2009 Fifth International Joint Conference on INC, IMS and IDC

Comparison on confidence bands of decision boundary between SVM and Logistic Regression

Xing Wang[1] Xin Wang[2]
School of statistics
Renmin University of China[1,2]
Beijing, 100872, China
wangxingscy@gmail.com[1] woxyc@126.com[2]

Zhaonan Sun[3]
Department of statistics
Purdue University[3]
West Lafayette, IN 47907-2066, USA
sunz@stat.purdue.edu

***Abstract*—Support Vector Machine (SVM) and Logistic Regression (LR) are two popular classification models. The main purpose of a classification algorithm is to figure out the estimator for the decision boundary. In this paper, we considered confidence bands of decision boundary generated from SVM and LR. Confidence bands of decision boundary are estimated through bootstrap methods. We compared the confidence band estimator of SVM with the estimator of the conventional LR. Our main result is that sample size of the observations makes effect on the stability of both SVM and LR, sample size ratio, central location and covariance matrix of the data bring less effects on the stability of SVM than that of LR.**

Keywords- confidence bands of decision boundary; Support Vector Machine; Logistic Regression; bootstrap

I. INTRODUCTION

Support Vector Machine (SVM) and Logistic Regression (LR) are two popular methods in classification literature. SVM provides a technique to build the optimal stable decision boundary based on the idea of minimizing empirical structure risk, and it solves the optimization problem of maximizing the margin between the two classes on training data. LR, on the other hand, maximizes the log-likelihood function. In terms of loss function, SVM applies the hinge loss whereas LR chooses the logit loss. Both of the two loss functions are typical for stabilizing models. However, Hastie Trevor (2005) suggested that different loss functions lead to different properties of the classifiers.

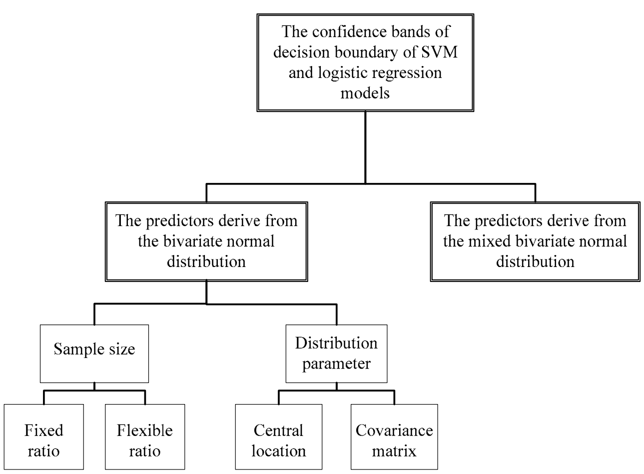

Figure 1. The research frame work

The concept of confidence band is traditional in statistics, it is used to characterize the stability of the estimator. In this paper, we discuss the properties of SVM and LR in terms of confidence bands of decision boundary and research on the effect of sample size and data distributions. This paper generally consists of three sections: first, we briefly review the principles of SVM and LR; second, we propose a method to compute the confidence band of decision boundary; finally, we design five kinds of experiments for both SVM and LR to compare how confidence bands vary with some factors changing. Fig.1 shows the research framework.

II. PRELIMINARIES

Suppose the objects we're interested in falls into two categories $Y \in \{-1,+1\}$, input space is p dimensional, i.e. $X \in \Re^p$, and let $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ be n independent random samples. Our goal is to model a hyperplane $\omega^T x + \omega_0 = 0$ as the decision boundary. For any $x_0 \in R^p$, the classification rule is:

$$G(x_0) = sign(\omega^{\mathrm{T}} x_0 + \omega_0). \tag{1}$$

A. Support Vector Machine

In this paper, we only discuss two cases: the two classes are linear separable and the two classes partially overlap in feature space.

