PROJECTION METHOD FOR SUPPORT VECTOR MACHINES WITH INDEFINITE KERNELS

Hao Jiang1, Wai-Ki Ching2, Yushan Qiu2, Xiaoqing Cheng2

1Department of Mathematics, School of Information, Renmin University of China, No.59 Zhong Guan Cun Street, Hai Dian District, Beijing, 100872
2Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, University of Hong Kong, Pokfulam Road, Hong Kong

jianghai@ruc.edu.cn, wching@hku.hk, u3001131@hku.hk, mechxqiaofen@gmail.com

Keywords: Semi-Definite Kernel, LogDet Divergence, Projection Method, Support Vector Machines (SVMs)

Abstract

In this paper, we tackle with indefinite kernels by introducing projection matrix to formulate a positive semi-definite kernel. The projection matrix has a nice property of sharing the same set of eigenvectors with the original kernel. The proposed model can be regarded as a generalized version of spectrum method (denoising method and flipping method) by varying parameter $\lambda$. The problem of selecting optimal $\lambda$ for optimizing the prediction performance is also considered. Using the Bregman matrix divergence theory, one can realize kernel learning by using unconstrained optimization. And our suggested $\lambda$ in projection matrix helps to exhibit optimal performance for different values of $\lambda$.

1 Introduction

Kernel methods are essentially a class of algorithms in machine learning. They work by embedding data instances into a feature space $F$. Due to their good performance in processing complicated data, kernel methods have gained increasing popularity in computational biology [1]. The best known member is Support Vector Machine (SVM) [2]. Positive Semi-Definite (PSD) is an important property [3] of a kernel matrix in SVM. It is required to ensure the existence of a Reproducing Kernel Hilbert Space (RKHS) where a convex optimization formulation can be deduced to yield an optimal solution. However, in practice, similarity matrices may not process such PSD property. For example, some popular functions evaluating pair-wise similarity between DNA and protein sequences induce indefinite kernel matrices [4, 5]. The histogram intersection kernel and its generalized version are not always positive definite [6]. There are also other indefinite kernels like sigmoid kernels [7], hyperbolic tangent kernels [8] etc. However, to our best knowledge, it is still not very clear how to take advantage of them effectively in the SVM framework. Therefore, it is a challenging optimization problem for training SVMs with indefinite kernels because in this learning scenario, convex solutions for standard SVMs are not valid.

A number of methods have been developed in dealing with indefinite kernels. Previous studies attempt to address the problem by altering the spectrum of an indefinite kernel matrix in order to create a PSD one. The denoising method is one of the representatives which deems all negative eigenvalues as noise and replaces them with zero [9]. Another representative is the flipping method which flips the sign of negative eigenvalues so as to form a PSD kernel matrix [10]. The diffusion method takes the data distribution into account by replacing the eigenvalues by an exponential form [11], and the shifting method shifts eigenvalues to ensure the nonnegativity of all the eigenvalues [12]. In [7], Lin and Lin developed a method for finding stationary points in non-convex dual formulation of SVMs with sigmoid kernels. In [13], learning with indefinite kernel is interpreted as a minimization problem in a pseudo-Euclidean space. In [14], a max-min optimization problem is proposed to find a proxy kernel for the the indefinite kernel under consideration. Based on confidence function, Guo and Schuurmans [15] suggested a simple generalization of SVMs. Kernel principal component analysis [16] is used as a kernel transformation method to deal with indefinite kernels.

We propose a projection method to transform an indefinite kernel to a positive semi-definite one. The projection method has its flexibility, by varying its parameters, different methods such as denoising method and flipping method can be obtained. The projection matrix also shows nice mathematical properties and one can demonstrate the connection between projection method and spectrum method. The rest of the paper is organized as follows. Section 2 describes the projection method and gives a theoretical illustration. The optimal $\lambda$ in projection matrix under unconstrained optimization framework is also presented. Section 3 presents experimental results on some real world datasets. Concluding remarks are then given in the last section.

2 Methods

Given a set of labeled patterns $(\mathbf{X}_i, y_i)_{i=1}^n$, where $\mathbf{X}_i \in \mathbb{R}^p$, and $y_i \in \{-1, 1\}$. Assume $\chi$ is the input space,
the function
\[ k : \chi \times \chi \to \mathbb{R} \]
is regarded as a kernel or a kernel function. Mercer’s theorem states that a valid kernel should be positive semi-definite.

