell(\beta_0,\beta)$ – iteratively reweightet least squares (IRLS).

Newton algorithm

If current parameter are $(\tilde{\beta}_0, \tilde{\beta})$, the quadratic approximation to $\ell(\beta_0, \beta)$ is,

$$\ell_Q(\beta_0, \beta) = -\frac{1}{2n} \sum_{i=1}^{n} w_i (z_i - \beta_0 - x_i^\top \beta)^2 + C(\tilde{\beta}_0, \tilde{\beta})^2 \tag{6}$$

where working response and wheight are,

$$\begin{aligned} z_i &= \tilde{\beta}_0 + x_i^\top \tilde{\beta} + \frac{y_i - \tilde{p}(x_i)}{\tilde{p}(x_i)(1 - \tilde{p}(x_i))} \\ w_i &= \tilde{p}(x_i)(1 - \tilde{p}(x_i)) \end{aligned}$$

Newton update is obtained by minimizing $\ell_Q(\beta_0, \beta)$.

Friedman approach

Coordinate descent is used to solve the penalized weighted least-square (PWLS)

$$\min_{\beta_0,\beta} \{-\ell_Q(\beta_0,\beta) + \lambda P_\alpha(\beta)\}$$

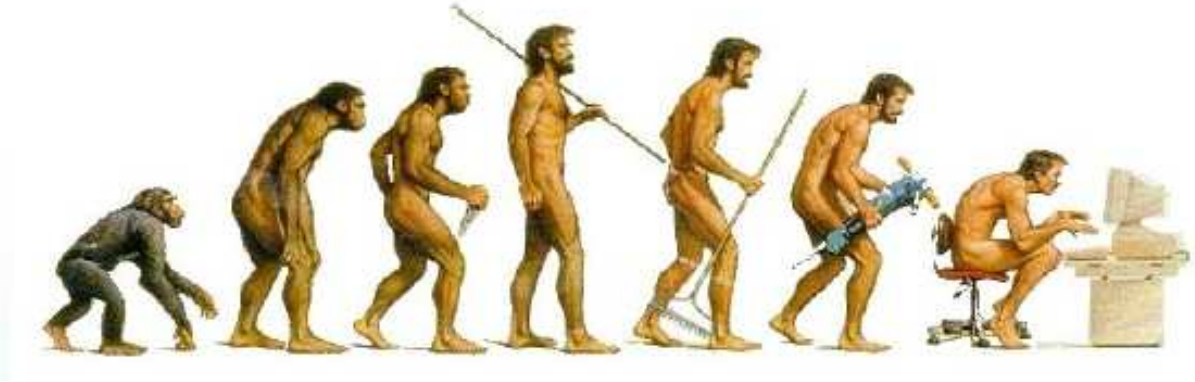
Sequence of nested loops:

- Outer loop: Decrement λ
- Middle loop: Update ℓ_Q using current parameter $(\tilde{\beta}_0, \tilde{\beta})$
- Inner loop: Run coordinate descent algorithm on PWLS

Global optimum

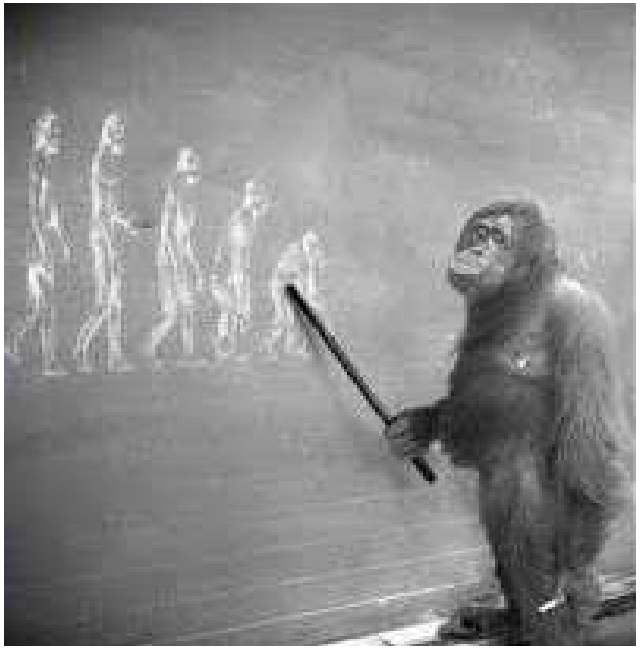
- Coordinate descent search local minimum
- How to choose α (and λ) ?
- More complicated

Evolutionary optimization



- Genetic Algorithm (GA)
- GA finds global optimum solution – parameters

What is a Genetic Algorithm ?



Genetics algorithm is search and optimization technique based on Darwin's principle on natural selection (Holland, 1975) ▸ GA

Classifier

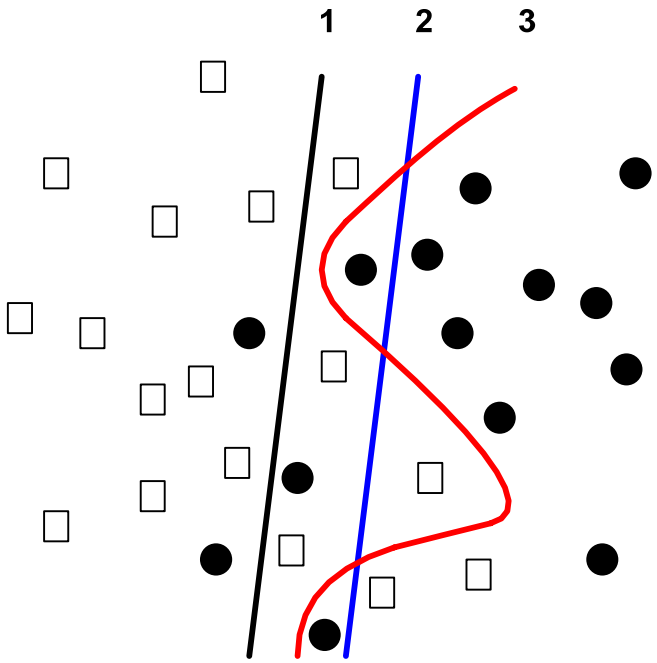


Figure 3: Linear classifier functions (1 and 2) and a non-linear one (3)

SVM

- Classification
  Data $D_n = \{(x_1, y_1), \ldots, (x_n, y_n)\} : \Omega \to (\mathcal{X} \times \mathcal{Y})^n$
  $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} \in \{-1, 1\}$

- Goal – to predict $\mathcal{Y}$ for new observation, $x \in \mathcal{X}$, based on information in D_n

