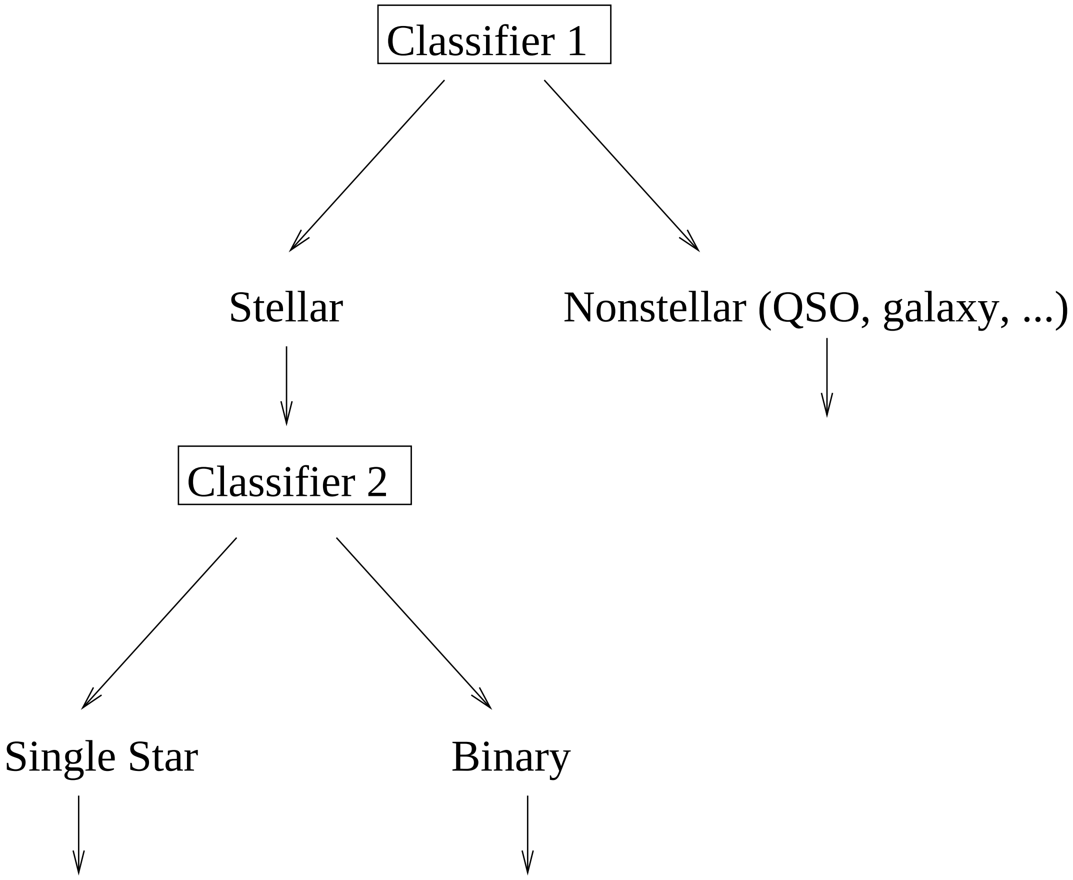
Identification of spectroscopic binaries from single shot medium band photometry

Philip G. Willemsen
Torsten A. Kaempf
Coryn A.L. Bailer-Jones
Klaas S. de Boer

ICAP-Variability-Alerts joint working group meeting

15-16 April 2004

1.1 Introduction



1.2 Introduction

Training set $\Omega = \{\mathbf{x}_i, y_i\}_{i=1}^{N}$ with y_i = targets (outputs):

'binary'/'single star'

Use *supervised learning* to approximate decision function

$$d_\Omega\colon\ \mathbb{R}^{11} \longrightarrow \{+1,-1\}$$

Predict class labels y_j from filter fluxes $\mathbf{x}_j$

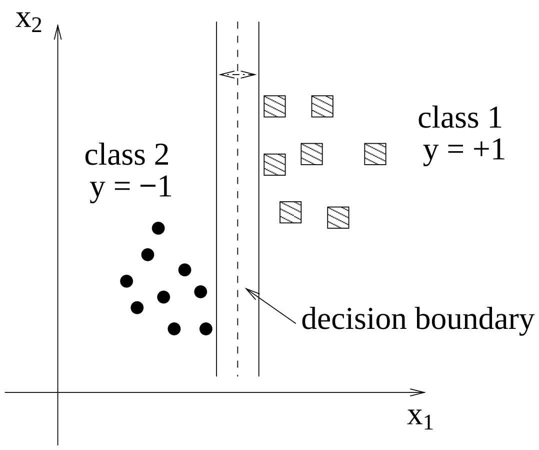
Support Vector Machines (SVM) (Vapnik 1995, 1998)