2.1 Related Works

Kernel transformation strategy gains increasing importance in dealing with non positive semi-definite kernel. If we assume \( K \) to be the originally constructed kernel matrix, i.e., it is symmetric but not positive semi-definite, we have the following decomposition:

\[ K = U \cdot P \cdot U' \]

where \( P \) is a diagonal matrix \( (P = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}) \) whose diagonal entries are not fully non negative, \( U \) is a unitary matrix, the \( i \)th column of which is the corresponding eigenvector for \( i \)th eigenvalue in \( P \). Here \( U' \) is the transpose of the matrix \( U \).

Representative methods in kernel transformation include:

**Denoising Method** [9]

\[ \hat{K} = U \cdot \hat{P} \cdot U', \text{ where} \]

\[ \hat{P} = \text{diag}\{\max(0, \lambda_1), \max(0, \lambda_2), \ldots, \max(0, \lambda_n)\} \]

**Flipping Method** [10]

\[ \hat{K} = U \cdot \hat{P} \cdot U', \text{ where} \]

\[ \hat{P} = \text{diag}\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|\} \]

**Diffusion Method** [11]

\[ \hat{K} = U \cdot \hat{P} \cdot U', \text{ where} \]

\[ \hat{P} = \text{diag}\{e^{\alpha \lambda_1}, e^{\alpha \lambda_2}, \ldots, e^{\alpha \lambda_n}\} \]

and \( \alpha \) is a parameter.

**Shifting Method** [12]

\[ \hat{K} = U \cdot \hat{P} \cdot U', \text{ where} \]

\[ \hat{P} = \text{diag}\{\alpha + \lambda_1, \alpha + \lambda_2, \ldots, \alpha + \lambda_n\} \]

where \( \alpha \) is a shifting parameter to ensure the non-negative of new eigenvalues.

2.2 Projection Method

Now in the following we present our projection method for turning an indefinite kernel to a positive semi-definite one.

**Theorem:**

Let \( K \) be an \( n \times n \) symmetric matrix which is not positive semi-definite, then there exists an \( n \times m \) \( (m < n) \) matrix \( B \) satisfying \( B'B = I_m \) such that \( (I_n - \lambda BB')K \) is a positive semi-definite kernel where \( \lambda \geq 1 \) is a regularization parameter.

**Proof:**

Since \( K \) is symmetric, there exists a unitary matrix \( P \) and a diagonal matrix \( D \) such that

\[ P'DP = K \]

where \( P = [\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_m] \) and \( D = \text{diag}\{d_1, d_2, \ldots, d_n\} \) is a diagonal matrix with the diagonal elements \( d_i \). Without loss of generality, we may assume all the eigenvalues are sorted in ascending order. If we assume negative inertia index is \( m \), then \( [\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_m] \) are the corresponding eigenvectors for \( \{d_1, d_2, \ldots, d_m\} \), if the positive inertia index is \( l \), then \( [\vec{p}_{n-l+1}, \vec{p}_{n-l+2}, \ldots, \vec{p}_n] \) are corresponding eigenvectors for \( \{d_{n-l+1}, d_{n-l+2}, \ldots, d_n\} \).

We construct \( B = [\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_m] \) and \( B'B = I_m \).

Since \( \vec{p}_i, i = 1, 2, \ldots, n \) are orthogonal eigenvectors, we have

\[ B'B = \left[ \begin{array}{c} \vec{p}_1 \\ \vec{p}_2 \\ \vdots \\ \vec{p}_m \end{array} \right] \left[ \begin{array}{c} \vec{p}_1 \\ \vec{p}_2 \\ \vdots \\ \vec{p}_m \end{array} \right]' = \left[ \begin{array}{cccc} (\vec{p}_1)' \vec{p}_1 & (\vec{p}_1)' \vec{p}_2 & \cdots & (\vec{p}_1)' \vec{p}_m \\ (\vec{p}_2)' \vec{p}_1 & (\vec{p}_2)' \vec{p}_2 & \cdots & (\vec{p}_2)' \vec{p}_m \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{p}_m)' \vec{p}_1 & (\vec{p}_m)' \vec{p}_2 & \cdots & (\vec{p}_m)' \vec{p}_m \end{array} \right] = I_m. \]