Linearly (Non-) Separable Case

details

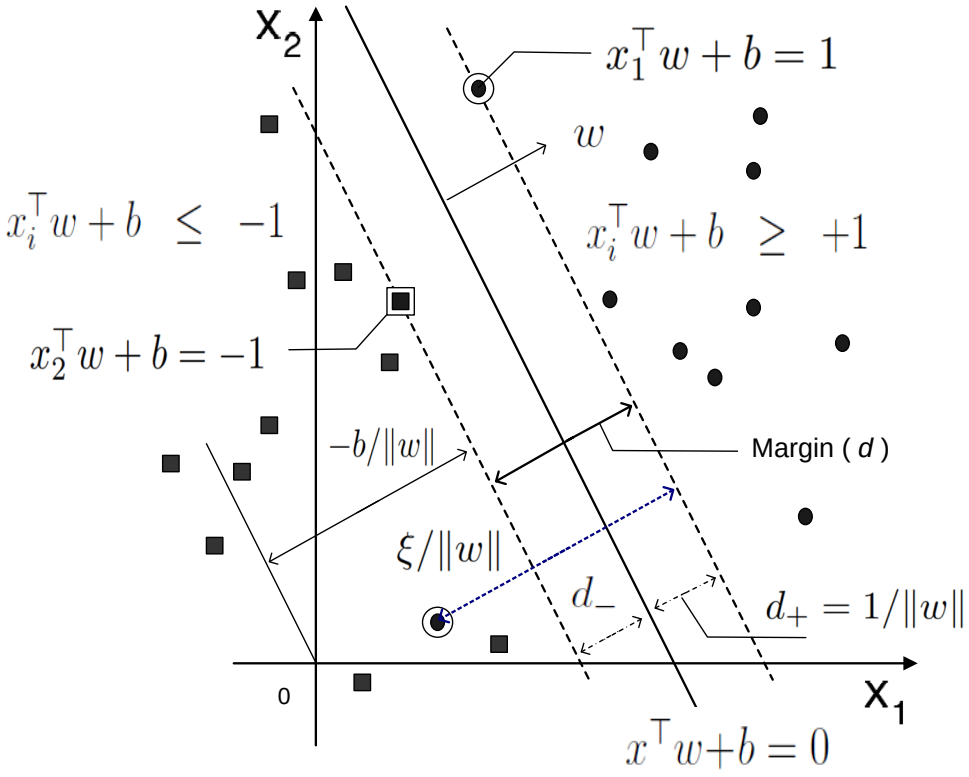


Figure 4: Hyperplane and its margin in linearly (non-) separable case

Loss function

$L(y, f(x))$	Loss type
$\{1 - yf(x)\}^2$	quadratic loss (ridge regression)
$\{1 - yf(x)\}_+ = \max\{0, 1 - yf(x)\}$	hinge loss (SVM)
$\log\{1 + exp(-yf(x))\}$	log-loss (logistic regression)
$\text{sign}\{-yf(x)\}$	$\{0, 1\}$ loss
$1/\{1 + exp(yf(x))\}$	sigmoidal loss

Table 1: Types of Loss function, with $f(x) = x^\top w + b$ is a score

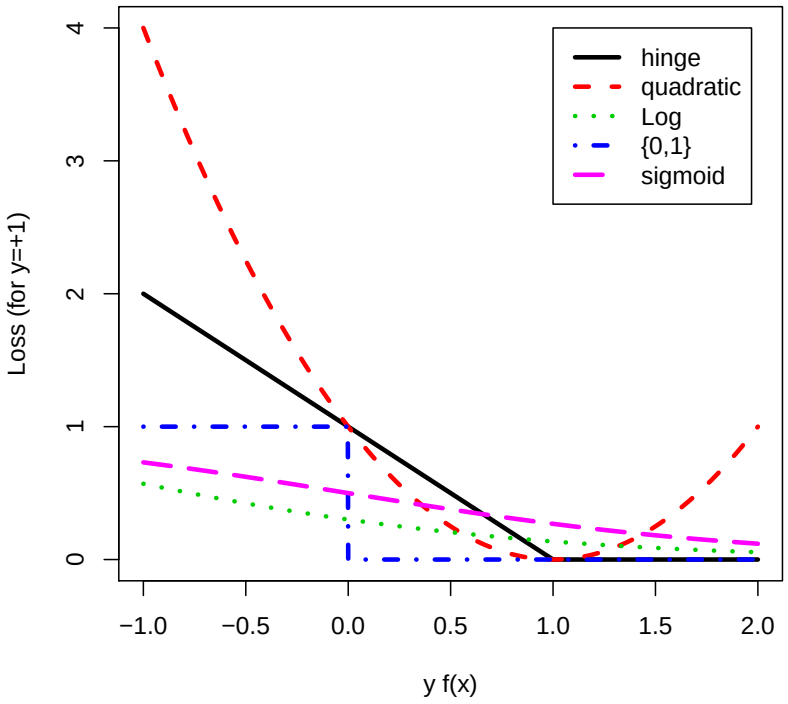


Figure 5: Plot of loss function for $y = 1, f(x) = 0.2x_1, x_1 \in [-5, 10]$. Similar plot for $y = -1$ but in opposite direction of $yf(x)$ axis

SVM dual problem

$$\max_{\alpha} L_D(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j \right\},$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

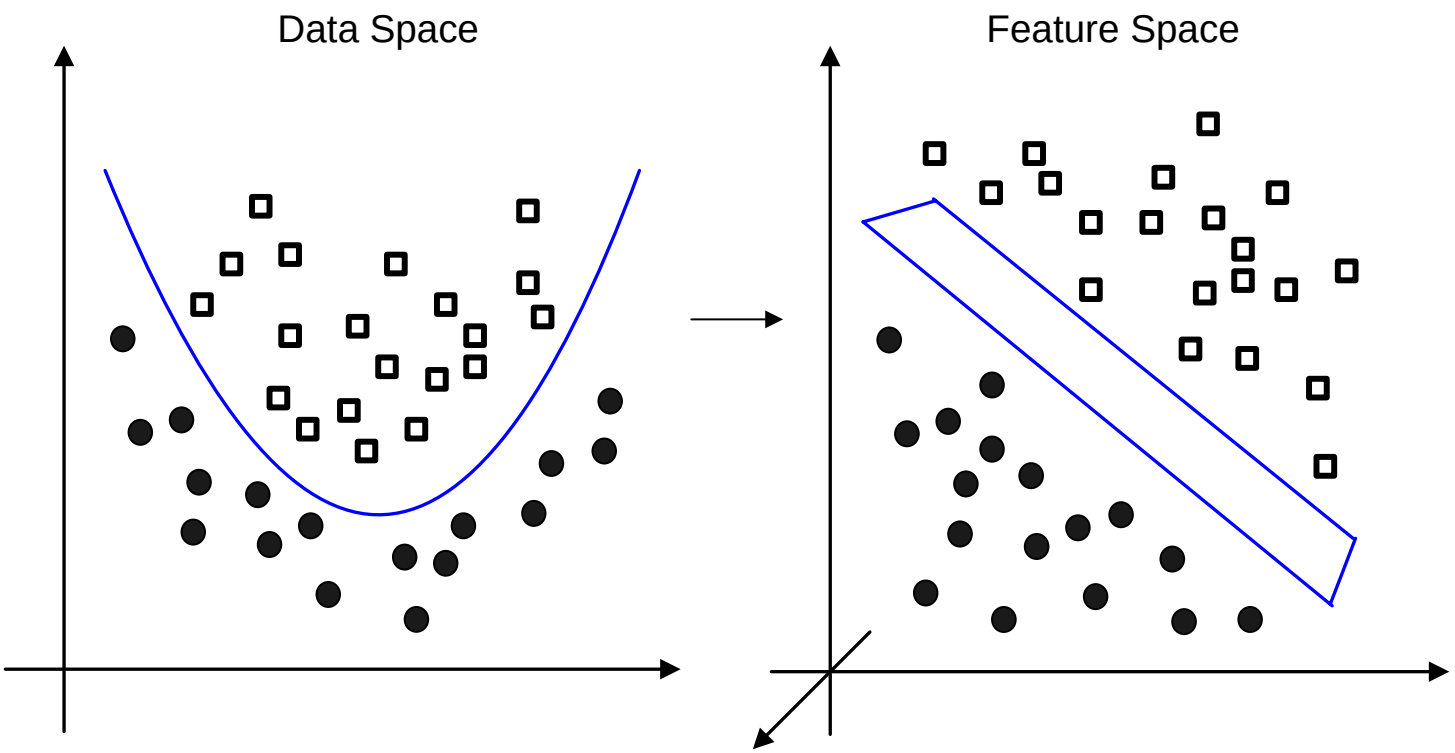


Figure 6: Mapping two dimensional data space into a three dimensional feature space, $\mathbb{R}^2 \mapsto \mathbb{R}^3$. The transformation $\Psi(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)^\top$ corresponds to $K(x_i, x_j) = (x_i^\top x_j)^2$

