KATHOLIEKE UNIVERSITEIT
LEUVEN

Classification of Ovarian Tumors Using Bayesian Least Squares Support Vector Machines

C. Lu[1], T. Van Gestel[1], J. A. K. Suykens[1], S. Van Huffel[1], D. Timmerman[2], I. Vergote[2]

[1]Department of Electrical Engineering,
Katholieke Universiteit Leuven, Leuven, Belgium,
[2]Department of Obstetrics and Gynecology,
University Hospitals Leuven, Leuven, Belgium

AIME03, Oct 21, 2003

SISTA

Overview

- Introduction
- Data
- Bayesian least squares support vector machines (LS-SVMs) for classification
  - LS-SVM classifier
  - Bayesian evidence framework
  - Input variable Selection
- Experiments
- Conclusions



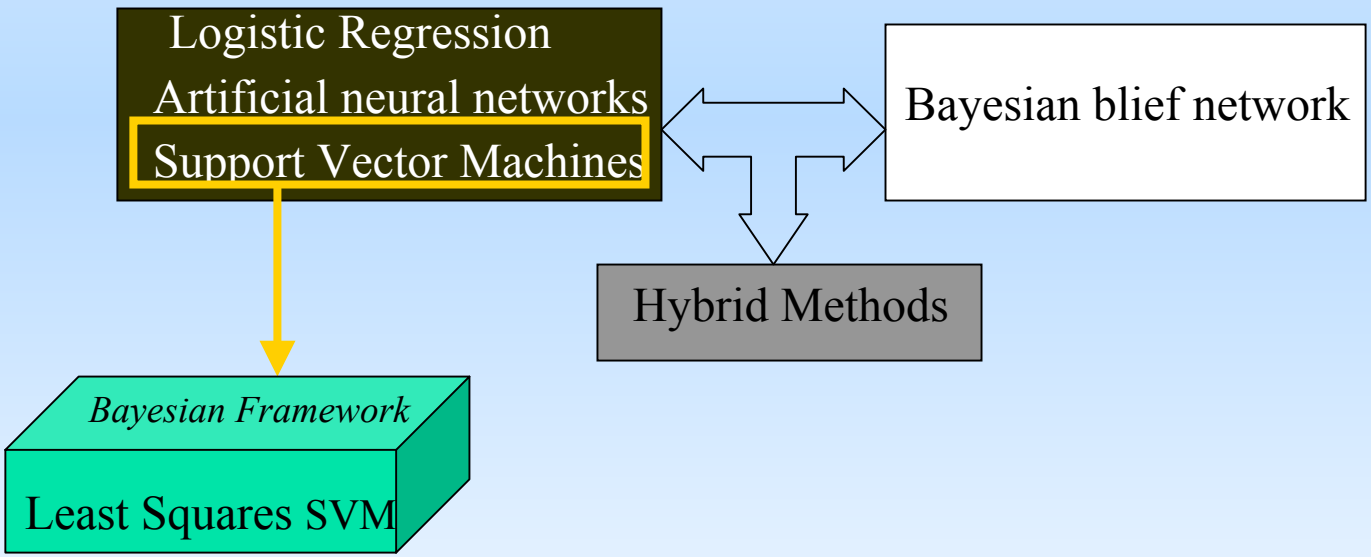
SISTA

Introduction

- Problem
  - ovarian masses: a common problem in gynecology.
    - ovarian cancer : high mortality rate
    - early detection of ovarian cancer is difficult
    - treatment and management of different types of ovarian tumors differs greatly.
  - develop a reliable diagnostic tool to preoperatively discriminate between benign and malignant tumors.
  - assist clinicians in choosing the appropriate treatment.
- Preoperative medical diagnostic methods
  - serum tumor maker: CA125 blood test
  - transvaginal ultrasonography
  - color Doppler imaging and blood flow indexing

Introduction

- Attempts to automate the diagnosis
  - Risk of malignancy Index (RMI) (Jacobs et al)
    $RMI = score_{morph} \times score_{meno} \times CA125$
  - Methematical models



Data

- Patient data collected at Univ. Hospitals Leuven, Belgium, 1994~1999
- 425 records (data with missing values were excluded), 25 features.
- 291 benign tumors, 134 (32%) malignant tumors
- Preprocessing: e.g.
  - CA_125->log,
  - Color_score {1,2,3,4} -> 3 design variables {0,1}..
- Descriptive statistics

