<table>
<thead>
<tr>
<th>Title</th>
<th>Estimation of High Dimensional Precision Matrix using Random Matrix Theory (Statistical Inference on Divergence Measures and Its Related Topics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>伊藤 翼</td>
</tr>
<tr>
<td>Citation</td>
<td>数理解析研究所講究録 2016年7月 1999: 86-94</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2016-07</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/224769">http://hdl.handle.net/2433/224769</a></td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Textversion</td>
<td>publisher</td>
</tr>
</tbody>
</table>

Kyoto University
Estimation of High Dimensional Precision Matrix using Random Matrix Theory
Tsubasa Ito
Graduate School of Economics, University of Tokyo

1 Introduction

About the problem of estimating the high-dimensional covariance matrix, it is well
known that we cannot invert the standard sample covariance matrix $S_p$ when $p > N$, and
even if $N > p$ but $p/N$ is relatively large it performs poorly. When we have no advance
information about the structure of the population covariance matrix $\Sigma_p$, shrinking $S_p$
to some stable statistics improves the performance. There are little research on the
direct estimation of the precision matrix and seems to be room for improvement over
the estimators, $U_pA_pU_p^T$, $\alpha(S_p + \gamma I_p)^{-1}$, $\alpha S_p^{-1} + \beta I_p$ proposed in recent years. Then we
propose $\alpha(S_p + \gamma I_p)^{-1} + \beta I_p$.

2 Preliminaries

We begin by stating the basic assumptions which are common in estimation of the
high-dimensional covariance matrix based on the random matrix theory. Throughout the
paper, and denote the spaces of real and complex numbers, respectively. Also, $+$ denotes
the half-plane of complex numbers with strictly positive imaginary part. The real and
imaginary parts of $z \in \mathbb{C}$ are denoted by $\Re(z)$ and $\Im(z)$, respectively.

(A1) $p/N \rightarrow y \in (0, 1) \cup (1, +\infty)$ as $p, N \rightarrow +\infty$.

(A2) $\Sigma_p$ is a non-random $p$-dimensional positive definite matrix. $X_p = (x_{p,1}, \ldots, x_{p,N})^T$
is an $N \times p$ random matrix, where $x_{p,1}, \ldots, x_{p,N}$ are mutually i.i.d as $E[x_{p,j}] = 0$ and
$\text{Cov}(x_{p,j}) = I_p$. $Y_p = (y_{p,1}, \ldots, y_{p,N})^T$, where $y_{p,j} = \Sigma_p^{1/2}x_{p,j}$.

(A3) $t_p = (t_{p,1}, \cdots, t_{p,p})^T$ is a system of eigenvalues of $\Sigma_p$, sorted in decreasing order.
The empirical spectral distribution (ESD) of $\Sigma_p$ is defined by

$$
H_p(t) \equiv \frac{1}{p} \sum_{i=1}^{p} I_{[t_{p,i}, +\infty)}, \quad \forall t \in \mathbb{R}.
$$
$H_p(t)$ converges to limit $H(t)$ at all points of continuity of $H$.

(A4) $\text{Supp}(H)$, the support of $H$, is the union of a finite number of closed intervals, bounded away from zero and infinity.

Let $S_p = N^{-1}Y_p^TY_p$. $\ell_p = (\ell_{p,1}, \ldots, \ell_{p,p})^T$ and $(u_1, \ldots, u_p)$ are a system of eigenvalues sorted in decreasing order and eigenvectors of $S_p$. The empirical spectral distribution (ESD) of $S_p$ is defined by

$$F_p(t) \equiv \frac{1}{p} \sum_{i=1}^{p} I[\ell_{p,i}, +\infty), \forall t \in \mathbb{R}.$$ 

For a nondecreasing function $G$ on the real line, the stieltjes transform $m_G$ of $G$ is defined by

$$m_G(z) \equiv \int \frac{1}{x-z}dG(x), \forall z \in \mathbb{C}^{+},$$

where $\mathbb{C}^{+}$ denotes the half-plane of complex numbers with strictly positive imaginary part.

