

Kernel machine methods in genomics

Debashis Ghosh

Departments of Statistics
Penn State University
ghoshd@psu.edu

May 22, 2009 / Rao Prize Conference Seminar

Outline

- 1 Introduction
 - Background
- 2 Support vector machines
- 3 RKHS
- 4 Bayesian Approach
 - Numerical examples
- 5 Semiparametric model
- 6 SVM and splines
 - SVMs and BLUPs
- 7 Simulation studies
- 8 Conclusions

Scientific Context

- High-dimensional genomic data are now very commonplace in the medical and scientific literature
- Gene expression microarrays, single nucleotide polymorphisms, next-generation sequencing
- Scientific goals: discovering new biology as well as targets for intervention

“Large p , small n ” problems

- Scientific studies with small sample sizes and high-dimensional measurements are increasingly common
- Examples
 - Spectroscopy
 - Bioinformatics
- Two goals: clustering and classification

Support vector machines

- One technique that has received a lot of attention: support vector machines (SVMs)
- Claimed by Vapnik to “avoid overfitting”
- Applications of SVMs:
 - microarray data (Brown et al., 2000, PNAS; Mukherjee et al., Bioinformatics, 2001)
 - Protein folds (Hua and Sun, Journal of Molecular Biology, 2001)
 - PubMed Search: 1443 hits

Support vector machines

- Suppose that we have two groups of observations
- Intuition behind SVMs: find the separating hyperplane that maximizes the **margin** between two groups and perfectly classifies observations
- margin: distance between the hyperplane and points
- Sometimes to achieve perfect classification, a mapping to a higher-dimensional space is required; this is achieved through use of a kernel function.

SVM optimization problem

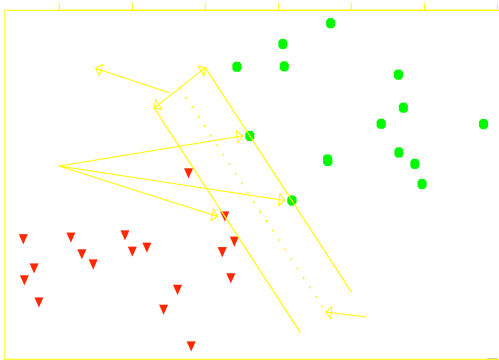
- $y = 1/-1$ (cancer/noncancer) \mathbf{z} =gene expression profile
- Use a gene expression profile to classify cancer/noncancer status.
- SVM classification problem formulation (separate case):

$$\begin{aligned} \max \quad & \frac{1}{\|\boldsymbol{\omega}\|} \\ \text{s.t.} \quad & y_i(\omega_0 + \mathbf{z}_i^T \boldsymbol{\omega}) \geq 1 \\ & i = 1, \dots, n \end{aligned}$$

- Classification rule:

$$\begin{aligned} \omega_0 + \mathbf{z}^T \boldsymbol{\omega} > 0 & \Rightarrow \hat{y} = 1 \\ \omega_0 + \mathbf{z}^T \boldsymbol{\omega} < 0 & \Rightarrow \hat{y} = -1 \end{aligned}$$

SVM: 2-D representation



Research goals

- Develop a formal inferential and statistical framework for SVMs and more general machine learning methods
- Advantages
 - 1 Probabilistic measures of predictiveness
 - 2 Avoid reliance on computationally intensive cross-validation
 - 3 Generalizations to nonlinear models

Aside: Reproducing Kernel Hilbert Spaces (RKHS)

- Let T be a general index set
- RKHS: Hilbert space of real-valued functions h on T with the property that for each $t \in T$, there exists an $M = M_t$ such that

$$|h(t)| \leq M \|h\|_H$$

- 1-1 correspondence between positive definite functions K defined on $T \times T$ with RKHS of real-valued functions on T with K as its reproducing kernel (H_K)

RKHS (cont'd.)