Data

Demographic, serum marker, color Doppler imaging and morphologic variables

	Variable (symbol)	Benign	Malignant
Demographic	Age (*age*)	45.6 ± 15.2	56.9 ± 14.6
	Postmenopausal (*meno*)	31.0 %	66.0 %
Serum marker	CA 125 (log) (*l_ca125*)	3.0 ± 1.2	5.2 ± 1.5
CDI	High blood flow (*colsc3,4*)	19.0%	77.3 %
Morphologic	Abdominal fluid (*asc*)	32.7 %	67.3 %
	Bilateral mass (*bilat*)	13.3 %	39.0 %
	Unilocular cyst (*un*)	45.8 %	5.0 %
	Multiloc/solid cyst (*mulsol*)	10.7 %	36.2 %
	Solid (*sol*)	8.3 %	37.6 %
	Smooth wall (*smooth*)	56.8 %	5.7 %
	Irregular wall (*irreg*)	33.8 %	73.2 %
	Papillations (*pap*)	12.5 %	53.2 %

Data

- Patient data collected at Univ. Hospitals Leuven, Belgium, 1994~1999
- 425 records (data with missing values were excluded), 25 features.
- 291 benign tumors, 134 (32%) malignant tumors
- Preprocessing: e.g.
  - CA_125->log,
  - Color_score {1,2,3,4} -> 3 design variables {0,1}..
- Descriptive statistics
- Visualization: Biplot

Data

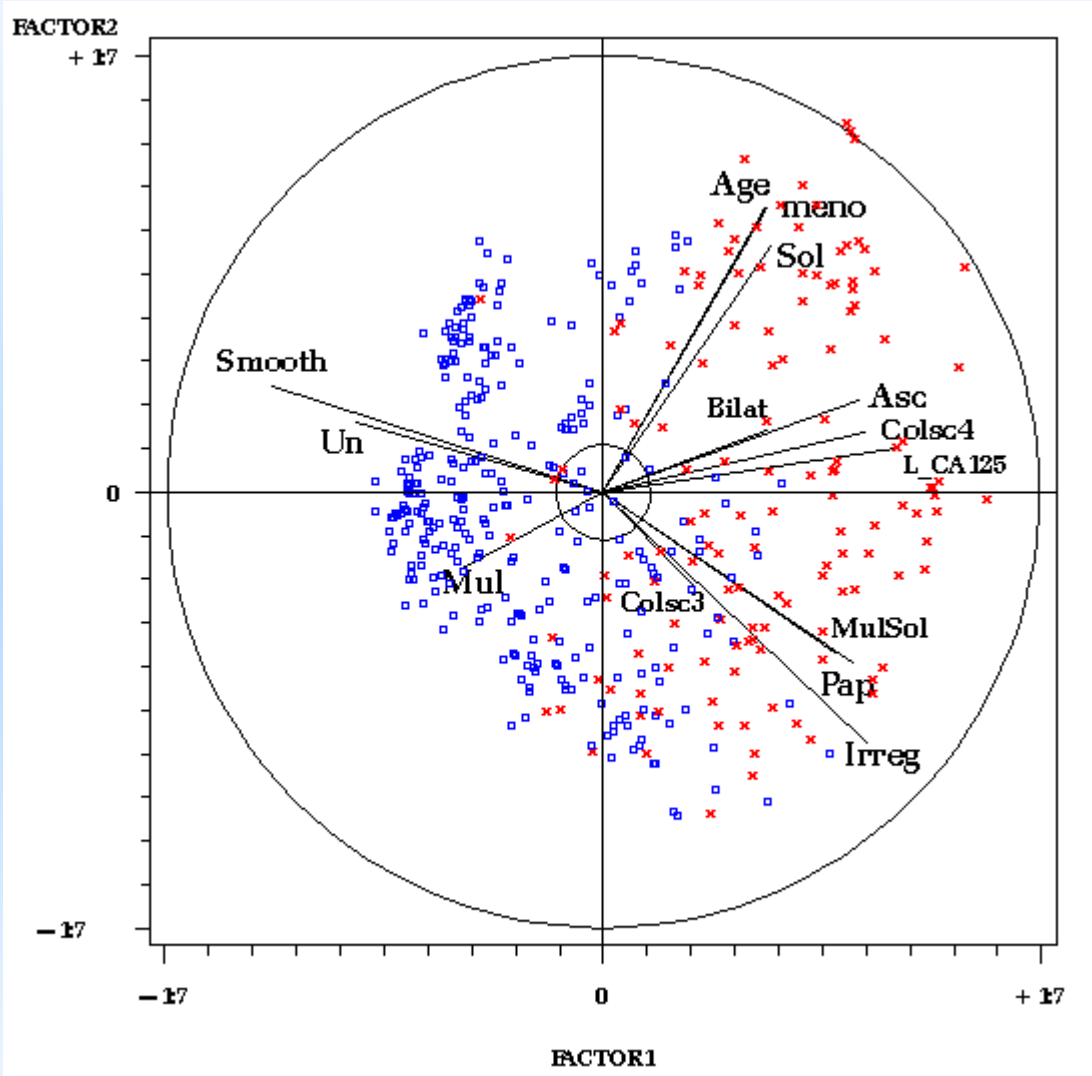


Fig. Biplot of Ovarian Tumor data.