The stieltjes transform has the well-known inversion formula

$$G[\{a, b\}] = \frac{1}{\pi} \lim_{\eta \to 0^+} \int_a^b \Im(m_G(\xi + i\eta))d\xi,$$

if $G$ is continuous at $a$ and $b$. Stieltjes transform of $F_p$ is

$$m_{F_p}(z) = \int \frac{1}{\lambda-z}dF_p(\lambda) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{\ell_i - z} = \frac{1}{p} \text{tr}(S_p - zI_p)^{-1}.$$ 

Under (A1)-(A4) and assumption that entries of $X_p$ are independent with common mean and variance and for any $\eta > 0$, as $p/N \to y$

$$\frac{1}{\eta^2 Np} \sum_{jk} E[|x_{jk}^{(p)}|^2 I(|x_{jk}^{(p)}| > \eta N^{1/2})] \to 0,$$

there exists a distribution function $F$ (limiting spectral distribution (LSD)) such that

$$F_p(x) \to F(x), \forall x \in \mathbb{R}\setminus\{0\}.$$ 

$F$ is everywhere continuous except at zero, and that the mass of $F$ at zero is

$$F(0) = \max\{1 - y^{-1}, H(0)\}.$$ 

Under the same assumptions, $m \equiv m_F(z)$ is the unique solution to the equation (Silverstein (1995))

$$m_F(z) = \int \frac{1}{t(1-y-yzm_F(z)) - z}dH(t), \forall z \in \mathbb{C}^{+}.$$
3 Estimation of the precision matrix

We consider the following loss function

\[ L_p(\Sigma_p^{-1}, \Omega_p) \equiv \frac{1}{p} \text{tr}(\Omega_p \Sigma_p - I_p)(\Omega_p \Sigma_p - I_p)^T \]

Instead of minimizing \( R(\Sigma_p^{-1}, \Omega_p) \equiv E[L_p(\Sigma_p^{-1}, \Omega_p)] \), we minimize the limit of \( L_p(\Sigma_p^{-1}, \Omega_p) \) obtained from RMT. We consider rotation-equivariant estimator.

\[ \Omega_p = U_p A_p U_p^T \]

where \( A_p \equiv \text{Diag}(a_1, \cdots, a_p) \)

finite-sample optimal \( a_i \) is

\[ a_i^* = \frac{u_i^T \Sigma_p u_i}{u_i^T \Sigma_p^2 u_i} \]

Ledoit and Wolf (2012) consider the limit of \( \hat{a}_i = u_i^T \Sigma_p u_i \) under \( \tilde{L}_p(\Sigma_p^{-1}, \Omega_p) = \frac{1}{p} \text{tr}(\Sigma_p^{-1} - \Omega_p)^2 \).

\[ \delta(\ell_i) \]

where \( \delta(\ell_i) = \begin{cases} \frac{\ell_i}{1 - y - \ell_i m_F(\ell_i)} & \text{if } \ell_i > 0 \\ \frac{1}{(y-1)m_F(0)} & \text{if } \ell_i = 0 \text{ and } y > 1 \\ 0 & \text{otherwise} \end{cases} \)

\( \phi(\ell_i) \)

where \( \phi(\ell_i) = \begin{cases} \frac{\ell_i}{1 - y - \ell_i m_F(\ell_i)} & \text{if } \ell_i > 0 \\ \frac{1}{(y-1)m_F(0)}(\frac{1}{m_F(z)} - \frac{1}{y m_F(0)}) & \text{if } \ell_i = 0 \text{ and } y > 1 \\ 0 & \text{otherwise} \end{cases} \)

\( F \) is LSD of \( \frac{1}{N} Y_p Y_p^T = \frac{1}{N} X_p \Sigma_p X_p^T \) and \( m_F(z) \) is the solution of \( m = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{t_i(1 - (p/N) - \ell_i m)} - \ell_i \).