Non-linear SVM

$$\max_{\alpha} L_D(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\}$$

$$\text{s.t.} \qquad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

- Gaussian RBF kernel – $K(x_i, x_j) = \exp\left(-\frac{1}{\sigma} \|x_i - x_j\|^2\right)$
- Polynomial kernel – $K(x_i, x_j) = \left(x_i^\top x_j + 1\right)^p$

Structural Risk Minimization (SRM)

Search for the model structure $\mathcal{S}_h$,

$$\mathcal{S}_{h_1} \subseteq \mathcal{S}_{h_2} \subseteq \ldots \subseteq \mathcal{S}_{h*} \subseteq \ldots \subseteq \mathcal{S}_{h_k} = \mathcal{F}$$

such that $f \in \mathcal{S}_{h*}$ minimises the expected risk bound, with $f \subseteq \mathcal{F}$ is class of linear function and h is VC dimension
i.e.

$$SVM(h_1) \subseteq \ldots \subseteq SVM(h*) \subseteq \ldots \subseteq SVM(h_k) = \mathcal{F}$$

with h correspond to the value of SVM (kernel) parameter

GA - SVM

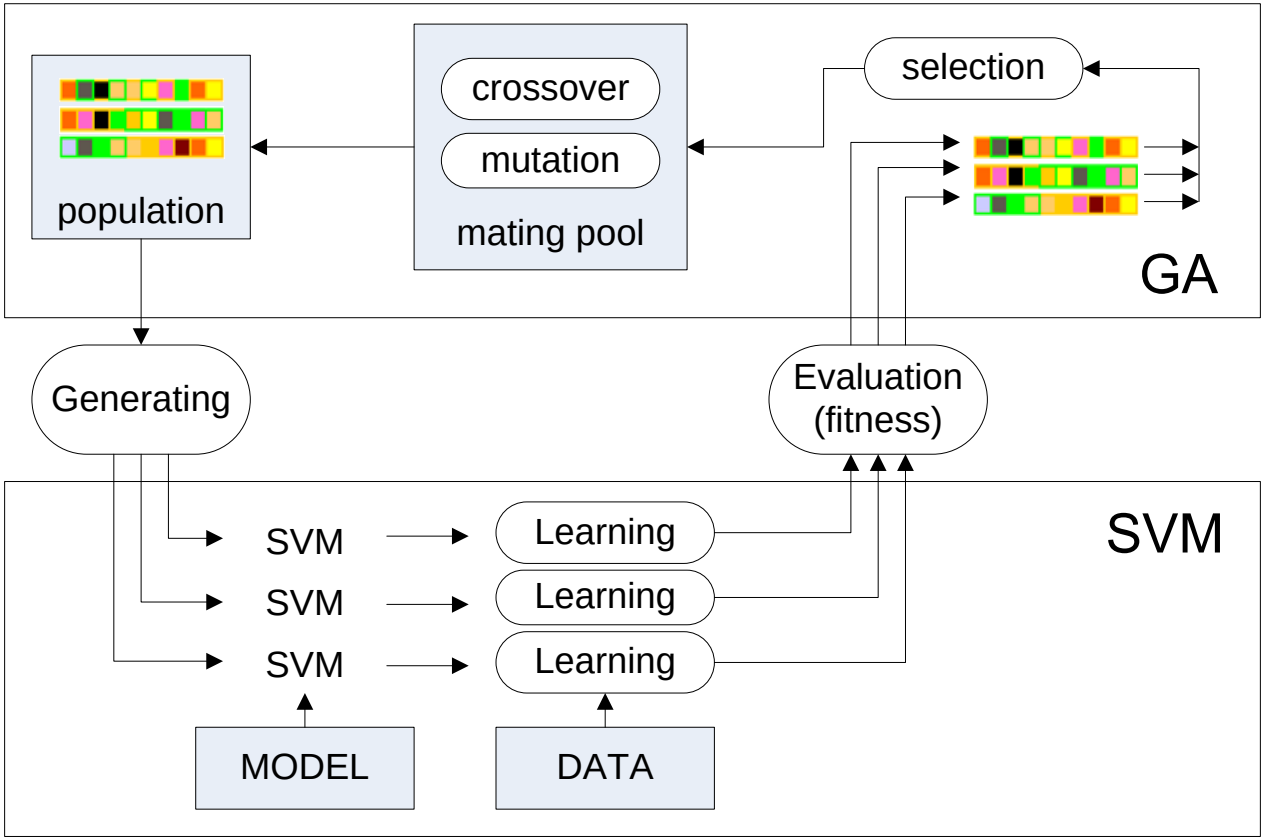


Figure 7: Iteration (generation) in GA-SVM

Validation of scores

Discriminatory power (of the score)

- Cumulative Accuracy Profile (CAP) curve
- Receiver Operating Characteristic (ROC) curve
- Accuracy, Specificity, Sensitivity

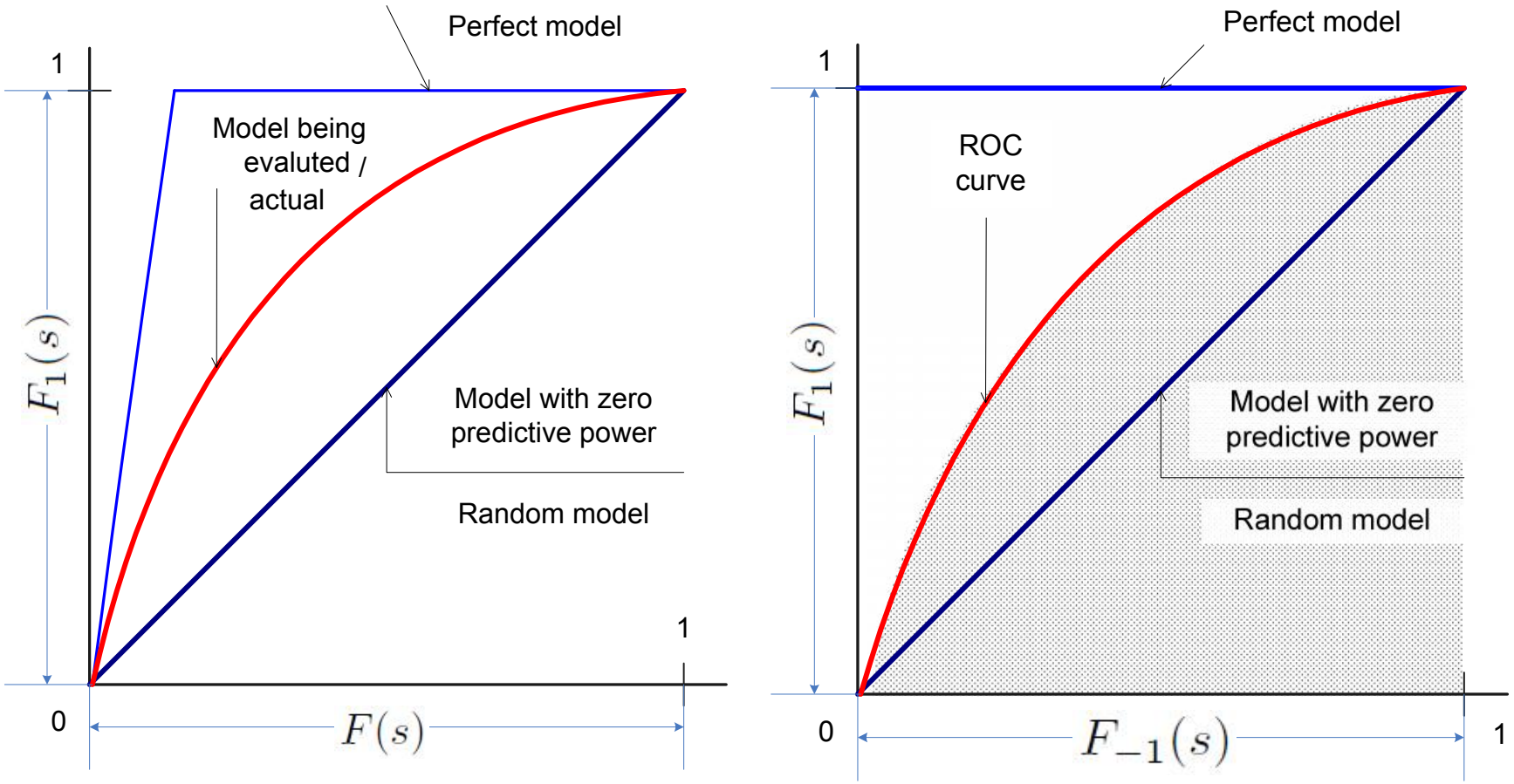


Figure 8: CAP curve (left) and ROC curve (right)