1) The two classes are linear separable: the optimization problem is:

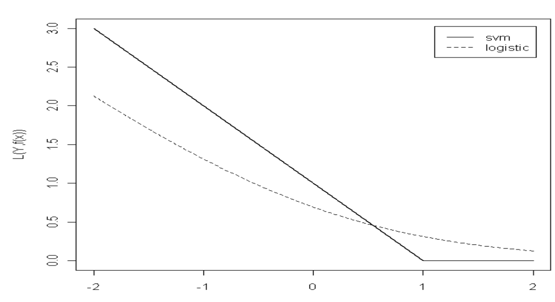

Figure 2. The solid line is hinge loss. The dotted line is logit loss

978-0-7695-3769-6/09 $26.00 © 2009 IEEE
DOI 10.1109/NCM.2009.281

$$\min_{\omega,\omega_0} \frac{1}{2}\|\omega\|^2$$

$$subject \quad to \quad y_i\left(\omega^{\mathrm{T}} x_i + \omega_0\right) \geq 1, \quad i = 1,2\ldots\ldots,n \cdot$$

By inducting Lagrangian multiplier α_i , $i = 1,2,\cdots,n$, where $\alpha_i \geq 0$, the linear programming can be transformed as a dual quadratic programming problem:

$$\max D = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j \quad , \tag{2}$$

$$\text{subjec to } \alpha_i \geq 0, \quad \sum_i \alpha_i y_i = 0,$$

where $i, j = 1,2,\cdots,n$.

According to the dual programming problem, we can get the solution

$$\omega = \sum_i \alpha_i y_i x_i \quad . \tag{3}$$

Those observations with $\alpha_i \neq 0$ are called support vectors. The decision boundary of SVM is uniquely characterized by support vectors.

2) The two classes may not be separable by linear boundary (overlap): define the slack variable ξ_i*, modify the constraint as:*

$$x_i \cdot \omega \geq +1 - \xi_i, \text{ for } y_i = +1,$$

$$x_j \cdot \omega \leq -1 + \xi_i, \text{ for } y_j = -1 .$$

where $\xi_i \geq 0$, $\forall i$, those points with $\xi_i = 0$ are on the correct side. The linear optimization problem can be written as:

$$\min L = \frac{1}{2}\|\omega\|^2 + C\sum \xi_i , \tag{4}$$

$$\text{subject to } y_i\left(\omega^T \cdot x^i + b\right) - 1 \geq \xi_i, \ \xi_i \geq 0 .$$

The dual programming is:

$$\max D = \sum_i \alpha_i^* - \frac{1}{2}\sum_i \sum_j \alpha_i^* \alpha_j^* y_i y_j x_i^T x_j , \tag{5}$$

$$\text{subject to } C \geq \alpha_i^* \geq 0, \quad \sum_i \alpha_i^* y_i = 0 \cdot$$

The solution is:

$$\omega = \sum_i \alpha_i^* y_i x_i \cdot \tag{6}$$

$L(Y, f(X)) = \max(0,1 - Y \cdot f(X))$ is the loss function of SVM, as shown in Fig. 2.

B. *Logistic Regression*

Given the data $\{(x_1, y_1),\cdots,(x_n, y_n)\}$, where n is the sample size, let $\pi(X) = P(Y = 1 \mid X)$, the logistics regression can be expressed as:

$$\pi(X) = \frac{\exp\left(\omega^T X + \omega_0\right)}{1 + \exp\left(\omega^T X + \omega_0\right)} \cdot \tag{7}$$

Define the decision function:

$$f(x) = \log\left(\frac{\pi(X)}{1 - \pi(X)}\right) = \omega^T X + \omega_0 = 0 \cdot \tag{8}$$

The classification rule is:

$$y = +1 , \ f(x) > 0 ,$$

$$y = -1 , \ f(x) < 0 .$$

$L(Y, f(X)) = \log(1 + \exp(Y \cdot f(X)))$ is the loss function of the LR, which is showed in Fig. 2.

C. *Confidence band of decision boundary*

Brand, Pinnock & Jackson (1973) suggested an approach to create confidence interval for LR coefficient in one dimensional space; Hauck (1983) proposed a method to create confidence band for LR coefficient in higher dimensional space, both of two methods made the interval estimation for prediction probability based on confidence interval of model coefficients. Jiang, Zhang and Cai (2008) discussed the confidence interval for prediction errors of SVM. However, few researches explore the properties of confidence band of decision boundary itself, which can directly reflect the stability of the classifier.

Russell (2005) proposed a method to build confidence bands on regression model, based on his idea, we proposed a method to compute confidence band of decision boundary in classification models.