- If $f(\mathbf{x}) = \beta_0 + h(\mathbf{x})$, then the estimate of h is obtained by minimizing

$$g(y, f(\mathbf{x})) + \lambda \|h\|_{H_K}^2,$$

where $g(\cdot)$ is a loss function, and $\lambda > 0$ is the smoothing parameter.

- RKHS theory guarantees minimizer has form

$$f_\lambda(\mathbf{x}) = \beta_0 + \sum_{i=1}^n \beta_i K(\mathbf{x}, \mathbf{x}_i);$$

$$\|h\|_{H_K}^2 \equiv \sum_{i,j=1}^n \beta_i \beta_j K(\mathbf{x}_i, \mathbf{x}_j).$$

RKHS \rightarrow Likelihood

- Can think of objective to function for minimizing as minimizing a penalized log-likelihood

$$-g(y, f(\mathbf{x})) - \lambda \|h\|_{H_K}^2,$$

where $-g(\cdot)$ is such that $\exp[-g(\cdot)]$ is proportional to the likelihood function; λ is the smoothing parameter, and $\|h\|_{H_K}^2$ is the penalty function.

Bayesian RKHS

- First level: $p(y_i|z_i) \propto \exp\{-g(y_i|z_i)\}$, $i = 1, \dots, n$, where the y_i are conditionally independent given z_i .
- $z_i = f(\mathbf{x}_i) + \epsilon_i$, where the ϵ_i are iid $N(0, \sigma^2)$ random variables.
- **Key difference from other Bayesian methods:** introduction of ϵ_i
- $f \in H_K \Rightarrow f(\mathbf{x}_i) = \beta_0 + \sum_{j=1}^n \beta_j K(\mathbf{x}_i, \mathbf{x}_j | \theta)$
 $\mathbf{K}'_i = (1, K(\mathbf{x}_i, \mathbf{x}_1 | \theta), \dots, K(\mathbf{x}_i, \mathbf{x}_n | \theta))$, $i = 1, \dots, n$,
 $\boldsymbol{\beta} = (\beta_0, \dots, \beta_n)$

Hierarchical model

$$p(y_i|z_i) \propto \exp\{-g(y_i|z_i)\} \quad (1)$$

$$z_i|\beta, \theta, \sigma^2 \stackrel{\text{ind}}{\sim} N_1(z_i|\mathbf{K}'_i\beta, \sigma^2) \quad (2)$$

$$\beta, \sigma^2 \sim N_{n+1}(\beta|0, \sigma^2\mathbf{D}_*^{-1})\text{IG}(\sigma^2|\gamma_1, \gamma_2)$$

$$\theta \sim \prod_{q=1}^p U(a_{q1}, a_{q2})$$

$$\lambda \sim \text{Gamma}(m, c),$$

where $\mathbf{D}_* \equiv \text{Diag}(\lambda_1, \lambda, \dots, \lambda)$ is a $(n+1) \times (n+1)$ diagonal matrix

- Can extend to have multiple smoothing parameters

Candidate likelihoods

- Logistic model:

$$g(y|z) = y - \log(1 + \exp(z))$$

- SVM likelihood:

$$\begin{aligned} g(y|z) &= \frac{1}{1 + \exp(-2yz)} \quad \text{for } |z| \leq 1; \\ &= \frac{1}{1 + \exp[-y(z + \text{sgn}(z))]} \quad \text{otherwise,} \end{aligned}$$

where $\text{sgn}(u) = 1, 0$ or -1 according as u is greater than, equal or less than 0.

Hierarchical model (cont'd.)

- Choices for K
 - (i) Gaussian kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\theta}\right\}$$

- (ii) polynomial kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^\theta,$$

where $\mathbf{a} \cdot \mathbf{b}$ denotes the inner product of two vectors \mathbf{a} and \mathbf{b} .