Discriminatory power

- **Cumulative Accuracy Profile** (CAP) curve
  - CAP/Power/Lorenz curve $\rightarrow$ Accuracy Ratio (AR)
  - Total sample vs. default sample
- **Receiver Operating Characteristic** (ROC) curve
  - ROC curve $\rightarrow$ Area Under Curve (AUC)
  - Non-default sample vs. default sample
- Relationship: **AR** = 2 **AUC** - 1

Discriminatory power (cont'd)

		sample	
		default (1)	non-default (-1)
predicted	(1)	True Positive (TP)	False Positive (FP)
	(-1)	False Negative (FN)	True Negative (TN)
total		P	N

- Accuracy, $\mathrm{P}(\widehat{Y} = Y) = \frac{TP+TN}{P+N}$
- Specificity, $\mathrm{P}(\widehat{Y} = -1 | Y = -1) = \frac{TN}{N}$
- Sensitivity, $\mathrm{P}(\widehat{Y} = 1 | Y = 1) = \frac{TP}{P}$

Credit reform data

Example

type	solvent (%)	insolvent (%)	total (%)
Manufacturing	27.37 (26.06)	25.70 (1.22)	27.29
Construction	13.88 (13.22)	39.70 (1.89)	15.11
Wholesale and retail	24.78 (23.60)	20.10 (0.96)	24.56
Real estate	17.28 (16.46)	9.40 (0.45)	16.90
total	83.31 (79.34)	94.90 (4.52)	83.86
others	16.69 (15.90)	5.10 (0.24)	16.14
#	20,000	1,000	21,000

Table 2: Credit reform data

Pre-processing

year	solvent # (%)	insolvent # (%)	total # (%)
1997	872 (9.08)	86 (0.90)	958 (9.98)
1998	928 (9.66)	92 (0.96)	1020 (10.62)
1999	1005 (10.47)	112 (1.17)	1117 (11.63)
2000	1379 (14.36)	102 (1.06)	1481 (15.42)
2001	1989 (20.71)	111 (1.16)	2100 (21.87)
2002	2791 (29.07)	135 (1.41)	2926 (30.47)
total	8964 (93.36)	638 (6.64)	9602 (100)

Table 3: Pre-processed credit reform data

Full model, $X_1, \ldots, X_{28}$

- Predictors – 28 financial ratio variables detail
- Population (# solutions) – 20
- Evolutionary iteration (generation) – 100
- Elitism – 0.2 of population
- Crossover rate – 0.5, mutation rate – 0.1
- Optimal SVM parameters – $\sigma = 1/178.75$ and $C = 63.44$

Scenario

Finding

scenario	training set	testing set
Scenario-1	1997	1998
Scenario-2	1997-1998	1999
Scenario-3	1997-1999	2000
Scenario-4	1997-2000	2001
Scenario-5	1997-2001	2002

Table 4: Training and testing data set

Quality of classification

training	TE (CV)	testing	TE (CV)
1997	0 (8.98)	1998	0 (9.02)
1997-1998	0 (8.99)	1999	0 (10.03)
1997-1999	0 (9.37)	2000	0 (6.89)
1997-2000	0 (8.57)	2001	5.43 (5.86)
1997-2001	0 (4.55)	2002	4.68 (4.61)

Table 5: Percentage of Training Error (TE) and Cross-Validation (5-fold CV)

Current findings

- SVM with optimal parameter is more robust to the imbalanced data set
- Evolutionary feature selection (using Genetic Algorithm) could find global solution of SVM parameter optimization
- More investigation to variable selection

Variable Selection and Optimization in Default Prediction

Dedy Dwi Prastyo
Wolfgang Karl Härdle

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

References

Chen, S., Härdle, W. and Moro, R.
Estimation of Default Probabilities with Support Vector Machines
Quantitative Finance, 2011, 11, 135 - 154

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R.
Least angle regression
The Annals of Statistics, 2004, 32(2), 407 - 449

Friedman, J., Hastie, T., Hoefling, H. and Tibshirani, R.
Pathwise coordinate optimization
The Annals of Applied Statistics, 2007, 2(1), 302 - 332

References

back

Friedman, J., Hastie, T. and Tibshirani, R.
Regularization path for generalized linear models via coordinate descent
Journal of Statistical Software, 2010, 33(1)

Härdle, Lee, Y-J., Schäfer, D. and Yeh, Y-R.
Variable Selection and Oversampling in the Use of Smooth Support Vector Machine for Predicting the Default Risk of Companies
Journal of Forecasting, 2009, 28, 512 - 534

Holland, J.H.
Adaptation in Natural and Artificial Systems
University of Michigan Press, 1975

References

Karatzoglou, A. and Meyer, D.
Support Vector Machines in R
Journal of Statistical Software, 2006, 15:9, 1-28

Rosset, S. and Zhu, J.
Piecewise linear regularized solution path
The Annals of Statistics, 2007, 35(3), 1012 - 1030

Tibshirani, R.
Regression shrinkage and selection via Lasso
Journal of the Royal Statistical Society B, 1996, 58, 267 - 288

References

Tseng, P.
Convergence of a block coordinate descent method for nondifferentiable minimization
Journal of Optimization Theory and Application, 2001, 109, 475 - 494

Van der Kooij, A.
Prediction accuracy and stability of regression with optimal scaling transformation
Ph.D. thesis, Department of data theory, University of Leiden, 2001