Besides, for \( (I - \lambda BB') \), it has \( (1 - \lambda) \) and \( 1 \) as its eigenvalues, the multiplicities of them are \( m \) and \( n - m \), respectively (see the following proposition). Furthermore, the eigenvectors of \( (I - \lambda BB') \) are exactly the same as the indefinite kernel \( K \). Hence,

\[ (I_n - \lambda BB')K = P' \text{diag}\{1, 1, \ldots, 1\} P \]

or

\[ P' \text{diag}\{(1 - \lambda)d_1, (1 - \lambda)d_2, \ldots, (1 - \lambda)d_m, 0, \ldots, 0, d_{n-l+1}, \ldots, d_n\} P. \]

Since \( \lambda \geq 1 \), we have

\[ (1 - \lambda)d_i \geq 0, \quad 1 \leq i \leq m. \]

This guarantees the positive semi-definiteness property of the matrix \((I_n - \lambda BB')K\).

**Proposition:**

The matrix \((I_n - \lambda BB')\) is symmetric and it has \((1 - \lambda)\) and 1 as its eigenvalues, the multiplicities for them are \( m \) and \( n - m \), respectively.

**Proof:**

- We first show that \( I_n - \lambda BB' \) is symmetric. We note that

\[ (I_n - \lambda BB')' = I_n - \lambda BB' \]

then it means that \((I_n - BB')\) is a symmetric matrix.
Hence the matrix \((I_n - \lambda BB')\) has 1 - \lambda and 1 as its eigenvalues, and their multiplicities are \(m\) and \(n - m\), respectively.

In the following, we state the relation between projection method and spectrum method. By varying the parameters \(\lambda\), one can obtain both denoising method and flipping method.

- When \(\lambda = 1\), \(I_n - \lambda BB' = I_n - BB'\),
  \[(I_n - BB')K = P'\text{diag}\{0, 0, \ldots, 0, d_{n-t+1}, \ldots, d_n\}P.\]

  This is exactly the denoising method for transforming indefinite kernel to a positive semi-definite one.

- When \(\lambda = 2\), \(I_n - \lambda BB' = I_n - 2BB'\),
  \[(I_n - 2BB')K = P'\text{diag}\{-d_1, -d_2, \ldots, -d_m, 0, \ldots, 0, 1, \ldots, 1\}P.\]

  This is exactly the flipping method for transforming an indefinite kernel to a positive semi-definite one.

**Optimal \(\lambda\) Determination**

Note that \(\lambda\) is a parameter embedded in Projection Method, in order to study the optimal choice of \(\lambda\), we begin with the definition of Bregman matrix divergence [17].

### Definition:

Bregman Matrix Divergences

Assume \(\phi(K)\) is a strictly convex differentiable function of \(K\). The Bregman Matrix Divergence of \(K\) is defined as follows:

\[
D_\phi(K, K_0) = \phi(K) - \phi(K_0) - \text{tr}(\nabla \phi(K_0)'(K - K_0)).
\]

Here \(\text{tr}(K)\) means the trace of matrix \(K\). There exist a number of matrix divergences [17].

- Frobenius Divergence: If \(\phi(K) = \|K\|_F^2\) then
  \[D_\phi(K, K_0) = \|K - K_0\|_F^2.\]

- Von Neumann Divergence: If \(\phi(K) = \text{tr}(K \log K - K)\) then
  \[D_\phi(K, K_0) = \text{tr}(K \log K - K \log K_0 - K_0).\]

- LogDet Divergence: If \(\phi(K) = -\log \det(K)\) then
  \[D_\phi(K, K_0) = \text{tr}(KK_0^{-1} - \log \det(KK_0^{-1}) - n).\]

As we have previously mentioned, \(K\) is our original kernel constructed. The projection method on \(K\) results in a new kernel \((I_n - \lambda BB')K\). In particular, when \(\lambda = 1\), the new kernel shows the same power as denoising method. When \(\lambda = 2\), the new kernel is identical with flipping method. Our aim here is to determine what value of \(\lambda\) which can yield a superior kernel that can guarantee excellent performance when \(\lambda > 0\) is considered.