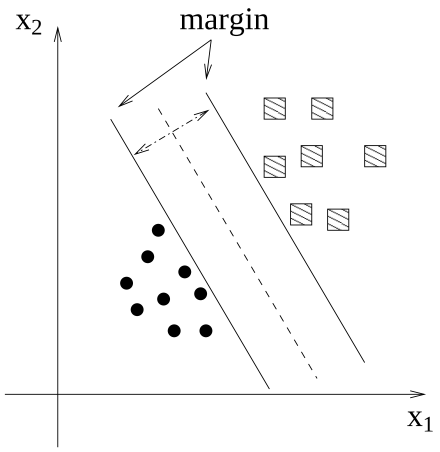
- universal function approximators
- Statistical Learning Theory (SLT),
  Structural Risk Minimization

1. Linear SVM - hard margins
2. Linear SVM - soft margins
3. Nonlinear SVM
4. Simulations - methods
5. Simulations - results

2.1 Linear SVM for linearly separable data

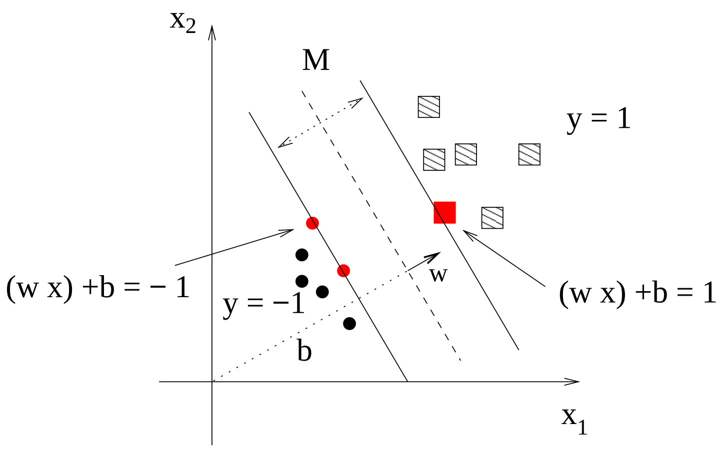
Hard margin classifier





SLT: restrict class of functions to one with a complexity suitable for the amount of training data
$\Rightarrow$ among all hyperplanes that minimize the training error, find the one with the largest margin

2.2 Optimal canonical hyperplane



(w x) + b = 0
y(x) = sign(w x + b)

since $\{\mathbf{x} \mid \mathbf{wx} + \mathrm{b} = 0\} \equiv \{\mathbf{x} \mid \lambda\mathbf{wx} + \lambda\mathrm{b} = 0\}$

canonical hyperplane : $\min_{i \in \Omega} |\mathbf{wx} + \mathrm{b}| = 1$

optimal hyperplane = canonical hyperplane with *maximal margin* M

for support vector $\mathbf{x}_s$: $\quad \Delta = M/2 = \frac{|\mathbf{w}\ \mathbf{x}_S + b)|}{\|\mathbf{w}\|} = \frac{1}{\|w\|}$

$$\Rightarrow M = \frac{2}{\|w\|}$$

2.3 Optimization

Learning: minimize $\|w\|^2$ with inequality constraints
$y_k[\mathbf{w}\ \mathbf{x}_k + b] \geq 1$

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}\ \mathbf{w} - \sum_{i=1}^{N} \alpha_i \{y_i[\mathbf{w}\mathbf{x}_i + b] - 1\}$$

with $\alpha_i \geq 0$

Karush-Kuhn-Tucker conditions

$$w_o = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i$$

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

$$\alpha_i \{y_i[\mathbf{w}\ \mathbf{x}_i + b] - 1\} = 0$$

$$L_d(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \mathbf{x}_i\ \mathbf{x}_j$$