Linearly Separable Case

back

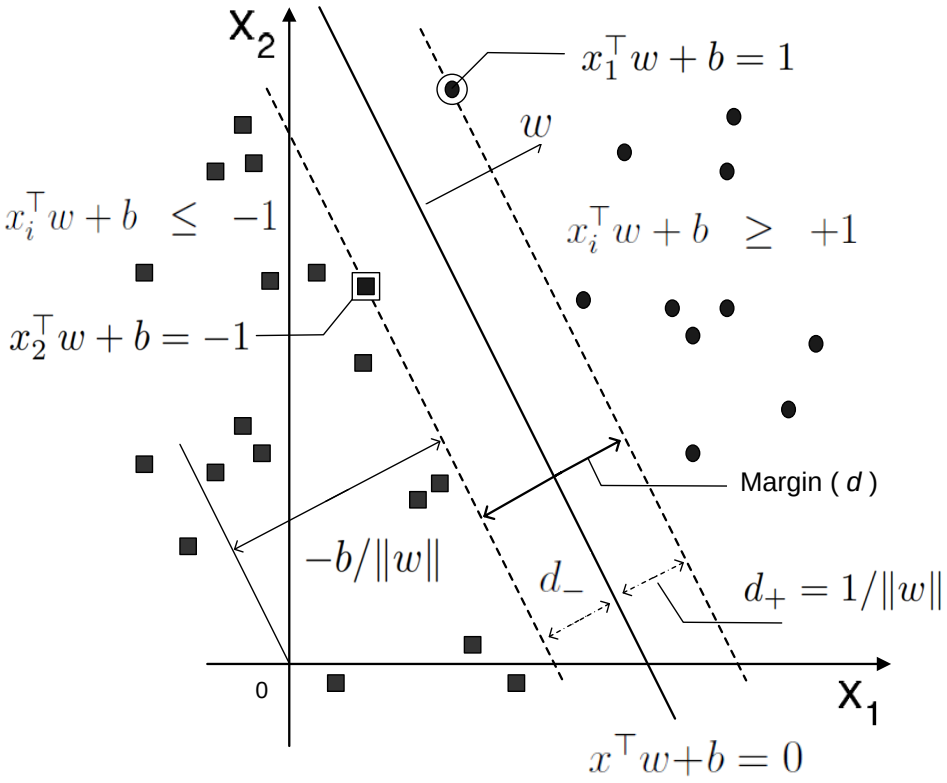


Figure 9: Separating hyperplane and its margin in linearly separable case

- Choose $f \in \mathcal{F}$ such that margin $(d_- + d_+)$ is maximal
- No error separation, if all $i = 1, 2, ..., n$ satisfy

$$x_i^\top w + b \geq +1 \qquad \text{for} \quad y_i = +1$$
$$x_i^\top w + b \leq -1 \qquad \text{for} \quad y_i = -1$$

- Both constraints are combined into

$$y_i(x_i^\top w + b) - 1 \geq 0 \qquad i = 1, 2, ..., n$$

- Distance between margins and the separating hyperplane is $d_+ = d_- = 1/\|w\|$
- Maximize the margin, $d_+ + d_- = 2/\|w\|$, could be attained by minimizing $\|w\|$ or $\|w\|^2$
- Lagrangian for the primal problem

$$L_P(w, b) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \{y_i(x_i^\top w + b) - 1\}$$

Karush-Kuhn-Tucker (KKT) first order optimality conditions

$$\frac{\partial L_P}{\partial w_k} = 0 : \qquad w_k - \sum_{i=1}^{n} \alpha_i y_i x_{ik} = 0 \qquad k = 1, ..., d$$

$$\frac{\partial L_P}{\partial b} = 0 : \qquad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$y_i(x_i^\top w + b) - 1 \geq 0 \qquad i = 1, ..., n$$

$$\alpha_i \geq 0$$

$$\alpha_i \{ y_i(x_i^\top w + b) - 1 \} = 0$$

- Solution $w = \sum_{i=1}^{n} \alpha_i y_i x_i$, therefore

$$\begin{aligned}
\frac{1}{2}\|w\|^2 &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^\top x_j \\
-\sum_{i=1}^{n}\alpha_i\{y_i(x_i^\top w + b) - 1\} &= -\sum_{i=1}^{n}\alpha_i y_i x_i^\top \sum_{j=1}^{n}\alpha_j y_j x_j + \sum_{i=1}^{n}\alpha_i \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^\top x_j + \sum_{i=1}^{n}\alpha_i
\end{aligned}$$

- Lagrangian for the dual problem

$$L_D(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^\top x_j$$

- Primal and dual problems

$$\min_{w,b} \quad L_P(w, b)$$

$$\max_{\alpha} \quad L_D(\alpha) \qquad \text{s.t.} \quad \alpha_i \geq 0, \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

- Optimization problem is convex, therefore the dual and primal formulations give the same solution
- Support vector, a point i for which $y_i(x_i^\top w + b) = 1$ holds

Linearly Non-separable Case

back

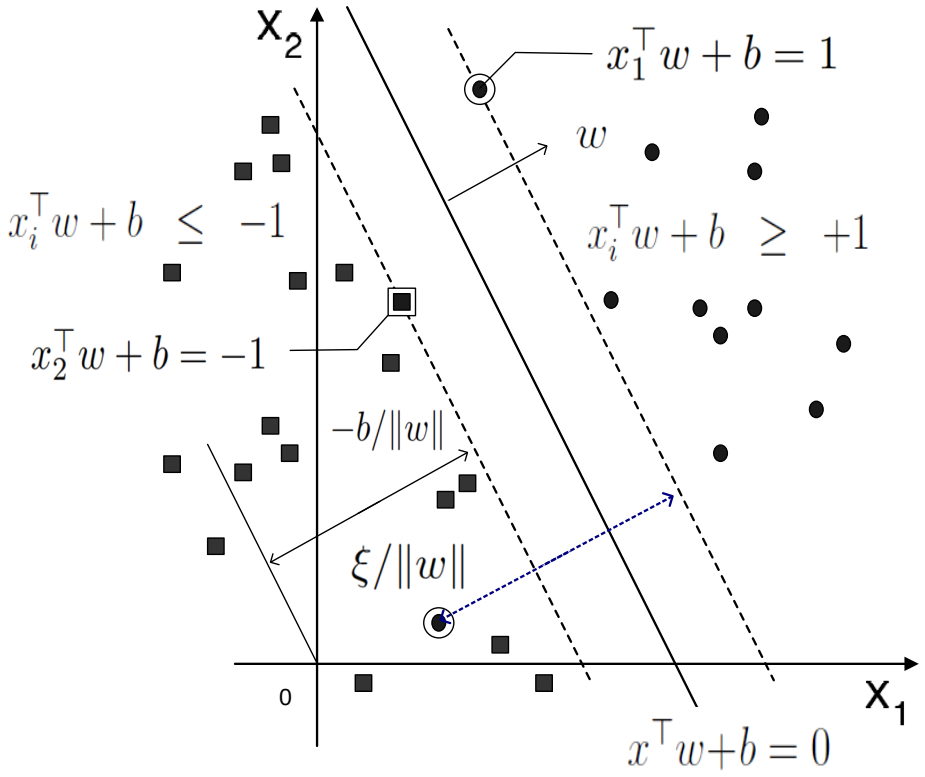


Figure 10: Hyperplane and its margin in linearly non-separable case

- Slack variables ξ_i represent the violation from strict separation

$$
\begin{aligned}
x_i^\top w + b &\geq 1 - \xi_i && \text{for} && y_i = 1, \\
x_i^\top w + b &\leq -1 + \xi_i && \text{for} && y_i = -1, \\
\xi_i &\geq 0
\end{aligned}
$$

- constraints are combined into

$$y_i(x_i^\top w + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0$$

- If $\xi_i > 0$, the objective function is

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$$

- Lagrange function for the primal problem

$$L_P(w, b, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\{y_i\left(x_i^\top w + b\right) - 1 + \xi_i\} - \sum_{i=1}^{n}\mu_i\xi_i,$$

where $\alpha_i \geq 0$ and $\mu_i \geq 0$ are Lagrange multipliers

- Primal problem

$$\min_{w,b,\xi} L_P(w, b, \xi)$$

First order conditions

$$\frac{\partial L_P}{\partial w_k} = 0 : \qquad w_k - \sum_{i=1}^{n} \alpha_i y_i x_{ik} = 0$$

$$\frac{\partial L_P}{\partial b} = 0 : \qquad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\frac{\partial L_P}{\partial \xi_i} = 0 : \qquad C - \alpha_i - \mu_i = 0$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad \mu_i \geq 0, \quad \mu_i \xi_i = 0$$

$$\alpha_i \{ y_i (x_i^\top w + b) - 1 + \xi_i \} = 0$$

- Note that $\sum_{i=1}^{n} \alpha_i y_i b = 0$. Translate primal problem into

$$L_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j + \sum_{i=1}^{n} \xi_i (C - \alpha_i - \mu_i)$$