In [18], the authors have proposed a framework of kernel learning by unconstrained optimization. We can similarly reformulate the problem as a kernel learning one. In this work, we consider LogDet Divergence[18]. Regard \(K\) as the original kernel, the new kernel can be written as \(\tilde{K} = K - \lambda BB'K\). Then the optimal \(\lambda\) can be obtained by minimizing \(D_\phi(\tilde{K}, K)\) where \(\tilde{K}\) is the optimal kernel that we want to construct and it is close to original kernel \(K\) in terms of divergence where

\[
K = P'DP = \sum_{i=1}^{n} d_i p_i^T p_i^T.
\]

We have

\[
\tilde{K} = \sum_{i=1}^{m} (1 - \lambda) d_i \bar{p}_i \bar{p}_i^T + \sum_{i=m+1}^{n} d_i \bar{p}_i \bar{p}_i^T.
\]

Then the minimization problem becomes:

\[
\min\limits_{\lambda} D_\phi(\tilde{K}, K) = \left\{ \begin{array}{ll}
\min\limits_{\lambda} & \sum_{i=1}^{m} \left[\log(1 - \lambda)d_i + (1 - \lambda)d_i \text{tr}(K^{-1}\bar{p}_i^T \bar{p}_i)\right] \\
& - \sum_{i=m+1}^{n} \left[\log d_i + d_i \text{tr}(K^{-1}\bar{p}_i^T \bar{p}_i)\right]
\end{array} \right\}
\]
It can be verified that the above minimization problem is equivalent to the followings:

\[
\begin{align*}
\min_{\lambda} & \quad -\sum_{i=1}^{m} [\log(1 - \lambda) d_i + (1 - \lambda) d_i \text{tr}(K^{-1} \hat{p}_i \hat{p}_i^t)] \\
\text{subject to} & \quad 1 - \lambda \sum_{i=1}^{m} d_i \text{tr}(K^{-1} \hat{p}_i \hat{p}_i^t) = 0
\end{align*}
\]

We can determine the optimal $\lambda$ as the following formula.

\[
1 - \lambda \sum_{i=1}^{m} d_i \text{tr}(K^{-1} \hat{p}_i \hat{p}_i^t) = 0
\]

where $X_{ji}$ is the $i$th element of $X_j$.

When $\alpha = \beta$, it can be proved that the kernel is a positive semi-definite matrix. In the experiment stage, we will use different values of $\alpha, \beta \in \{1, 2, 3\}$ to test the performance of our proposed projection method.

### 3 Results

#### 3.1 Data Sets

In this section, we evaluate the performance of our projected kernel and the original kernel constructed from real world data sets which can be obtained from libsvm data sets at http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets. There are 3 life science data sets involved for experimental purpose. For sonar data set, there are 208 data instances, 60 are positive, the remaining 111 are negative. Live disorder data set has 345 data instances, 6 features. 145 are positive and 200 are negative. In Breast Cancer data set, there are 683 data instances in total with 10 features. 444 are negative and 239 are positive. Detailed information on the data sets can be found in Table 1.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Number of Instances</th>
<th>Number of Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonar</td>
<td>208</td>
<td>60</td>
</tr>
<tr>
<td>Live Disorder</td>
<td>345</td>
<td>6</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>683</td>
<td>10</td>
</tr>
</tbody>
</table>

#### 3.2 Performance Evaluation

The performance of the model is measured through 5-fold cross-validation, we use the Area Under Curve (AUC) for accuracy evaluation which is commonly used for model evaluation. It can be calculated as the area under the ROC curve. We run 10 times 5-fold cross-validation to measure the averaged AUC values for the considered methods. Here we use Generalized Histogram Intersection (GHI) kernel [6] for illustration purpose. It has been successfully applied in image classification problem but it has not yet been used in glycan classification or cancer prediction. The kernel function is defined as follows:

\[
k(X_j, X_k) = \sum_{i=1}^{p} \min(|X_{ji}|^\alpha, |X_{ki}|^\beta)
\]