$\alpha_i \geq 0$ and $\sum_{i}^{N} \alpha_i y_i = 0$

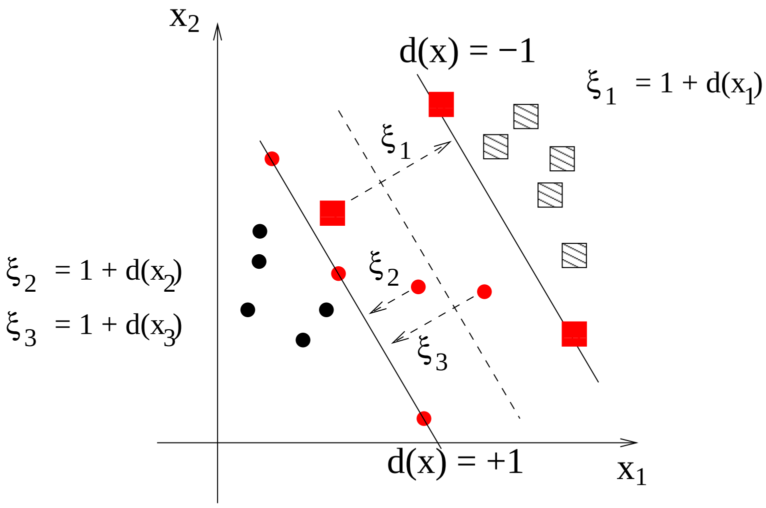
for support vectors (on the boundary): $\alpha_i \neq 0$

for all other points: $\alpha_i = 0$

indicator function: $y(\mathbf{x}) = sign(w\mathbf{x} + b)$
$= sign(\sum_{i=1}^{N_{SV}} \alpha_i y_i (\mathbf{x}_i \mathbf{x}) + b)$

3 Linear SVM for overlapping classes

Soft margin classifier



introduce *slack* variables $\xi_i \geq 0$, (i=1,N)

hyperplane must satisfy

$$y_i[\mathbf{w}\ \mathbf{x}_i + b] \geq 1 - \xi$$

minimize cost function with *penalty* term

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

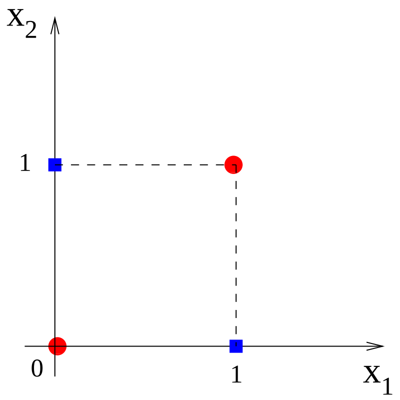
$$L_d(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \mathbf{x}_i\ \mathbf{x}_j$$

$$\mathrm{C} \geq \alpha_i \geq 0$$

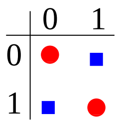
$$\sum_{i=1}^{N} y_i \alpha_i = 0$$

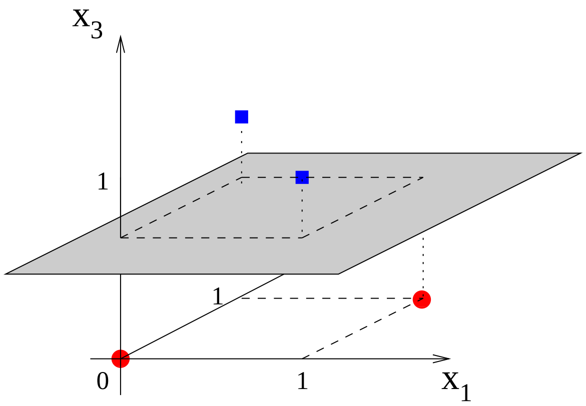
4.1 Nonlinear SVM - general case

Cover's theorem (1965): '*A complex pattern-classification problem cast in a high dimensional space nonlinearly is more likely to be linearly separable than in a low-dimensional space*'



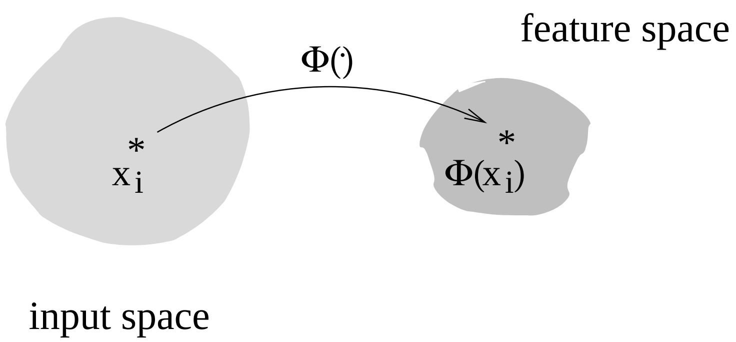
XOR problem





4.2 Nonlinear SVM - general case

1. $\mathbf{x} \in \mathbb{R}^N \rightarrow \Phi(x_i) \in \mathbb{R}^M$



2. construct optimal separating hyperplane in feature space

$$\sum_{j=1}^{M} w_j \phi_j(\mathbf{x}) + b = 0$$

$$d(\mathbf{x}) = \sum_{i=1}^{N_{SV}} \alpha_i y_i \Phi\mathbf{(x)}\ \Phi\mathbf{(x}_i\mathbf{)} + b$$

problems:

- choice of mapping Φ
- 'curse of dimensionality'

4.3 Nonlinear SVM - Inner-Product Kernel

$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)$

constraints:

- Mercer conditions
- positive definit

polynomial	$K(\mathbf{x}, \mathbf{x}_i) = [\mathbf{x} \cdot \mathbf{x}_i + 1]^d$
sigmoid	$K(\mathbf{x}, \mathbf{x}_i) = tanh[\beta_o \mathbf{x} \cdot \mathbf{x}_i + \beta_1]$
Gaussian RBF	$K(\mathbf{x}, \mathbf{x}_i) = exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|)$

$$L_d(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$$

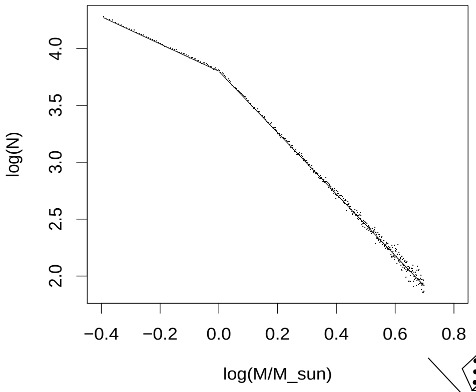
$$\mathsf{C} \geq \alpha_i \geq 0$$

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

indicator function: $y(\mathbf{x}) = sign(w\mathbf{x} + b)$

$$= sign(\sum_{i=1}^{N_{SV}} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b)$$

5.1 Simulation of single and double stars



$0 < \text{age} <= \text{Age}_{max}$ [Gyrs]

& $\text{age} <= 13.6$ [Gyrs]

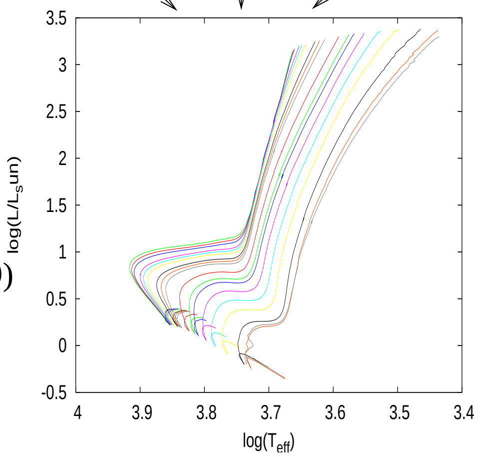
$-3.3 <= [\text{Fe/H}] <= 0.6$ [dex]

$\Gamma_1 = -0.2$ for 0.1 to 1 M_{sun}

$\Gamma_2 = -1.7$ for 1 to 10 M_{sun}

(Scalo, 1998, ASP, 241, 201)

M_1 M_2 age_s $[\text{Fe/H}]_s$



(Yi et al., 2003, APJS, 144, 259)

L_1 & T_{eff1} ---> $\log(g_1)$

L_2 & T_{eff2} ---> $\log(g_2)$

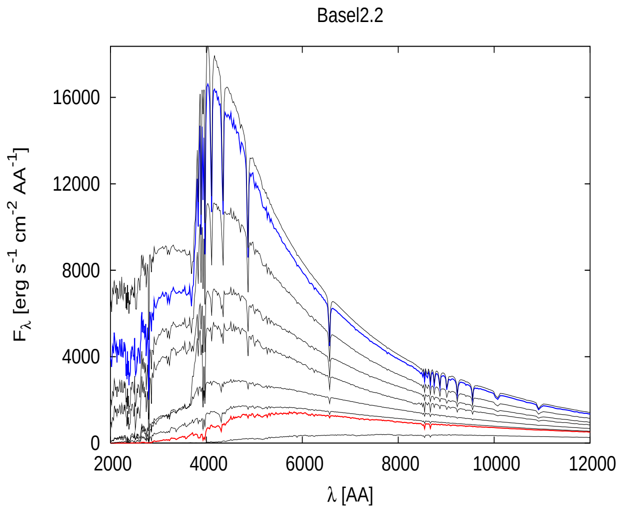
$Teff_1, \log(g_1)$, $[\text{Fe/H}]_s$

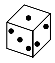
$Teff_2, \log(g_2)$, $[\text{Fe/H}]_s$

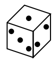
5.2 Simulation of single and double stars