The observations are plotted as points (o - benign, x - malignant), the variables are plotted as vectors from the origin.

- visualization of the correlation between the variables

- visualization of the relations between the variables and clusters.

Bayesian LS-SVM Classifiers

- Least square support vector machines (LS-SVM) for classification
  - Kernel based method:
    - Map the input data into a higher dimensional feature space $\mathbf{x} \rightarrow \varphi(\mathbf{x})$
    - good generalization performance, unique solution, statistical learning theory

Bayesian LS-SVM Classifiers

- LS-SVM classifier
  - Given data $D = \{(x_i, y_i)\}_{i=1,..,N}$, with binary targets $y_i = \pm 1$ (+1: malignant, -1: benign }

The following model is taken:

$$f(\mathbf{x}) = w^T \varphi(x) + b$$

$$\min_{w,b} J(w,b) = \frac{\mu}{2} w^T w + \frac{\varsigma}{2} \sum_{i=1}^{N} e_i^2,$$

$$\text{S.T. } y_i[w^T \varphi(x_i) + b] = 1 - e_i$$

$$i = 1,..., N$$

with regularizer $\gamma = \varsigma / \mu$.

Denote $\theta = [\mu, \varsigma]$

solved in dual space

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \gamma^{-1} I_N \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}$$

$$Y = [y_1,..., y_N]^T, 1_v = [1,...,1]^T, e = [e_1,..., e_N]^T,$$

$$\alpha = [\alpha_1,..., \alpha_N]^T, \Omega_{ij} = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j)$$

e.g. RBF kernel: $K(\mathbf{x}, \mathbf{z}) = \exp\{-\|\mathbf{x} - \mathbf{z}\|^2 / \sigma^2\}$

Linear kernel: $K(\mathbf{x}, \mathbf{z}) = \mathbf{z}^T \mathbf{x}$

Resulting classifier:

$$y(x) = sign[\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b]$$

Bayesian LS-SVM classifiers

- Integrate Bayesian evidence framework with LS-SVM
  - Need of probabilistic framework
    - Tune the regularization and kernel parameters
    - To judge the uncertainty in predictions, which is critical in medical environment
  - Maximizing the posterior probabilities of the models ← marginalizing over the model parameters.

Bayesian LS-SVM classifiers

- Bayesian Inference (model H : kernel parameter, e.g. σ for rbf kernels)
  - Find the maximum a posteriori estimates of model parameters w_{MP} and b_{MP}, using conventional LS-SVM training
  - The posterior probability of the parameters can be estimated via marginalization using Gaussian probability at w_{MP}, b_{MP}

Level 1: infer w,b for given H, θ

$$p\left(w,b\middle|D,\theta,H\right)=\frac{p(D\middle|w,b,\theta,H)p(w,b\middle|\theta,H)}{P(D\middle|\theta,H)}\propto \exp(-J(w,b))$$

Level 2: Infer hyperparameter θ

$$p(\theta\middle|D,H)=\frac{p(D\middle|\theta,H)p(\theta\middle|H)}{p(D\middle|H)}\propto p(D\middle|\theta,H)$$

  - Assuming a uniform prior $p(H_j)$ over all model, rank the model by the evidence $p(D|H_j)$ evaluated using Gaussian approximation.

Level 3: Compare models

$$p(H_j\middle|D)=\frac{p(D\middle|H_j)p(H_j)}{p(D)}\propto p(D\middle|H_j)$$

Bayesian LS-SVM classifiers

- Class probability for LS-SVM classifiers
  - Conditional class probabilities computed using Gaussian distributions.
  - Posterior class probability

$$p(y|x, D, \mathcal{H}) = \frac{p(y)p(x|y, D, \mathcal{H})}{\Sigma_{y'=\pm 1}\, p(y')p(x|y', D, \mathcal{H})}.$$

  - The probability of tumor being malignant ***p(y=+1|x,D,H)*** will be used for final classification (by thresholding).
  - Cases with higher uncertainty can be rejected.