By replacing \( m_F(\ell_i) \) and \( m_F(0) \) with their estimator \( \hat{m}_F(\ell_i) \) and \( \hat{m}_F(0) \), we obtain \( \hat{\Omega}_p^{LW} = U_p \hat{A}_p U_p^T \hat{a}_i^* = \hat{\delta}(\ell_i)/\hat{\phi}(\ell_i) \). We use a package QuEST on Matlab introduced in Ledoit and Wolf to estimate \( \hat{m}_F(\ell_i) \). In this algorithm, we obtain \( \hat{t}_p \), the consistent estimator of eigenvalues of \( \Sigma_p \) and solve

\[ m = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{t_i(1 - (p/N) - (p/N)\ell_i m)} - \ell_i \]

When \( N, p \) are relatively small, the approximations of \( u_i^T \Sigma_p u_i, \hat{u}_i^T \Sigma_p^2 u_i \) by \( \hat{\delta}(\ell_i), \hat{\phi}(\ell_i) \) become bad, and \( \hat{\Omega}_p^{LW} \) performs poorly. We propose the following estimator of the precision matrix.

\[ \hat{\Omega}_p^{LR} = \alpha(S_p + \gamma I_p)^{-1} + \beta I_p \]
In the case of $N > p$, consider the following hierarchical bayes model.

$$V(=NS_p) | \Sigma_p \sim \mathcal{W}_p(N, \Sigma_p)$$

$$\Sigma_p^{-1} | \eta \sim (1 - \eta) \mathcal{W}_p(k, \Lambda_1) + \eta \delta_{\Lambda_0}(\Sigma_p^{-1})$$

$$\eta \sim Ber(\theta)$$

Denote pdf of $V$ and prior distribution of $\Sigma_p^{-1}$ by

$$V \mid \Sigma_p^{-1} \sim f(V \mid \Sigma_p^{-1})$$

$$\Sigma_p^{-1} \mid \eta \sim (1 - \eta) \pi(\Sigma_p^{-1} \mid \Lambda_1) + \eta \delta_{\Lambda_0}(\Sigma_p^{-1})$$

The joint distribution of $(V, \Sigma_p^{-1})$ and marginal distribution of $V$ are

$$f(V, \Sigma_p^{-1}) = f(V \mid \Sigma_p^{-1})((1 - \theta) \pi(\Sigma_p^{-1} \mid \Lambda_1) + \theta \delta_{\Lambda_0}(\Sigma_p^{-1}))$$

$$f(V) = (1 - \theta) \int f(V \mid \Sigma_p^{-1})\pi(\Sigma_p^{-1} \mid \Lambda_1)d\Sigma_p^{-1} + \theta f(V \mid \Lambda_0).$$

$$\Omega_p^{Bayes} = E[\Sigma_p^{-1} \mid V]$$

is

$$\Omega_p^{Bayes} = \frac{\int \Sigma_p^{-1} f(V, \Sigma_p^{-1})d\Sigma_p^{-1}}{f(V)}$$

$$= \frac{(1 - \theta) \int f(V \mid \Sigma_p^{-1})\pi(\Sigma_p^{-1} \mid \Lambda_1)d\Sigma_p^{-1} + \theta \delta_{\Lambda_0}(V, \Lambda_0)}{(1 - \theta) \int f(V \mid \Sigma_p^{-1})\pi(\Sigma_p^{-1} \mid \Lambda_1)d\Sigma_p^{-1} + \theta f(V \mid \Lambda_0)}$$

$$= (1 - w_0) \frac{\int f(V \mid \Sigma_p^{-1})\pi(\Sigma_p^{-1} \mid \Lambda_1)d\Sigma_p^{-1}}{f(V \mid \Sigma_p^{-1})\pi(\Sigma_p^{-1} \mid \Lambda_1)d\Sigma_p^{-1}} + w_0 \Lambda_0,$$

where

$$w_0 = \frac{\theta f(V \mid \Lambda_0)}{(1 - \theta) \int f(V \mid \Sigma_p^{-1})\pi(\Sigma_p^{-1} \mid \Lambda_1)d\Sigma_p^{-1} + \theta f(V \mid \Lambda_0)}.$$