Bayesian Analysis

- Introduction of $\epsilon_1, \dots, \epsilon_n$ facilitates use of MCMC methods
- Iterate through steps of
 - (i) update \mathbf{z} (Metropolis-Hastings);
 - (ii) update $\mathbf{K}, \beta, \sigma^2$ (Metropolis-Hastings for K , standard conjugate for β and σ^2);
 - (iii) update λ . (standard)

Prediction and Model Choice

- For a new sample with gene expression \mathbf{x}_{new} , the posterior predictive probability that its tissue type, denoted by y_{new} , is cancerous is given by

$$p(y_{new} | \mathbf{x}_{new}, \mathbf{y}) = \int p(y_{new} = 1 | \mathbf{x}_{new}, \phi) p(\phi | \mathbf{y}) d\phi \quad (3)$$

where ϕ is the vector of all the model parameters. The integral given in (3) can be approximated by its Monte Carlo estimate as

$$\sum_{i=1}^M p(y_{new} = 1 | \mathbf{x}_{new}, \phi^{(i)}) / M, \quad (4)$$

Prediction and Model Choice (cont'd.)

- To select from the different models, we will generally use misclassification error.
- If test set is available, build model on training set, use them to classify test samples.
- No test set available, use method of Gelfand (1996) for estimating cross-validation predictive density.

Simulation study

- Simulation parameters for θ and β based on posterior mean from analyzing Golub et al. (data) (38 training, 34 test)
- Generate data from the logistic and CSVM models
- Run MCMC chain for 10,000 iterations for twenty-five simulated datasets
- Look at average number of misclassifications

Simulated data

Generation Model

Model	Logistic	CSVM
Logistic	2.5	3.8
BSVM	2.7	2.2
CSVM	3.2	2.1

Misclassification errors for the three data sets studied

Method	Ripley's	Pima	Crabs	WBC
Logistic (single)	13.0(11,17)	21.4 (20.1,24.3)	5 (4,6)	12.1 (10.2,14.3)
Logistic (multiple)	9.2 (9,12)	19.4 (18.9,21.4)	2 (1,3)	8.3 (8.1,11.2)
BSVM (single)	12.4(11.1,16.8)	21 (20,23.9)	4 (2,5)	11.8 (10.1,14.4)
BSVM (multiple)	8.8(8.4,11.6)	18.9 (18.3,20.6)	1 (0,4)	8.2 (8.0,11.1)
CSVM (single)	12.7(10.8,16.7)	21.3 (19.9,24.1)	4 (2,5)	11.9 (10.0,14.5)
CSVM (multiple)	9.1(8.9,12)	19.2 (18.9,21.6)	2 (1,4)	8.3 (8.1,11.2)
RVM	9.3	19.6	2	8.8
VRVM	9.2	19.6	N/A	N/A
Jeff(Figrd)	9.6	18.5	0	8.5
Neural Networks	N/A	22.5	3	N/A
SVM*	13.2	21.2	4	12

Semiparametric Model for High-Dimensional data

- Model some covariate effects parametrically and other (e.g., gene expression) effects nonparametrically.

$$y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + h(\mathbf{z}_i) + \mathbf{e}_i$$

where $i = 1, \dots, n$, $\mathbf{x}_i = (x_{i1}, \dots, x_{iq})^T$,

$\mathbf{z}_i = (z_{i1}, \dots, z_{ip})^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)^T$,

$h(\mathbf{z}_i)$ = unknown smooth nonparametric function,

$h \in \mathcal{H}$ = some functional space, $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I})$.

- Estimate $\boldsymbol{\beta}$ and $h(\cdot)$ by minimizing the penalized RSS:

$$\sum_{i=1}^n \{y_i - (\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + h(\mathbf{z}_i))\}^2 + \lambda \|h\|_{\mathcal{H}}^2$$

- Suppose $\{\phi_i(\mathbf{z})\}_{i=1}^{\infty}$ is an orthonormal basis of \mathcal{H} . Then

$$h(\mathbf{z}) = \sum_{j=1}^{\infty} \omega_j \phi_j(\mathbf{z}) \triangleq \phi(\mathbf{z})^T \boldsymbol{\omega}$$

and $\|h\|_{\mathcal{H}}^2 = \|\boldsymbol{\omega}\|^2$ (Parseval's Theorem).