In classification models, the decision boundary is

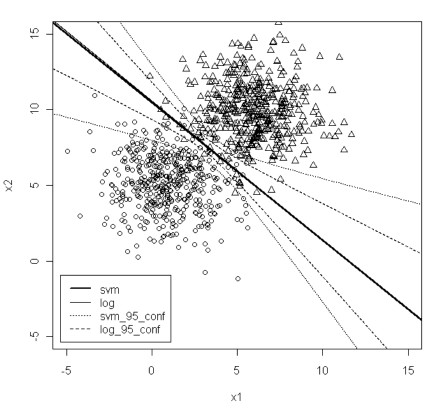

Figure 3. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples in class -1 are from $N_1(1,5,4,4,0)$; samples in class +1 are from $N_2(6,10,4,4,0)$. n_1=n_2=450.

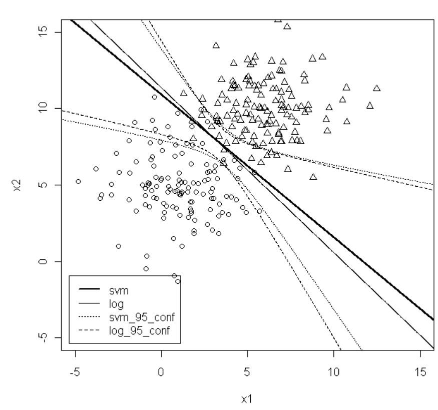

Figure 4. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples in class -1 are from $N_1(1,5,4,4,0)$, samples in class +1 are from $N_2(6,10,4,4,0)$. n_1=n_2=12

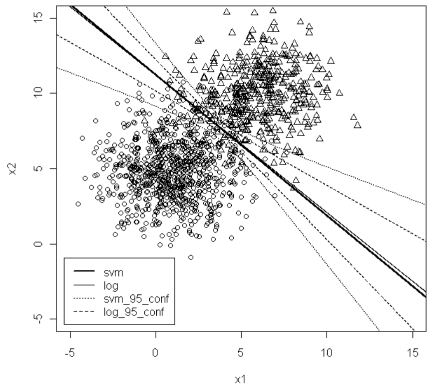

Figure 5. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples in class -1 are from $N_1(1,5,4,4,0)$, n_1=900; samples in class +1 are from $N_2(6,10,4,4,0)$, n_2=450.

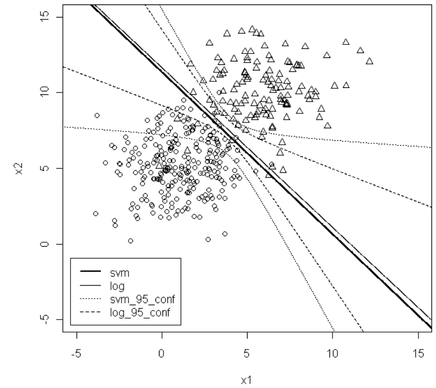

Figure 6. he dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples in class -1 are from $N_1(1,5,4,4,0)$, n_1=256; samples in class +1 are from $N_2(6,10,4,4,0)$, n_2=128.

usually expressed as $f(x)=\beta^T X+\beta_0=0$. In this paper, we only consider the cases of $X\in\Re^2$, thus the decision boundary can be indicated as $f(x)=\beta_1 x_1+\beta_2 x_2+\beta_0=0$, which is a straight line in two-dimension space.

After simple transformation, we get

$$x_2=\theta_0+\theta_1 x_1, \tag{9}$$

where

$$\theta_0=-\frac{\beta_0}{\beta_2},\ \theta_1=-\frac{\beta_1}{\beta_2}. \tag{10}$$

The decision boundary can be expressed as a function of x , $\eta(x)=\theta_0+\theta_1 x$, it is quite similar to the form of regression function. We can get the estimator of decision boundary from sample data: $\hat{f}(x)=\hat{\beta}_1 x_1+\hat{\beta}_2 x_2+\hat{\beta}_0=0$, then we derive the $\hat{\eta}(x)=\hat{\theta}_0+\hat{\theta}_1 x$, where $\hat{\theta}_0=-\hat{\beta}_0/\hat{\beta}_2$, $\hat{\theta}_1=-\hat{\beta}_1/\hat{\beta}_2$. We labeled $\theta=(\theta_0,\theta_1)$, $\hat{\theta}=(\hat{\theta}_0,\hat{\theta}_1)$.