Bayesian LS-SVM Classifiers

- Input variable selection
  - Select the input variable according to model evidence $p(D|H_j)$
  - Performs a forward selection (greedy search).
    - Starting from zero variables,
    - Iteratively select the variable which gives the greatest increase in the current model evidence.
    - Stop the selection when addition of any remaining variables can no longer increase the model evidence.

Experiments

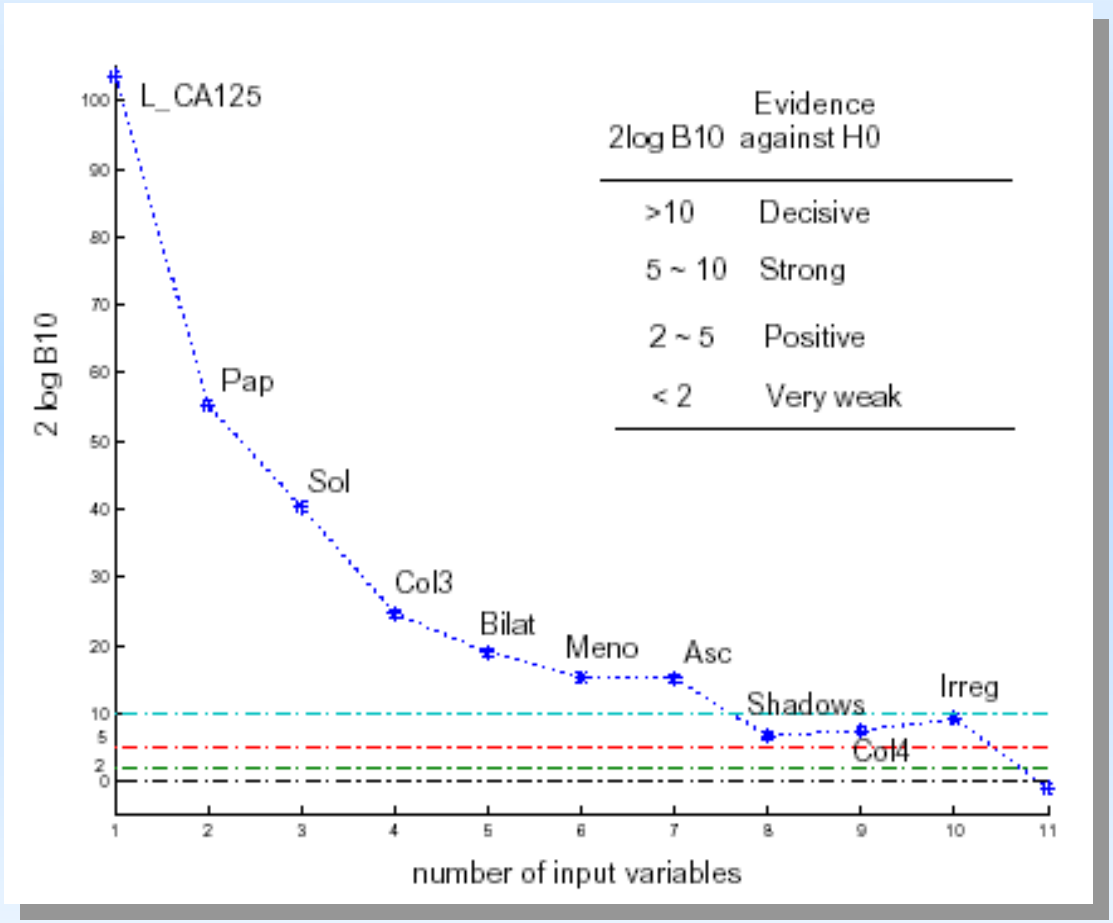
- Performance evaluation
  - **Receiver operating characteristic (ROC) analysis**
  - Goal:
    - **high sensitivity for malignancy ↔ low false positive rate.**
    - **Providing probability of malignancy for individual**
- 'Temporal' cross-validation
  - Training set : 265 data (1994~1997).
  - Test set: 160 data (1997~1999).
- Compared models
  - Bayesian LS-SVM classifiers
  - Bayesian MLPs : 10-2-1
  - Linear discriminant analysis (LDA)