- Re-formulate the objective function (Primal Formulation):

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^n e_i^2 + \frac{1}{2} \lambda \|\boldsymbol{\omega}\|^2 \\ \text{s.t.} \quad & e_i = y_i - \{\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \phi(\mathbf{z}_i)^T \boldsymbol{\omega}\} \end{aligned}$$

- Difficulties of directly minimizing of the primal formulation:
 - Need to specify the basis $\{\phi_j(\mathbf{z})\}_{j=1}^{\infty}$ (high dimension).
 - Computation of $\hat{\boldsymbol{\omega}}$ involves inverting a high dimensional matrix.

Dual Formulation

- Introduce the Lagrangian multiplier (dual parameters) γ and form the Lagrangian function $\mathcal{L}(\omega, \beta, \mathbf{e}; \gamma) =$

$$\frac{1}{2} \sum_{i=1}^n e_i^2 + \frac{1}{2} \lambda \|\omega\|^2 - \sum_{i=1}^n \gamma_i \{ \beta_0 + \mathbf{x}_i^T \beta + \phi(\mathbf{z}_i)^T \omega + e_i - y_i \}$$

- The dimension of $\gamma = n$ (low dimension).
- The dual formulation is obtained by removing the high dimensional parameters ω and writing $\mathcal{L}(\omega, \beta, \mathbf{e}; \gamma)$ as a function of dual parameters γ and β alone.

- Optimality conditions:

$$\left\{ \begin{array}{l} \nabla_{\omega} \mathcal{L} = 0 \rightarrow \omega = \frac{1}{\lambda} \sum_{i=1}^n \gamma_i \phi(\mathbf{z}_i) \\ \frac{\partial \mathcal{L}}{\partial \mathbf{e}_i} = 0 \rightarrow \mathbf{e}_i = \gamma_i \\ \nabla_{\beta} \mathcal{L} = 0 \rightarrow \sum_{i=1}^n \gamma_i \mathbf{x}_i = 0 \\ \frac{\partial \mathcal{L}}{\partial \gamma_i} = 0 \rightarrow \mathbf{x}_i^T \beta + \phi(\mathbf{z}_i)^T \omega + \mathbf{e}_i - y_i = 0 \end{array} \right.$$

- The dual formulation is obtained by substituting $\hat{\omega}$ and $\hat{\mathbf{e}}$ into the last equation:

$$\left\{ \begin{array}{l} y_i - \mathbf{x}_i^T \beta - \frac{1}{\lambda} \sum_{i'=1}^n \gamma_{i'} \phi(\mathbf{z}_i)^T \phi(\mathbf{z}_{i'}) - \gamma_i = 0 \\ \sum_{i=1}^n \gamma_i \mathbf{x}_i = 0 \end{array} \right.$$

- Estimation in the dual formulation is low dimensional.
- The estimator $\hat{h}(\mathbf{z}) = \lambda^{-1} \sum_{i=1}^n \hat{\gamma}_i \phi(\mathbf{z})^T \phi(\mathbf{z}_i)$.

- Computation of $\hat{\gamma}$ and $\hat{h}(\mathbf{z})$ hence only requires evaluating the kernel function

$$k(\mathbf{z}, \mathbf{z}') = \langle \phi(\mathbf{z}), \phi(\mathbf{z}') \rangle = \phi(\mathbf{z})^T \phi(\mathbf{z}').$$

- If $k(\mathbf{z}, \mathbf{z}')$ is specified, no need to explicitly know the basis $\{\phi_j(\mathbf{z})\}_{j=1}^{\infty}$.