For a parameter θ, let $\hat{\theta}_{ML}$ be its maximum likelihood estimation, it retains the property of the asymptotic normality. When $n\to\infty$,

$$\left(\hat{\theta}_{ML}-\theta\right)^T V^{-1}(\theta)\left(\hat{\theta}_{ML}-\theta\right)\sim\chi_p^2,$$

where $V(\theta)^{-1}$ is the inverse matrix of the information matrix. Usually θ is unknown, we use the approximation $V\left(\hat{\theta}_{ML}\right)^{-1}$ instead:

$$\left(\hat{\theta}_{ML}-\theta\right)^T V^{-1}\left(\hat{\theta}_{ML}\right)^{-1}\left(\hat{\theta}_{ML}-\theta\right)\sim\chi_p^2.$$

Then we get

$$R(\alpha)=P\left(\left(\hat{\theta}_{ML}-\theta\right)^T V\left(\hat{\theta}_{ML}\right)^{-1}\left(\hat{\theta}_{ML}-\theta\right)\le\chi_\alpha{}^2(p)\right)=1-\alpha, \tag{11}$$

where $\chi_p^2(\alpha)$ is the upper α chi-squared quantile for p degrees of freedom.

We are actually interested in the confidence band of function $\eta(x)=\theta_0+\theta_1 x$.

Suppose we can find two estimators $\theta_L(X)$ and $\theta_U(X)$ defined as follows:

$$\theta_L(X)=\arg\left[\min_{\theta\in R(\alpha)}\eta(X;\theta)\right], \tag{12}$$

$$\theta_U(X)=\arg\left[\max_{\theta\in R(\alpha)}\eta(X;\theta)\right]. \tag{13}$$

Now we solve $\eta(X;\theta_L(X))$ and $\eta(X;\theta_U(X))$ by using Taylor expansion about θ, then we minimize and maximize the linear approximation subject to (11), the problem is thus:

$$\text{min/ max：}\ \eta(X;\theta)=\eta\left(X;\hat{\theta}_{ML}\right)+\left(\frac{\partial f(X;\theta)}{\partial\theta}\right)^T\left(\theta-\hat{\theta}_{ML}\right), \tag{14}$$

subject to $\left(\hat{\theta}_{ML}-\theta\right)^T V\left(\hat{\theta}_{ML}\right)^{-1}\left(\hat{\theta}_{ML}-\theta\right)\le\chi_\alpha{}^2(p)$.

The solution is

$$\eta(X;\theta_L(X))=\eta\left(X;\hat{\theta}_{ML}\right)-h\left(X;\hat{\theta}_{ML}\right), \tag{15}$$

$$\eta(X;\theta_U(X))=\eta\left(X;\hat{\theta}_{ML}\right)+h\left(X;\hat{\theta}_{ML}\right), \tag{16}$$

where

$$h\left(X;\hat{\theta}_{ML}\right)=\sqrt{\chi_p{}^2(\alpha)\left(\frac{\partial(\eta(X;\theta))}{\partial\theta}\right)^{\mathrm{T}} V\left(\hat{\theta}_{ML}\right)\left(\frac{\partial(\eta(X;\theta))}{\partial\theta}\right)}. \tag{17}$$

Note that $\hat{\theta}$ of decision boundary is not MLE, thus $D=\left(\hat{\theta}-\theta\right)^T V\left(\hat{\theta}\right)^{-1}\left(\hat{\theta}-\theta\right)$ is not asymptotically chi-square

TABLE I. PARAMETER ESTIMATION & 95% WCB

Model		PE[a]		95% WCB	
		$\hat{\theta}_0$	$\hat{\theta}_1$	***Class -1***	***Class +1***
$n_1=n_2$=450 $N_1(1,5,4,4,0)$	SVM	10.48	-0.92	1.77	1.78
$N_2(6,10,4,4,0)$	LR	10.75	-0.99	1.16	1.18
$n_1=n_2$=128 $N_1(1,5,4,4,0)$	SVM	11.29	-1.08	3.49	3.63
$N_2(6,10,4,4,0)$	LR	11.72	-1.21	2.76	2.90
n_1=900 n_2=450 $N_1(1,5,4,4,0)$	SVM	11.22	-0.94	1.85	1.83
$N_2(6,10,4,4,0)$	LR	11.19	-0.91	0.95	0.96
n_1=256 n_2=128 $N_1(1,5,4,4,0)$	SVM	11.22	-1.07	4.00	3.66
$N_2(6,10,4,4,0)$	LR	11.65	-1.06	2.03	1.98