Experiments

– input variable selection

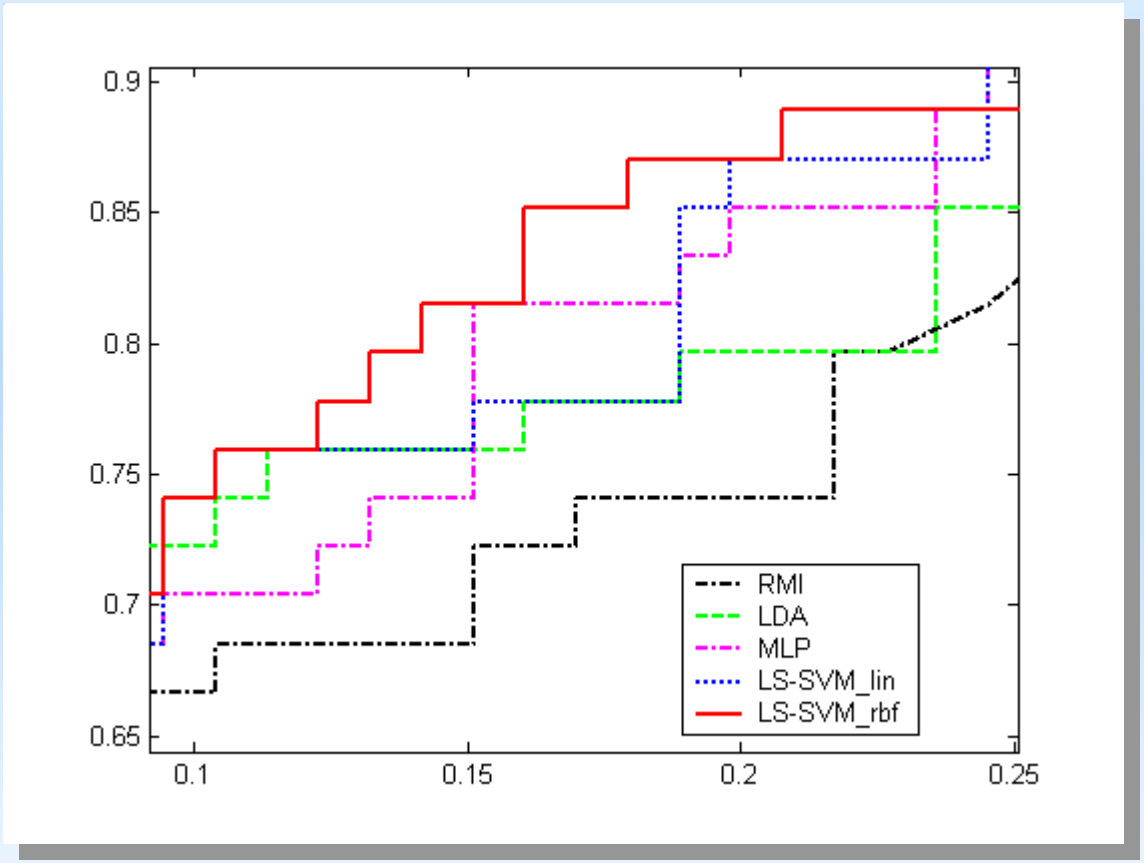
10 variables were selected based on the training set (first treated 265 patient data), using an RBF kernel.

Evolution of the model evidence



Model Evaluation

Performance on Test Set: ROC curves



Model Evaluation

Performance on Test set

MODEL TYPE	AUC	Accuracy	Sensitivity	Specificity
RMI	0.8733	78.13	74.07	80.19
		76.88	81.48	74.53
LDA	0.9034	84.38	75.93	88.68
		81.87	77.78	83.96
MLP	0.9174	82.50	77.78	84.91
		81.87	83.33	81.13
LS-SVM (LIN)	0.9141	82.50	77.78	84.91
		81.88	83.33	81.13
LS-SVM (RBF)	0.9184	84.38	77.78	87.74
		84.38	85.19	83.96

* Probability cutoff value: 0.5 and 0.3

Model Evaluation

Performance (LS-SVM_RBF) on Test set with rejection based on

$$|P(y=+1|x,\ D,\ H)\ -\ 0.5|\ \propto \text{uncertainty}$$

Reject	AUC	Accuracy	Sensitivity	Specificity
10% (16)	0.9420	88.97	83.72	91.4
5% (8)	0.9343	87.50	82.61	89.8
0% (0)	0.9184	84.38	77.78	87.74

- The rejected patients need further examination by human experts
- Posterior probability essential for medical decision making

Conclusions

- Summary
  - Within the Bayesian evidence framework, the hyperparameter tuning, input variable selection and computation of posterior class probability can be done in a unified way, without the need of selecting additional validation set.
  - The proposed forward variable selection procedure which tries to maximize the model evidence can be used to identify the subset of important variables for model building.
  - Posterior class probability enables us to assess the uncertainty in classification, important for medical decision making.
  - Bayesian LS-SVMs have the potential to give reliable preoperative prediction of malignancy of ovarian tumors.
- Future work
  - Application of the model to the multi-center data in a larger scale.
  - Possibly further subclassify the tumors.