Two most popular kernel functions

- Gaussian kernel:

$$k(\mathbf{z}_i, \mathbf{z}_{i'}) = \exp\left(\frac{-\|\mathbf{z}_i - \mathbf{z}_{i'}\|^2}{\rho}\right)$$

Functional space: radial basis

- d^{th} degree polynomial kernel:

$$k(\mathbf{z}_i, \mathbf{z}_{i'}) = (\langle \mathbf{z}_i, \mathbf{z}_{i'} \rangle + c)^d$$

Functional space: d^{th} polynomial basis

- We choose to use the Gaussian kernel in our model.
 - Its form is the same as the Gaussian density function.
 - It is an infinitely smooth function.
 - The space \mathcal{H}_k it spans is dense in L_2 .

General Support Vector Machine (SVM)

- The general form of SVM:

$$\min \sum_{i=1}^n L\{y_i, f(\mathbf{z}_i)\} + \lambda \|\boldsymbol{\omega}\|^2,$$

where $f(\mathbf{z}) = \omega_0 + h(\mathbf{z}) = \omega_0 + \boldsymbol{\phi}(\mathbf{z})^T \boldsymbol{\omega}$.

- Examples of the error (loss) function $L\{y_i, f(\mathbf{z}_i)\}$:
 - SVM classification with Hinge loss function

$$L\{y, f(\mathbf{z}_i)\} = \max \{0, 1 - yf(\mathbf{z}_i)\}$$

- Least squares SVM regression

$$L\{y, f(\mathbf{z}_i)\} = \{y - f(\mathbf{z}_i)\}^2$$

Semiparametric Model: Revisit

- Recall the semiparametric model:

$$y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + h(\mathbf{z}_i) + \mathbf{e}_i, \quad i = 1, \dots, n$$

where $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I})$ and $h(\cdot) \in \mathcal{H}_k$.

- Write $h(\mathbf{z}_i) = \boldsymbol{\phi}(\mathbf{z}_i)^T \boldsymbol{\omega}$.
- Modified Primal Formulation of LS SVM

$$-\frac{1}{2\sigma^2} \sum_{i=1}^n \{y_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta} - \boldsymbol{\phi}(\mathbf{z}_i)^T \boldsymbol{\omega}\}^2 - \frac{1}{2\tau} \boldsymbol{\omega}^T \boldsymbol{\omega},$$

where $\tau = 1/\lambda$.

- Modified dual formulation:

$$\begin{bmatrix} \mathbf{0} & \mathbf{X}^T \\ \mathbf{X} & \tau\mathbf{K} + \sigma^2\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}$$

where the $n \times n$ matrix $\mathbf{K} = \mathbf{K}(\rho) = \{k(\mathbf{z}_i, \mathbf{z}_{i'})\}$.

- Given (τ, ρ, σ^2) , we have

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}^T(\tau\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{X}]^{-1}\mathbf{X}^T(\tau\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y}$$

$$\hat{\boldsymbol{\gamma}} = (\tau\mathbf{K} + \sigma^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$\hat{\mathbf{h}} = \tau\mathbf{K}\hat{\boldsymbol{\gamma}} = \tau\mathbf{K}(\tau\mathbf{K} + \sigma^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

- Estimation of (τ, ρ, σ^2) is challenging, e.g.,
 - Estimation of the smoothing parameter $\tau = 1/\lambda$ is often done using CV or requires a separate validation set.
 - No systematic way to estimate ρ .

Connection of LS SVM and Linear Mixed Model

- Key Message: LS SVM (semi)non-parametric regression can be fitted using linear mixed models by PROC MIXED.
- The forms of the SVM estimators $\hat{\beta}$ and $\hat{\mathbf{h}}$ are identical to the BLUP estimators under the linear mixed effects model

$$\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{b} + \boldsymbol{\epsilon} \quad (5)$$

where the $n \times 1$ random effect vector $\mathbf{b} \sim N\{\mathbf{0}, \tau \mathbf{K}(\rho)\}$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