Let $v_0 = (N + k)/N$,

$$\frac{\int \Sigma_p^{-1} f(V \mid \Sigma_p^{-1})\pi(\Sigma_p^{-1} \mid \Lambda_1)d\Sigma_p^{-1}}{\int f(V \mid \Sigma_p^{-1})\pi(\Sigma_p^{-1} \mid \Lambda_1)d\Sigma_p^{-1}} = (N + k)(V + \Lambda_1)^{-1}$$

$$= v_0(S_p + N^{-1}\Lambda_1)^{-1},$$

then, we get

$$\Omega_p^{Bayes} = (1 - w_0)v_0(S_p + N^{-1}\Lambda_1)^{-1} + w_0 \Lambda_0.$$

where $v_0 > 1$, $0 < w_0 < 1$. Letting $\Lambda_1 = N\gamma I_p$, $\Lambda_0 = (1/\bar{\ell})I_p$, $\bar{\ell} = \sum_{i=1}^{p} \ell_i/p = tr[S_p]/p$, $\alpha = v_0(1 - w_0)$, $\beta = w_0/\bar{\ell}$, we obtain

$$\Omega_p^{LR} = \alpha(S_p + \gamma I_p)^{-1} + \beta I_p.$$
We estimate $\alpha$, $\beta$, $\gamma$ to satisfy $v_0 > 1$, $0 < w_0 < 1$.

Under $L_p(\Sigma_p^{-1}, \Omega_p) \equiv \frac{1}{p}tr(\Omega_p\Sigma_p - I_p)(\Omega_p\Sigma_p - I_p)^T$,

$$
\alpha^*(\gamma) = \frac{\text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p]\text{tr}[\Sigma_p^2] - \text{tr}[(S_p + \gamma I_p)^{-2}\Sigma_p^2]\text{tr}[\Sigma_p]}{\text{tr}[(S_p + \gamma I_p)^{-2}\Sigma_p^2]\text{tr}[\Sigma_p^2] - \text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p]^2} \\
\beta^*(\gamma) = \frac{\text{tr}[(S_p + \gamma I_p)^{-2}\Sigma_p^2]\text{tr}[\Sigma_p] - \text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p]^2\text{tr}[\Sigma_p^2] - \text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p]^2\text{tr}[\Sigma_p]}{\text{tr}[(S_p + \gamma I_p)^{-2}\Sigma_p^2]\text{tr}[\Sigma_p^2] - \text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p]^2\text{tr}[\Sigma_p]} \\
L_p^*(\gamma) = L_p(\Sigma_p^{-1}, \Omega_p^{LR}(\alpha^*(\gamma), \beta^*(\gamma), \gamma)) = \frac{1}{p}\left[\text{tr}[(S_p + \gamma I_p)^{-2}\Sigma_p^2]\text{tr}[\Sigma_p] - \text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p]^2\text{tr}[\Sigma_p] + 1\right]
$$

Wang, et.al (2014) shows, for $\gamma > 0$

$$\frac{1}{p}\text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p] \text{ a.s.} \frac{1 - \gamma m_F(-\gamma)}{1 - y(1 - \gamma m_F(-\gamma))}$$

Wang, et.al (2014) shows this by considering the limit of $F^\Sigma_{p}^{1/2}(S_p + \gamma I_p)S_p^{-1/2}$. From slide 11, we know

$$\frac{1}{p}\text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p^2] \text{ a.s.} \frac{-\gamma + \gamma^2 m_F(-\gamma)}{(1 - y(1 - \gamma m_F(-\gamma)))^2} + \frac{\int tdH(t)}{1 - y(1 - \gamma m_F(-\gamma))}.$$