#### 3.3 Experiments on Life Science Data

Table 2 to 4 summarize the performance for Projection Method and GHI Kernel SVM for three different life science data sets. When $\alpha = \beta \in \{1, 2, 3\}$, Projection Method has identical performance as GHI Kernel. GHI kernel is indefinite when $\alpha \neq \beta$ for the 3 considered data sets (data not shown). When values of $\alpha$ and $\beta$ differ from each other, we can observe different performance in the two considered methods. For example, Table 2 reports the performance for sonar data set. When $\alpha = 1, \beta = 2$, the averaged AUC values for Projection Method is 81.47% with standard deviation 0.99%, the averaged AUC values for GHI Method is 53.42% with standard deviation 4.94%. When $\alpha = 1, \beta = 3$, the averaged AUC values for Projection Method is 84.02% with standard deviation 1.19%, the averaged AUC values for GHI Method is 54.10% with standard deviation 4.92%. When $\alpha = 2, \beta = 3$, the averaged AUC values for Projection Method is 84.31% with standard deviation 1.56%, the averaged AUC values for GHI Method is 83.06% with standard deviation 2.04%.

For live disorder data set, results show that when $\alpha \neq \beta$, Projection method gives much better performance than GHI kernel method.

For breast cancer data set, results show that when $\alpha \neq \beta$, Projection method gives much better performance than GHI kernel method except when $\alpha = 2, \beta = 3$. It is not surprising as GHI kernel sometimes with indefinite kernel can also yield good performance. Our Projection method can still demonstrate equivalent performance with GHI kernel in this case.

#### 3.4 Experiments on Optimal $\lambda$ for Projection Method

Projection method can transform an indefinite kernel into a positive semi-definite one when $\lambda \geq 1$. In the previous section, the optimal $\lambda$ for Projection Method

---

2015 ISORA 978-1-78561-086-8 ©2015 IET

140

Luoyang, China, August 21–24, 2015
Table 3: Classification Accuracies (%) of Projection Method and Generalized Histogram Intersection Kernel Using Live Disorder Data.

<table>
<thead>
<tr>
<th>PARAMETERS</th>
<th>Projection Method</th>
<th>GHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 1, \beta = 1$</td>
<td>76.00± 0.80</td>
<td>75.99± 0.81</td>
</tr>
<tr>
<td>$\alpha = 1, \beta = 2$</td>
<td>74.39± 0.92</td>
<td>47.48± 3.59</td>
</tr>
<tr>
<td>$\alpha = 1, \beta = 3$</td>
<td>71.67± 1.47</td>
<td>56.49± 3.02</td>
</tr>
<tr>
<td>$\alpha = 2, \beta = 2$</td>
<td>75.69± 1.23</td>
<td>75.69± 1.23</td>
</tr>
<tr>
<td>$\alpha = 2, \beta = 3$</td>
<td>74.41± 1.83</td>
<td>56.08± 1.11</td>
</tr>
<tr>
<td>$\alpha = 3, \beta = 3$</td>
<td>75.53± 1.23</td>
<td>75.53± 1.22</td>
</tr>
</tbody>
</table>

Table 4: Classification Accuracies (%) of Projection Method and Generalized Histogram Intersection Kernel Using Breast Cancer Data.

<table>
<thead>
<tr>
<th>PARAMETERS</th>
<th>Projection Method</th>
<th>GHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 1, \beta = 1$</td>
<td>96.73± 0.11</td>
<td>96.73± 0.11</td>
</tr>
<tr>
<td>$\alpha = 1, \beta = 2$</td>
<td>97.06± 0.01</td>
<td>90.12± 4.78</td>
</tr>
<tr>
<td>$\alpha = 1, \beta = 3$</td>
<td>97.01± 0.01</td>
<td>75.61± 7.44</td>
</tr>
<tr>
<td>$\alpha = 2, \beta = 2$</td>
<td>96.71± 0.11</td>
<td>96.71± 0.11</td>
</tr>
<tr>
<td>$\alpha = 2, \beta = 3$</td>
<td>96.92± 0.01</td>
<td>96.96± 0.01</td>
</tr>
<tr>
<td>$\alpha = 3, \beta = 3$</td>
<td>96.63± 0.1</td>
<td>96.63± 0.1</td>
</tr>
</tbody>
</table>

when focusing on LogDet Divergence is given as a function of original kernel $K$.

In this subsection, we conduct some experiments on the considered data sets, to check the performance of Projection Method with suggested optimal $\lambda$ and performance of Projection Method with various values of $\lambda > 0$. We take $\alpha \neq \beta \in \{1, 2, 3\}$ and $\lambda \in [0.1, 20]$ with step size 0.1. Figures 1 to 9 plot the performance of Projection Method with different values of $\lambda$. The dot ‘o’ with blue color gives the suggested optimal $\lambda$ obtained using kernel learning with LogDet Divergence. The following table lists the optimal $\lambda$ for the 3 considered data sets.