- The LS SVM estimator $\hat{\mathbf{h}}(\mathbf{z}) = \{\hat{h}(\mathbf{z}_1), \dots, \hat{h}(\mathbf{z}_n)\}^T$ is the BLUP:

$$\hat{\mathbf{h}} = \hat{\mathbf{b}}$$

- Unified estimation of (τ, ρ, σ^2) by REML

Test for the Nonparametric Function

- Hypothesis of interest:

$$H_0 : h(\mathbf{z}) = 0 \quad \text{vs} \quad H_1 : h(\mathbf{z}) \in \mathcal{H}_k.$$

- This hypothesis is equivalent to

$$H_0 : \tau = 0 \quad \text{vs} \quad H_1 : \tau > 0.$$

- The null hypothesis is on the boundary of the parameter space and \mathbf{K} is not a block diagonal matrix. So the LRT and Wald tests are not $0.5\chi_0^2 + 0.5\chi_1^2$.
- \mathbf{K} involves an unknown scale parameter ρ for the Gaussian kernel, which is unestimable under H_0 .

Score Test When ρ Is Given

- Score statistic:

$$U_{\tau}(\mathbf{y}; \beta, \sigma^2, \rho) = \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{K}(\rho) (\mathbf{y} - \mathbf{X}\beta).$$

- Under H_0 , $U_{\tau}(\mathbf{y}; \beta, \sigma^2)$ follows a mixture of χ_1^2 's.
- Use the Satterthwaite method to approximate the mixture of the chi-squares by $\kappa\chi_{\nu}^2$ where κ and ν are calculated by matching the first two moments of $U_{\tau}(\cdot)$ and $\kappa\chi_{\nu}^2$.
- The test statistic $S = \frac{U_{\tau}(\mathbf{y}; \beta, \sigma^2)}{\kappa}$ is approx χ_{ν}^2 under H_0 .
- Later, employ Davies's (1978, 1987) sup approximation: Treat the test statistic as a χ^2 process and calculate the p-value of its sup $P(\sup_{\rho} S > c)$.

Simulation Study

- Setting 1:
 - Sample size $n=300$ and $p=15$ genes.
 - Model:

$$y = x_1 + h(z_1, \dots, z_{15}) + e \quad (6)$$

where $e \sim N(0, 1)$ and $h(\cdot)$ has a complicated form.

- Use the 15 right genes and fit (2) using LS SVM via linear mixed model with the Gaussian kernel.
- 300 runs.

Simulation results of point estimation

Model	Parameter Estimates				Reg ression of h on \hat{h}		
	$\hat{\beta}_1$	$\hat{\sigma}^2$	$\hat{\tau}$	$\hat{\rho}$	int	slope	R^2
$y = \mathbf{x}^T \boldsymbol{\beta} + e$	2.07	70.03					
$y = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{z}^T \boldsymbol{\omega} + e$	2.02	47.36					
$\rho = 64$	1.07	1.77	403.51	64	-0.05	1.01	0.99
$\rho = 100$	1.08	2.68	651.06	100	-0.14	1.02	0.98
$\rho = 225$	1.13	5.18	1562.75	225	-0.26	1.04	0.96
$\rho = 400$	1.18	7.55	2482.36	400	-0.29	1.05	0.94
$\rho = 625$	1.21	8.70	4472.89	625	-0.29	1.05	0.92
$\rho \text{ est}' d$	1.07	1.07	333.67	43.30	0.05	1.00	0.99

- Setting 2:
 - Sample size $n=150$ and $p=5$ active genes.
 - Model: $y = x_1 + h(z_1, \dots, z_5) + e$ where $e \sim N(0, 1)$
 $h(\mathbf{z}) = 10 \cos(z_1) + 3z_2^2 - 2\sqrt{|z_3|}z_4 + 6 \sin(z_5) + e$
 - Use the 5 right genes and 5 junk genes and fit
 $y = x_1 + h(z_1, \dots, z_{10}) + e$ using LS SVM via the linear mixed model.
- 1000 runs.