- Last term is 0, therefore the dual problem is

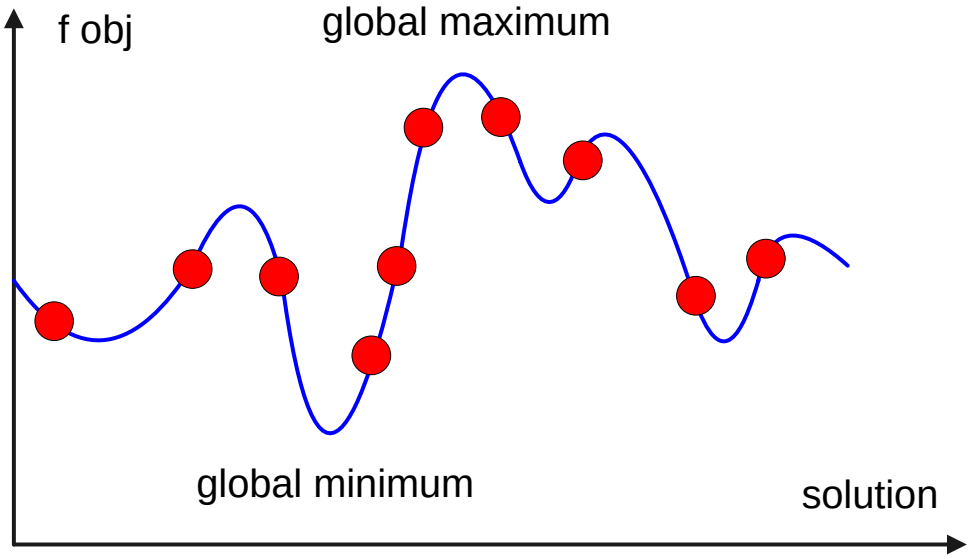
$$\max_{\alpha} L_D(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j \right\},$$

$$\text{s.t.} \qquad 0 \le \alpha_i \le C, \qquad \sum_{i=1}^{n} \alpha_i y_i = 0$$

back

GA – Initialization

Back



decoding

real number (solution)	binary encoding population
122	1 1 1 1 0 1 0
77	0 1 1 0 1 1 0
54	0 0 1 0 0 1 1
19	1 0 0 0 1 0 1
69	0 1 0 1 1 0 0
44	1 1 1 0 0 0 1
113	1 1 0 1 1 1 0
110	1 0 1 1 0 0 1
89	0 0 1 0 1 1 1
23	1 1 0 1 0 0 1
105	

Figure 11: GA at first generation

GA – Convergency

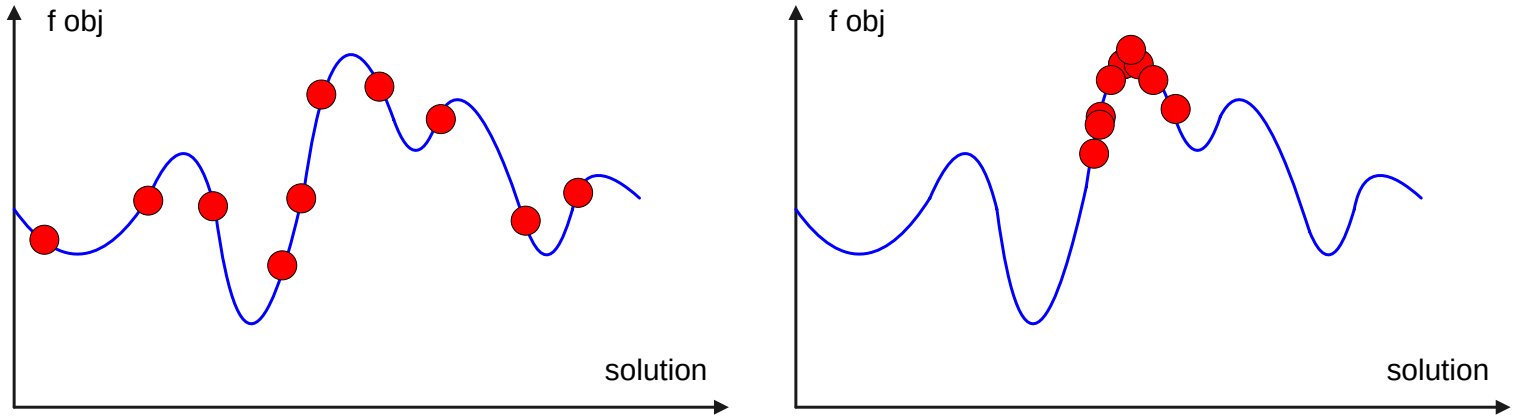


Figure 12: Solutions at 1st generation (left) and r^{th} generation (right)

GA – Decoding

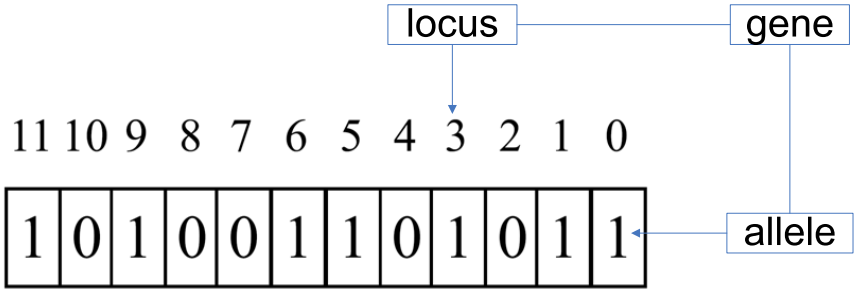


Figure 13: Decoding

$$\theta = \theta_{lower} + (\theta_{upper} - \theta_{lower}) \frac{\sum_{i=0}^{l-1} a_i 2^i}{2^l}$$

where θ is solution (i.e. parameter), a is allele

GA – Fitness evaluation

- Calculate $f(\theta_i)$, $\quad i = 1, \ldots, popsize$
- Evaluate fitness, $f_{dp}(\theta_i)$
  $f_{dp}(\theta_i)$ – AR, AUC, accuracy, specificity, sensitivity
- Relative fitness, $p_i = \frac{f_{dp}(\theta^i)}{\sum_{k=i}^{popsize} f_{dp}(\theta^i)}$

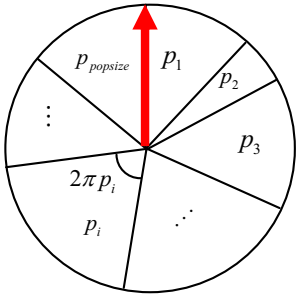
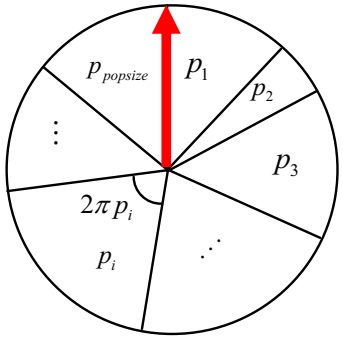


Figure 14: Proportion to be choosen in the next iteration (generation)

GA – Roulette wheel



- $rand \sim \mathrm{U}(0,1)$
- Select i^{th} chromosome if $\sum_{i=1}^{k} p_i < rand < \sum_{i=1}^{k+1} p_i$
- Repeat *popsize* times to get *popsize* new chromosomes