Since $p^{-1}\text{tr}[(S_p + \gamma I_p)^{-2}\Sigma_p^2] = -(d/d\gamma)p^{-1}\text{tr}[(S_p + \gamma I_p)^{-1}\Sigma_p^2]$, we have

$$\frac{1}{p}\text{tr}[(S_p + \gamma I_p)^{-2}\Sigma_p^2] \rightarrow \frac{d}{d\gamma} \left\{ \frac{-\gamma + \gamma^2 m_F(-\gamma)}{(1 - y(1 - \gamma m_F(-\gamma)))^2} + \frac{\int tdH(t)}{1 - y(1 - \gamma m_F(-\gamma))} \right\}$$

We estimate $m_F(-\gamma) \approx m_F'(-\gamma)$ by $p^{-1}\text{tr}[(S_p + \gamma I_p)^{-1}]$, $p^{-1}\text{tr}[(S_p + \gamma I_p)^{-2}]$. Consistent estimator of $p^{-1}\text{tr}(\Sigma_p) \rightarrow \int tdH(t)$ is $p^{-1}\text{tr}(S_p)$. For $p^{-1}\text{tr}(\Sigma_p^2) \rightarrow \int t^2dH(t)$, $\hat{a}_2 = (N-1)^{-1}\frac{(N-2)^{-1}(N-3)^{-1}(N-4)^{-1}(N-5)^{-1})\text{tr}(S_p)^2}{(N-1)^{-1}\sum_{i=1}^{N}((y_i - \bar{y})^2)}$ is a consistent estimator which proposed by Himeno and Yamada (2014).

We look at two estimators: the ridge and the linear shrinkage estimators and check the optimal values of the parameters in these estimators with respect to our loss function.


The ridge estimator is of the form $\Omega_p^{ridge} = \alpha(S_p + \gamma I_p)^{-1}$ Given $\gamma$, the optimal $\alpha$ is
\( \alpha^{\text{ridge}}(\gamma) = \frac{\text{tr}[(S_p + \gamma I_p)^{-1} \Sigma_p]}{\text{tr}[(S_p + \gamma I_p)^{-1} S_p (S_p + \gamma I_p)^{-1}]} \), which leads to the reduced loss function
\[
L_p(\Sigma_p^{-1}, \Omega_p^{\text{ridge}}(\alpha^{\text{ridge}}(\gamma), \gamma)) = 1 - \frac{1}{p} \frac{\{[(S_p + \gamma I_p)^{-1} \Sigma_p]\}^2}{\text{tr}[(S_p + \gamma I_p)^{-1} S_p (S_p + \gamma I_p)^{-1}]}.
\]


The linear shrinkage estimator is of the form
\[
\Omega_p^{\text{linear}} = \begin{cases} 
\alpha S_p^{-1} + \beta I_p & \text{if } N > p \\
\alpha S_p + \beta I_p & \text{if } N < p.
\end{cases}
\]

In the case of \( N > p \),
\[
L_p(\Sigma_p^{-1}, \Omega_p^{\text{linear}}) = \frac{1}{p} \{ \alpha^2 \text{tr}[S_p^{-2} \Sigma_p^2] + 2\alpha \beta \text{tr}[S_p^{-1} \Sigma_p^2] + \beta^2 \text{tr}[\Sigma_p^2] - 2\alpha \text{tr}[S_p^{-1} \Sigma_p] - 2\beta \text{tr}[\Sigma_p] \} + 1
\]

In the case of \( N < p \), Bodnar, (2014) cannot provide estimators for general \( \Sigma_p \), because the limit of \( p^{-1} \text{tr}[(S_p^+ \Sigma_p^{-1})^2] \) is needed, which cannot be obtained without assuming a structure such as \( \Sigma_p = \sigma^2 I_p \). Without assuming such a structure, however, we can obtain estimators of the optimal \( \alpha \) and \( \beta \) in our situation. The loss function is
\[
L_p(\Sigma_p^{-1}, \Omega_p^{\text{linear}}) = \frac{1}{p} \{ \alpha^2 \text{tr}[(S_p^+)^2 \Sigma_p^2] + 2\alpha \beta \text{tr}[S_p^+ \Sigma_p^2] + \beta^2 \text{tr}[\Sigma_p^2] - 2\alpha \text{tr}[S_p^+ \Sigma_p] - 2\beta \text{tr}[\Sigma_p] \} + 1
\]