Table 5: Optimal $\lambda$ suggested in Projection Method

<table>
<thead>
<tr>
<th>PARAMETERS</th>
<th>SONAR DATA</th>
<th>LIVE DATA</th>
<th>BREAST DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 1, \beta = 2$</td>
<td>2.00</td>
<td>2.37</td>
<td>4.60</td>
</tr>
<tr>
<td>$\alpha = 1, \beta = 3$</td>
<td>2.00</td>
<td>2.45</td>
<td>6.57</td>
</tr>
<tr>
<td>$\alpha = 2, \beta = 3$</td>
<td>2.00</td>
<td>2.38</td>
<td>4.06</td>
</tr>
</tbody>
</table>

In Figures 1 to 3, the suggested optimal $\lambda$ is 2.0, we can see clearly that the performance of Projection Method decreases steadily when $\lambda > 2$, implying that we have selected a good $\lambda$. When $\lambda < 1$, the performance is quite unstable which is consistent with our analysis (we suggest $\lambda \geq 1$) because the positive semi-definite property cannot be guaranteed.

Figures 4 to 6 record the performance of Projection Method for Live Disorder Data. For example, in Figure 5 the optimal $\lambda$ in experiment is 1.6, with averaged AUC value 0.7441. Our suggested $\lambda$ is 2.3659, with averaged AUC value 0.7409. With the increment of $\lambda$, the performance is becoming slightly worse. The performance between Projection Method with our suggested optimal $\lambda = 2.3659$ and $\lambda = 1.6$ is very little: 0.003.

Figures 7 to 9 record the performance of Projection Method for Breast Cancer Data. For example, in Figure 8 the optimal $\lambda$ in experiment is 1.0, with averaged AUC value 0.9700. Our suggested $\lambda$ is 6.5738, with averaged AUC value 0.9664. With the increment of $\lambda$, the performance is becoming slightly worse. The performance between Projection Method with our suggested optimal $\lambda = 6.5738$ and $\lambda = 1.0$ is very little: 0.0006.

We note that when $\lambda$ lies in $(0, 1)$ where the positive semi-definiteness cannot be assured, the performance is very unstable. There is no fixed optimal $\lambda$ as well, we cannot conclude denoising method is better or worse than flipping method. Our suggested theoretical $\lambda$ sometimes cannot guarantee the best performance, but we can see that its performance is close to the optimal. One possible reason is that, our framework of optimal $\lambda$ determination is kernel learning by unconstrained optimization where it assumes the positive definiteness of the kernels, however, in our case, we use indefinite kernels. And the inverse of kernel was substituted by pseudo inverse as well. It is gratifying that our suggested $\lambda$ can always yield near optimal performance.

4 Conclusion

In this paper, we propose a projection method to transform an indefinite kernel to a positive semi-definite one. The projection method is very flexible and comprehensive, by varying the parameter $\lambda$, to change from denoising method to flipping method. They are two well-known techniques for dealing with indefinite kernels. By introducing GHI kernel method which is not generally positive semi-definite, we compare the performance of projection method with GHI kernel in terms of AUC values through 5-fold cross-validations. We also determine the optimal $\lambda$ in Projection Method.
through introducing kernel learning by unconstrained optimization with LogDet Divergence. Results show that experimental studies are consistent with the theoretical analysis and our suggested $\lambda$ can always yield near optimal performance for $\lambda \in (0, 20]$ with step size 0.1. This can be regarded as a good choice for dealing with indefinite kernels. Future research may contribute to the development of more precise method in determination of optimal $\lambda$ and the applications of projection method in more wide areas.

Acknowledgments
This research is supported in part by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China, GRF Grant, Hung Hing Ying Physical Research Grant and HKU CERG Grants, National Natural Science Foundation of China Grant Nos. 10971075 and S201201009985.

References


[19] D.Conforti, R.Guido, “Kernel based support vec-

Figure 8: Averaged AUC Values for Different Values of $\lambda$ in Projection Method Using Breast Cancer Data Set

Figure 9: Averaged AUC Values for Different Values of $\lambda$ in Projection Method Using Breast Cancer Data Set