GA – Crossover

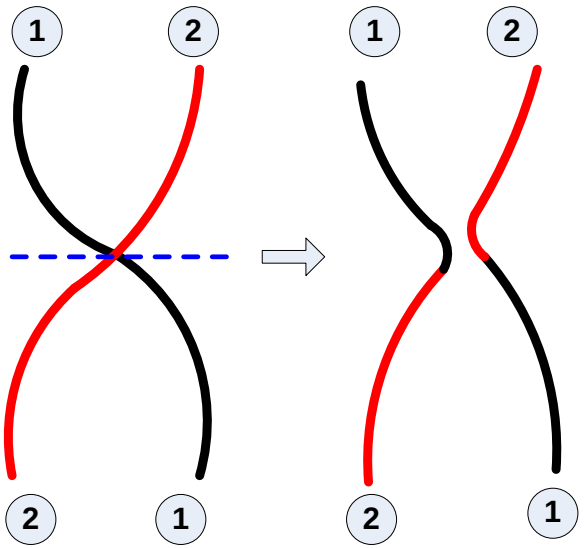


Figure 15: Crossover in nature

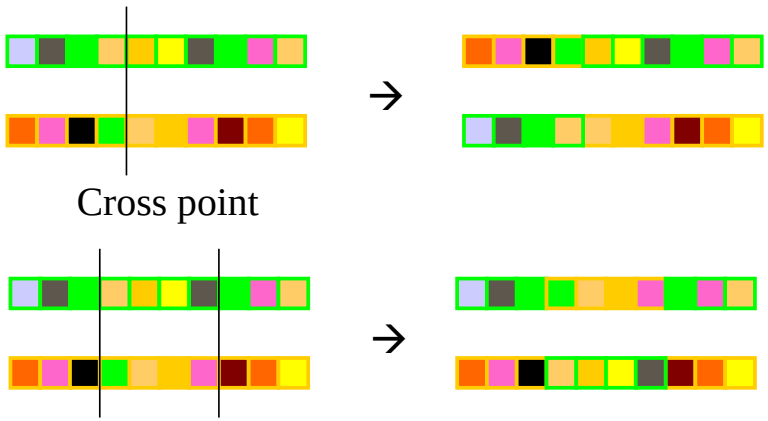


Figure 16: Randomly choosen one-point crossover (top) and two-points crossover (bottom)

GA – Reproductive operator

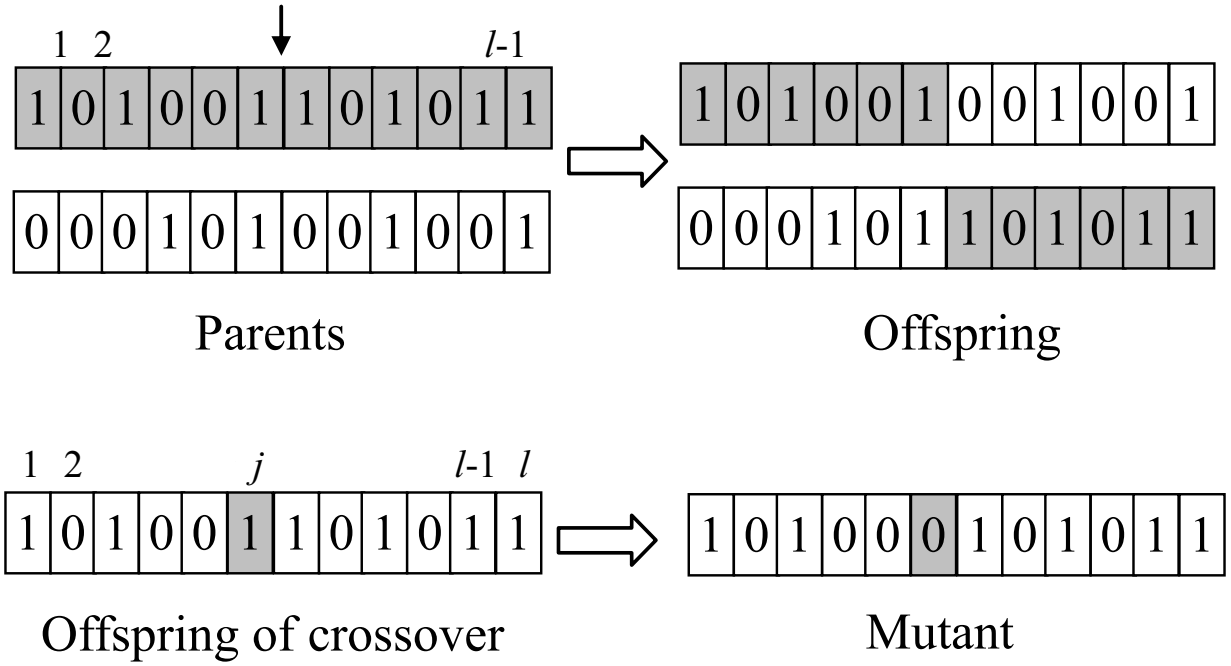


Figure 17: One-point crossover (top) and bit-flip mutation (bottom)

GA – Elitism

- Best solution in each iteration is maintained in another memory place
- New population replaces the old one, check whether best solution is in the population

  If not, replace any one in the population with best solution

Nature to Computer Mapping

Back

Nature	GA-SVM
Population	Set of parameter
Individual (phenotype)	Parameters
Fitness	Discriminatory power
Chromosome (genotype)	Encoding of parameter
Gene	Binary encoding
Reproduction	Crossover
Generation	Iteration

Table 6: Nature to GA-SVM mapping

Examples

- Small sample: 100 solvent and insolvent companies
- Credit reform data
- X3 – Operating Income / Total Asset
- X24 – Account Payable / Total Asset

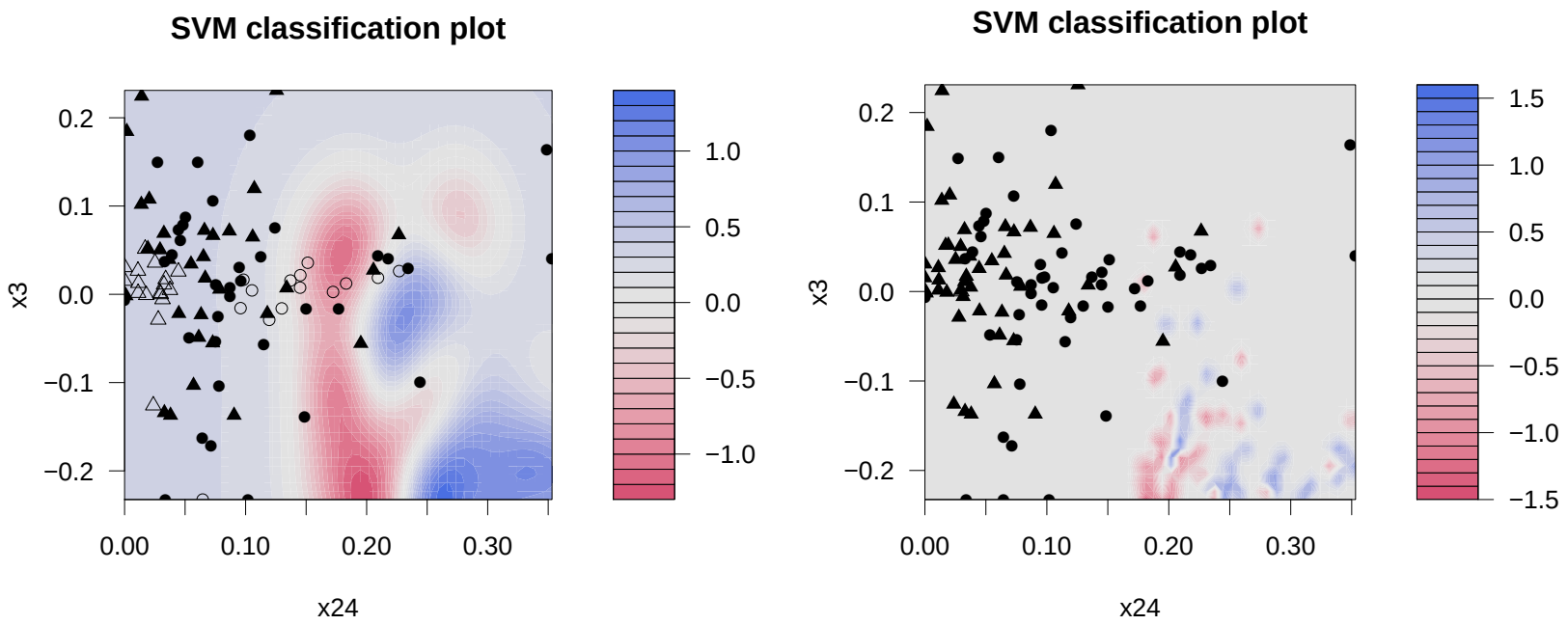


Figure 18: SVM plot, $C = 1$ and $\sigma = 1/2$, misclass. rate 0.19 (left) and GA-SVM, $C = 14.86$ and $\sigma = 1/121.61$, misclass. rate 0 (right).