a. Parameter Estimator

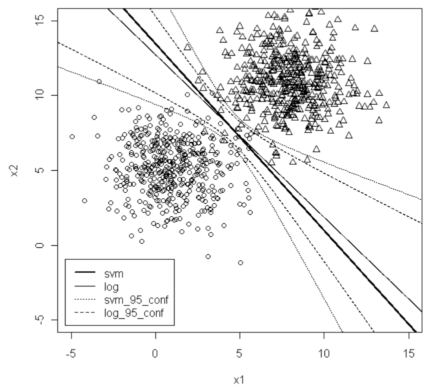

Figure 7. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples in class -1 the same as that in Fig. 3. In class +1, we made all first component of x_i is that of Fig.3 plus 2, and all second component of x_i is that of Fig.3 plus 1, n_1=n_2=450.

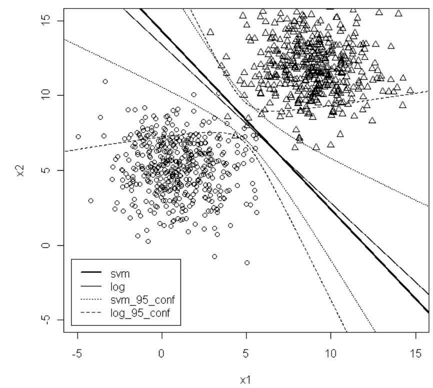

Figure 8. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples in class -1 the same as that in Fig. 3. In class +1, we made all first component of x_i is that of Fig.3 plus 3, and all second component of x_i is that of Fig.3 plus 2, n_1=n_2=450.

distributed. Here we use bootstrap method to get an empirical estimation instead. Resample B times, we get:

$$\mathrm{D}^* = \left(\theta^* - \hat{\theta}\right)^T V^{*-1}\left(\hat{\theta}\right)\left(\theta^* - \hat{\theta}\right), \quad (18)$$

where θ^* is the estimation of θ obtained from each bootstrap sample, $\hat{\theta}$ is the estimation of θ obtained from training data, $V^*\left(\hat{\theta}\right)$ is the bootstrap variance. Sort D^* as $\left\{\mathrm{D}_1^*, \cdots, \mathrm{D}_B^*\right\}$,we use $\mathrm{D}^*_{\mathrm{B(1-\alpha)}}$ instead of $\chi_p^2(\alpha)$, and use $V^*\left(\hat{\theta}\right)$ instead of $V\left(\hat{\theta}_{ML}\right)$.

The confidence band of decision boundary is:

$$\eta^*\left(X;\theta_L(X)\right) = \eta\left(X;\hat{\theta}\right) - h^*\left(X;\hat{\theta}\right), \quad (19)$$

$$\eta^*\left(X;\theta_U(X)\right) = \eta\left(X;\hat{\theta}\right) + h^*\left(X;\hat{\theta}\right), \quad (20)$$

where

$$h^*\left(X;\hat{\theta}_{ML}\right) = \sqrt{D^*_{B(1-\alpha)}\left(\frac{\partial(\eta(X;\theta))}{\partial\theta}\right)^{\mathrm{T}} V^*\left(\hat{\theta}\right)\left(\frac{\partial(\eta(X;\theta))}{\partial\theta}\right)}. \quad (21)$$

Thus, we have the approximation

$$P\left(\eta^*\left(X;\theta_L(X)\right) \le \eta(X:\theta) \le \eta^*\left(X;\theta_U(X)\right)\right) = (1-\alpha)\% . \quad (22)$$

III. Results on Simulation Experiments

We generate various simulated data and compare the confidence bands of decision boundary of SVM and LR. Here we define the width of 95% confidence band (95% WCB) to make the results numerically comparable.

Suppose we have data $\left\{(x_1, y_1), \cdots, (x_n, y_n)\right\}$, where $x_i \in \Re^2$, $y_i \in \{-1, +1\}$, the predictors in class -1 and $+1$ are from two different distribution respectively. We set $\sigma^2_{-1,1}$ and $\sigma^2_{-1,2}$ as the diagonal elements of sample covariance matrix of class -1, $\sigma^2_{+1,1}$ and $\sigma^2_{+1,2}$ as the diagonal elements of sample covariance matrix of class $+1$. We define $\sigma^* = \max\left(\sigma_{-1,1}, \sigma_{-1,2}, \sigma_{+1,1}, \sigma_{+1,2}\right)$