Simulation results of point estimates

Model	Parameter Estimates				Regression of h on \hat{h}		
	$\hat{\beta}_1$	$\hat{\sigma}^2$	$\hat{\tau}$	$\hat{\rho}$	int	slope	R^2
$y = \mathbf{x}^T \boldsymbol{\beta} + e$	2.23	44.37					
$y = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{z}^T \boldsymbol{\omega} + e$	2.26	31.91					
$\rho = 9$	1.45	0.14	855.80	9	0.77	1.02	0.96
$\rho = 25$	1.14	0.57	708.45	25	0.21	1.01	0.98
$\rho = 64$	1.07	1.34	741.75	64	0.02	1.01	0.98
$\rho = 100$	1.07	1.38	741.75	100	0.02	1.01	0.98
$\rho = 225$	1.09	2.67	3674.85	225	-0.09	1.03	0.97
ρ <i>est'</i> d	1.09	0.85	575.81	36.72	0.12	1.01	0.98

Simulation Results for the Nonparametric Function

- Model: $y = x + h(\mathbf{z}) + e$ where $e \sim N(0, 1)$, $h(\mathbf{z}) = \alpha h_0(\mathbf{z})$, and takes a similar form to $h_0(\mathbf{z})$ to setting 2.
- $H_0 : h(\mathbf{z}) = 0$ ($\tau = 0$) vs $H_1 : h(\mathbf{z}) \neq 0$ ($\tau > 0$).
- Study the size and power of the score test.
- Set $n = 50$ and $\alpha = 0, 0.2, 0.4, 0.6, 0.8, 1.0$.
- Number of simulations = 2000 (size), 1000 (power).

Simulation Results of the Score Test

Scale ρ	Size	Power				
	$\alpha = 0$	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1.0$
1	0.057	0.081	0.212	0.400	0.619	0.770
4	0.048	0.152	0.455	0.796	0.945	0.992
25	0.053	0.164	0.427	0.673	0.855	0.943
64	0.042	0.141	0.394	0.600	0.786	0.900
100	0.044	0.139	0.378	0.609	0.747	0.879
200	0.044	0.145	0.367	0.578	0.738	0.866

Conclusions for Bayesian approach

- Development of a general hierarchical model for classification that extends previous work
- Development of MCMC procedures for Bayesian analysis of the model
- Rigorous comparisons with existing classification methods shows that it is quite competitive
- Multiple smoothing parameters helps (automatic relevance determinance)

SVM Conclusions

- We develop a semiparatic regression model where clinical covariates are modeled parametrically and high-dimensional gene expressions are modeled nonparametrically using LS SVM.
- The LS SVM can be fit using a linear mixed model in a unified framework, where the regression coefficients and the nonparametric function can be estimated by the BLUPs and the smoothing parameters and the kernel scale parameter can be estimated using REML.
- Obvious Extensions to GLMMs

Kernels

- Kernels here are presented in a mathematical manner
- However, kernels can be used to encode biological information
- Intuition, need a notion of “dissimilarity” between objects \Rightarrow a kernel matrix \mathbf{K} (like what was used in last part of the talk) can be developed.
- Main conditions: \mathbf{K} is symmetric and positive definite
- Link between distance and \mathbf{K} was given by Gower

Acknowledgments

- Bayesian SVM: Bani Mallick, Texas A&M and Malay Ghosh, University of Florida
- SVM/BLUP: Dawei Liu, University of Iowa and Xihong Lin, Harvard University
- Publications:
 - Mallick, B., Ghosh, D. and Ghosh, M. (2005). Bayesian kernel-based classification of microarray data. *Journal of the Royal Statistical Society Series B* **2**, 219 – 234.
 - Liu, D., Lin, X. and Ghosh, D. (2007). Semiparametric regression of multi-dimensional genetic pathway data: least squares kernel machines and linear mixed models. *Biometrics* **63**, 1079 – 1088.