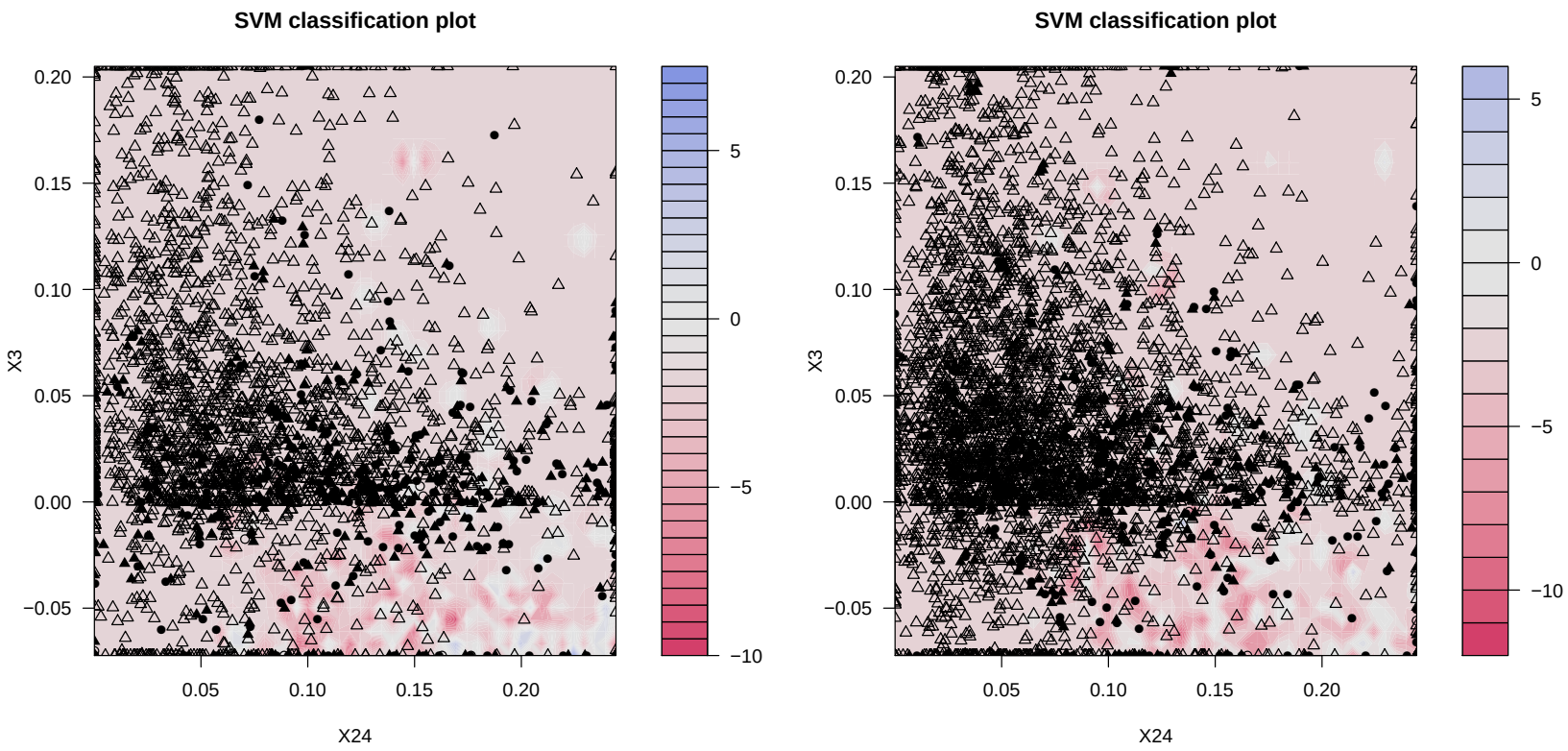


Figure 19: GA-SVM ($C = 187.93$ and $\sigma = 1/195.16$) plot of training data, misclass. rate 2.38%, and testing data, misclass. rate 1.37%. Back

FR: Profitability

Ratio No.	Definition	Ratio
$x1$	NI/TA	Return on assets (ROA)
$x2$	NI/Sales	Net profit margin
$x3$	OI/TA	Operating Income/Total assets
$x4$	OI/Sales	Operating profit margin
$x5$	EBIT/TA	EBIT/Total assets
$x6$	(EBIT+AD)/TA	EBITDA
$x7$	EBIT/Sales	EBIT/Sales

Table 7: Defintions of financial ratios.

FR: Leverage

Ratio No.	Definition	Ratio
$x8$	Equity/TA	Own funds ratio (simple)
$x9$	(Equity-ITGA)/ (TA-ITGA-Cash-LB)	Own funds ratio (adjusted)
$x10$	CL/TA	Current liabilities/Total assets
$x11$	(CL-Cash)/TA	Net indebtedness
$x12$	TL/TA	Total liabilities/Total assets
$x13$	Debt/TA	Debt ratio
$x14$	EBIT/ Interest expenses	Interest coverage ratio

Table 8: Defintions of financial ratios.

FR: Liquidity

Ratio No.	Definition	Ratio
$x15$	Cash/TA	Cash/Total assets
$x16$	Cash/CL	Cash ratio
$x17$	QA/CL	Quick ratio
$x18$	CA/CL	Current ratio
$x19$	WC/TA	Working Capital
$x20$	CL/TL	Current liabilities/ Total liabilities

Table 9: Defintions of financial ratios.

FR: Activity

Ratio No.	Definition	Ratio
$x21$	TA/Sales	Asset turnover
$x22$	INV/Sales	Inventory turnover
$x23$	AR/Sales	Account receivable turnover
$x24$	AP/Sales	Account payable turnover
$x25$	Log(TA)	Log(Total assets)

Table 10: Defintions of financial ratios.

FR

back

Ratio No.	Definition
$x26$	increase (decrease) in inventories / inventories
$x27$	increase (decrease) in liabilities / total liabilities
$x28$	increase (decrease) in cash flows / cash and cash equivalent

Table 11: Defintions of financial ratios.

Findings

- Härdle et al. (2009): Smooth SVM – overall mean of correct predictions ranging from 70% to 78% (misclassification: 22% to 30%)
- Chen, Härdle and Moro (2011):
  - Most of the models tested, AR in 43.50% and 60.51%
  - SVM (grid search optiization): percentage of correctly classified out-of-sample 71.85%
  - Logit model: percentage of correctly classified out-of-sample 67.24%

Findings

Back

- Zang and Härdle (2010):

Performance measure	Logit (%)	CART (%)	BACT (%)
Ovearal misclass. rate	30.2	33.8	26.6
Type I misclass. rate	28.3	27.2	27.6
Type II misclass. rate	30.3	34.3	26.5
AR	52.1	58.7	60.4

Table 12: Average value (of bootstrap) of performance measures