$Teff_1, log(g_1)$, $[Fe/H]_s$ - - - - - →

$Teff_2, log(g_2)$, $[Fe/H]_s$ - - - - - →

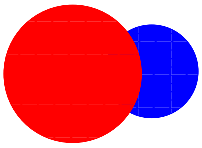


$[0, 2\pi]$  → phase ϕ

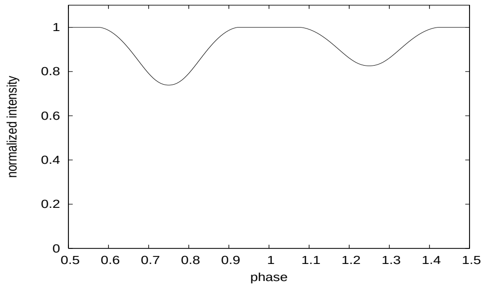
$R_1 + R_2 <= a <= a_{max}$  → semimajor axis a

$[0°, 90°]$  → inclination ρ

$A_1 = \pi R_1 * R_1$



$A_2 = \pi R_2 * R_2$



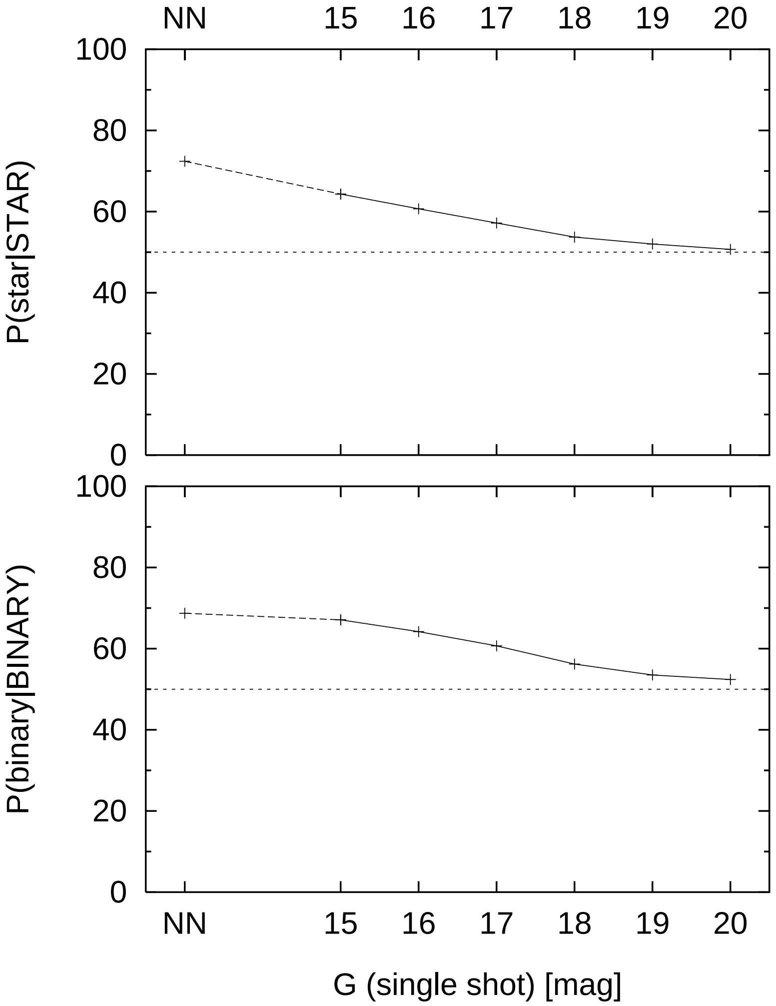
$$F_{binary} = \frac{A_1 * F_1}{A_1 + A_2} + \frac{A_2 * F_2}{A_1 + A_2}$$

$0 <= A_V <= 5$ mag  → extinction A_V

6 Classification results

photometry simulation: **photsim** (ICAP-CBJ-004,version 2) 1X system

training set : $2 \cdot 2500$, # validation set: $2 \cdot 10000$



References

- Chang, C.-C., & Li, C.J., 2001, LIBSVM: a library for support vector machines, software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm
- Cristianini, N. & Shawe-Taylor, J. 2000, An introduction to Support Vector Machines and other Kernel Based learning methods (Cambridge University Press)
- Haykin, S. 1999, Neural Networks, A comprehensive Foundation (Prentice Hall)
- Kecman, V. 2001, Learning and Soft Computing (Massachusetts Institute of technology)
- The R-project : http://www.r-project.org/ ⇒ e1071
- Vapnik, V. 1995, The nature of statistical learning theory (New York, Springer - Verlag)
- Vapnik, V. 1998, Statistical Learning Theory (New York: Wiley)