In class -1, we set $P_{-1}\left(x_{-1,1}, x_{-1,2}\right)$ as the central point, where $x_{-1,1}$ is the mean of the first component of x_i , $i = 1,2,\cdots,n$, and $x_{-1,2}$ is the mean of the second component of x_i , $i = 1,2,\cdots,n$. Let NL_{-1} denote the normal line of the decision boundary that passes through $P_{-1}\left(x_{-1,1}, x_{-1,2}\right)$. According to (19), (20) and (21), one boundary line of the 95% confidence band is a curve and on the side of class -1. Thus we can find two points on this boundary line, where the Euclidean distance from the two points to NL_{-1} is $3\sigma^*$ respectively. Then we calculate the Euclidean distance from each point to the decision boundary, label as $DL_{-1,1}$ and $DL_{-1,2}$. The 95% WCB of class -1 is defined as the mean of $DL_{-1,1}$ and $DL_{-1,2}$.

In class $+1$, the 95% WCB is similarly defined.

A. Predictors in class -1 and +1 are generated from two different bivariate normal distribution respectively

In this section, in class $Y = -1$, n_1 is sample size, X is from $N_1\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho_1\right)$; in class $Y = +1$, n_2 is sample size, X is from $N_2\left(\mu_3, \mu_4, \sigma_3^2, \sigma_4^2, \rho_2\right)$. We resample $n_1/2$ and $n_2/2$ data in class $Y = -1$ and $Y = +1$ respectively, B=1000.

1) The effect of the sample size on the confidence bands of decision boundary.

When the sample size is large (at least 30^2 in two dimensional space), whether the sample ratio of two classes is 1:1 or 2:1, the 95% WCB of SVM is wider than that of LR, as shown in Fig. 3 and Fig. 5.

In Fig. 3 and Fig. 4, the sample ration of two classes is

TABLE II. Parameter Estimation & 95% WCB

Model		PE		95% WCB	
		$\hat{\theta}_0$	$\hat{\theta}_1$	*Class -1*	*Class +1*
n_1=n_2=450 N_1(1,5,4,4,0) N_2(8,11,4,4,0)	SVM	13.08	-1.17	2.07	2.04
	LR	13.20	-1.21	2.05	2.07
n_1=n_2=450 N_1(1,5,4,4,0) N_2(9,12,4,4,0)	SVM	14.48	-1.21	2.27	2.17
	LR	13.81	-1.11	5.09	5.67

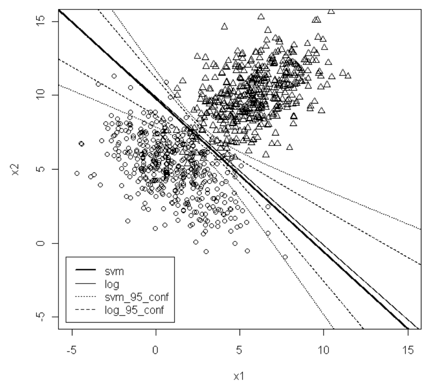

Figure 9. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples in class -1 are from $N_1(1,5,4,4,-0.5)$; samples in class +1 are from $N_2(6,10,4,4,0.5)$, $n_1=n_2=450$

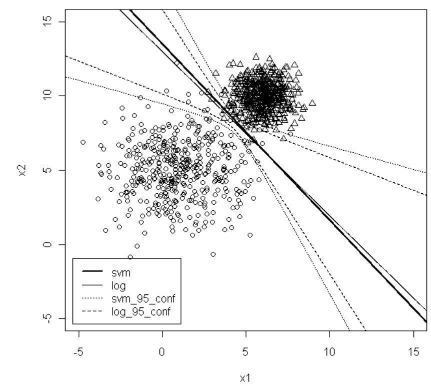

Figure 10. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples in class -1 are from $N_1(1,5,4,4,0)$; samples in class +1 are from $N_2(6,10,1,1,0)$, $n_1=n_2=450$

1:1, we see that as sample size gets smaller, 95% WCB of two models both become larger and get closer. If we add more samples in either class, which we change ratio of two classes from 1:1 to 2:1, whether the samples size is large or small, the 95% WCB of LR become narrower, but that of SVM doesn't change much, as shown in Fig. 5 and Fig.6. The results of parameter estimation and 95% WCB are shown in Table I.

Briefly, larger sample size makes the WCB narrower, and the confidence band of decision boundary of LR is more sensitive to sample size.

2) The effect of the central location of predictors on the confidence bands of decision boundary

Here we define the central location as the sample mean. Samples of class +1 in Fig. 7 and Fig. 8 are moving result of that in Fig. 3, where we increase the Euclidean distances of central location between class +1 and class -1. According to Table II, we can find that the central location has a little effect on the 95% WCB of SVM. But the confidence band of LR obviously become wider when the distance of central location gets larger; they finally become approximately 2 times wider than those of SVM, as shown in Fig. 8.

3) The effect of the covariance matrix of predictors on the confidence bands of decision boundary

Compared with Fig. 3, the covariance matrix of class -1 and class +1 in Fig. 9 both change, and are different from each other, but the Table III shows parameter estimations and 95% WCB are similar with those in Table I.

Fig. 10 illustrates a result quite different from that of Fig. 3, where we only change the variance of class +1 and 95% WCB of both models become more than 30% larger.

B. Prediciors in class -1 and + 1 are generated from two different mixed bivariate normal distribution respectively

Fig.11 shows that when data is from mixed normal distribution, the 95% WCB of LR is much wider than that of SVM. Then we use the sample mean and sample variance to create normal distribution settings, as shown in Fig.12. We find that 95% WCB of LR changes a lot, and it becomes narrower than that of SVM, whereas 95%WCB of SVM is quite stable in the two cases.

In sum, compared with LR, the confidence band of decision boundary of SVM less depends on the assumption of distribution.

IV. MAIN CONCLUSIONS

We designed various simulation experiments to explore the factors that have effects on the confidence band of decision boundary of SVM and that of LR.

The width of confidence band provides the insight of stability and reliability for classifiers. The narrower the confidence band, more stable and reliable the decision boundary is. Thus, we can evaluate the performance of SVM and LR in terms of comparison on confidence bands of decision boundary.

The simulation experiments show that sample size has a notable impact on the confidence bands of decision boundary of both SVM and LR.

With the condition of normal distribution and large sample, the confidence band of decision boundary of LR is narrower than that of SVM. However, the confidence band of decision boundary of LR is more sensitive to the factors including sample ratio, central location and covariance matrix of the data. On the contrary, in various cases, the width of confidence band of SVM doesn't change much, this can be explained as that SVM is generally a more stable classification model and needs less distribution assumptions.

TABLE III. PARAMETER ESTIMATION & 95% WCB

Model		**PE**		**95% WCB**	
		$\hat{\theta}_0$	$\hat{\theta}_1$	***Class -1***	***Class +1***
$n_1=n_2=450$ $N_1(1,5,4,4,-0.5)$	SVM	9.77	-1.04	1.95	1.94
$N_2(6,10,4,4,0.5)$	LR	9.89	-1.00	1.17	1.14
$n_1=n_2=450$ $N_1(1,5,4,4,0)$	SVM	13.49	-1.18	2.74	2.71
$N_2(6,10,1,1,0)$	LR	12.98	-1.11	2.02	2.06

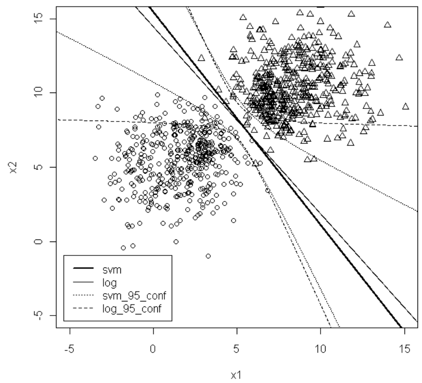

Figure 11. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. Samples of two classes are from mixed bivariate normal distribution. We indicate the bivariate normal distribution as $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. In class -1, 300 samples are from *N*(1,5,4,4,0), 150 samples are from *N*(3,6.5,1,1,0). In class +1, 300 samples are from *N*(9,11,4,4,0), 150 samples are from *N*(7,9.5,1,1,0). When resampling, we random select 225 samples in two classes respectively, *B*=1000.

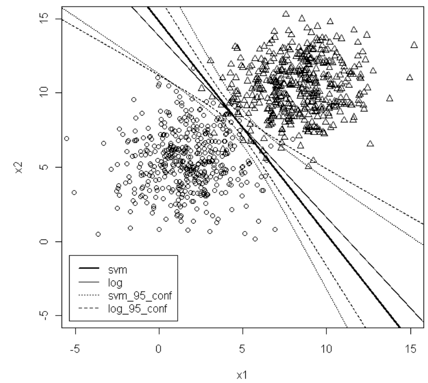

Figure 12. The dark solid line is the decision bounary of SVM, the dotted line is its 95% confidence band; The light solid line is the decision bounary of LR, the dashed line is its 95% confidence band. In class -1, 450 samples are from bivariate normal distribution, where the mean is the sample mean of class -1 in Fig. 11, covariance matrix is the sample covariance matrix of class -1 in Fig. 11. In class +1, 450 samples are from bivariate normal distribution, where the mean is the sample mean of class +1 in Fig. 11, covariance matrix is the sample covariance matrix of class +1 in Fig. 11. When resampling, we random select 225 samples in two classes respectively, *B*=1000

V. DISCUSSION

When considering the effects of central location and covariance of matrix, it is possible that distribution of overlapped points effectively cause the change of 95% WCB, for different central location and covariance matrix lead to different type of overlapping. This needs further study.

In this paper, we consider two dimensional problems, where decision boundary is a straight line in feature space. In higher dimensional problems, the decision boundary becomes a hyperplane and its confidence band will be more complicated. Thus, some conclusions we get in this paper may not be simply expanded to the higher dimensional space.

Another significant feature of SVM is that it solves the nonlinear problem by introducing kernel function. In order to compare SVM with LR, we only discuss the cases that build the linear decision boundary. Further study may explore the confidence band of the nonlinear decision boundary.

Last, our conclusions are not based on real datasets but only the simulation datasets. And some theoretical proof can be helpful to make clear of the principles of our conclusions.

ACKNOWLEDGEMENTS

This research was supported by funding No. 22382050 from Renmin University of China, and project No. 08JJD910248 from Center of Applied Statistics, Ministry of Education of the Peoples' Republic of China.

TABLE IV. PARAMETER ESTIMATION & 95% WCB

Model		PE		95% WCB	
		$\hat{\theta}_0$	$\hat{\theta}_1$	*Class -1*	*Class +1*
Mixed bivariate normal distribution	SVM	15.45	-1.43	2.20	2.16
	LR	14.45	-1.26	4.65	4.60
Empirical bivariate normal distribution	SVM	14.82	-1.43	1.65	1.64
	LR	13.81	-1.22	1.53	1.62

REFERENCES

[1] X. Wang, Non-parametric Statistics (Chinese), Beijing: TsingHua University Press and Springer, 2009.

[2] N. Cristianini and J. S. Taylor, An introduction to support vector machines: and other kernel-based learning methods, Cambridge: Cambridge University Press, 2004.

[3] J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, No. 2, June. 1998, pp. 121-167.

[4] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: data mining, inference and prediction, New York: Springer, 2005.

[5] R. J. Brand, D.E. Pinnock, K.L. Jackson, "Large Sample Confidence Bands for the Logistic Response Curve and Its Inverse," The American Statistician, vol. 27, No. 4, Oct. 1973, pp. 157-160.

[6] W. W. Hauck, "A Note on Confidence Bands for the Logistic Regression Curve," The American Statistician, vol. 37, No. 2, May 1983, pp. 158-160.

[7] B. Jiang, X. Zhang, T. Cai, "Estimation Confidence Interval for Prediction Errors of Support Vector Machine Classifiers," Journal of Machine Learning Research, vol. 9, Mar. 2008, pp. 521-540.

[8] C. H. Russell, "Bootstrapping Simultaneous Confidence Bands," Proc. 2005 Winter Simulation Conference, Dec. 2005, pp.-, doi: 10.1109/WSC.2005.1574257.

[9] D.C. Bowden, "Simultaneous Confidence Bands for Linear Regression Models," Journal of American Statistical Association, vol. 65, No. 329, Mar. 1970, pp. 413-421.

[10] J. Shao, and D. Tu, The Jackknife and Bootstrap, 2rd ed., New York: Springer-Verlag., 1996.

[11] V. Vapnik, Estimation of Dependences Based on Empirical Data, New York: Springer-Verlag, Reprint, 2006.

[12] R. A. Fisher, "The use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, No. 7, Sep. 1936, pp. 179-188.

[13] L. Wasserman, All of Nonparametric Statistics, Springer, New York: Springer, 2005.