so that we need the limit of \( p^{-1} \text{tr}[(S_p^+)^2 \Sigma_p^2] \), \( p^{-1} \text{tr}[S_p^+ \Sigma_p^2] \) and \( p^{-1} \text{tr}[S_p^+ \Sigma_p] \). By Theorem 3.3 in Bodnar, (2014), one gets

\[
\lim_{N,p \to \infty} p^{-1} \text{tr}[(S_p^+)^2 \Sigma_p^2] = \lim_{N,p \to \infty} p^{-1} \sum_{i=1}^{N} \frac{\phi(\ell_i)}{\ell_i^2} = \int \frac{\phi(x)}{x^2} dF(x)
\]

\[
\lim_{N,p \to \infty} p^{-1} \text{tr}[S_p^+ \Sigma_p^2] = \lim_{N,p \to \infty} p^{-1} \sum_{i=1}^{N} \frac{\phi(\ell_i)}{\ell_i} = \int \frac{\phi(x)}{x} dF(x)
\]

\[
\lim_{N,p \to \infty} p^{-1} \text{tr}[S_p^+ \Sigma_p] = \frac{1}{y-1}
\]

\( p^{-1} \sum_{i=1}^{N} \frac{\phi(\ell_i)}{\ell_i^2} \) is the estimator of \( p^{-1} \text{tr}[(S_p^+)^2 \Sigma_p^2] \).

4 Numerical Results

We compare estimators with \( \alpha(S_p + \gamma I_p)^{-1} \)(Wang, et.al (2014)), \( \alpha S_p^{-1} + \beta I_p \)(Bodnar, et.al (2014)). Data is as follows. \( y_i = \Sigma_p^{1/2} x_i \).
\( x_{ij} \) i.i.d \( \sim N(0, 1) \), \( i = 1, \ldots, N \), \( j = 1, \ldots, p \)

\( x_{ij} = \sqrt{(m - 2)/m} z_{ij} \), \( z_{ij} \) i.i.d \( \sim t_m \), \( i = 1, \ldots, N \), \( j = 1, \ldots, p \), \( m = 10 \)

L.S.D of \( \Sigma_p \) is based on Beta distribution

\[
H_{(a,b)}(x) = \frac{\Gamma(a + b)}{\Gamma(a) \Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}dt, \quad x \in [0, 1],
\]

and the population eigenvalues are generated by

\[
1 + 9 H_{(a,b)}^{-1}\left(\frac{i}{p} - \frac{1}{2p}\right), \quad i = 1, \ldots, p.
\]

Risk is evaluated by the averaging the empirical losses from 1000 times simulation.

<table>
<thead>
<tr>
<th>( p )</th>
<th>oracle</th>
<th>LW</th>
<th>LR</th>
<th>ridge</th>
<th>linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.1538</td>
<td>0.1710</td>
<td>0.1665</td>
<td>0.1681</td>
<td>0.1830</td>
</tr>
<tr>
<td>70</td>
<td>0.1703</td>
<td>0.1782</td>
<td>0.1770</td>
<td>0.1813</td>
<td>0.8705</td>
</tr>
<tr>
<td>Normal</td>
<td>0.1769</td>
<td>0.1854</td>
<td>0.1856</td>
<td>0.1901</td>
<td>0.8081</td>
</tr>
<tr>
<td>( (a,b) = (1,1) )</td>
<td>0.1791</td>
<td>0.1933</td>
<td>0.1902</td>
<td>0.1951</td>
<td>0.9056</td>
</tr>
<tr>
<td>500</td>
<td>0.1800</td>
<td>0.2342</td>
<td>0.1981</td>
<td>0.2209</td>
<td>0.9757</td>
</tr>
<tr>
<td>700</td>
<td>0.1799</td>
<td>0.3789</td>
<td>0.2116</td>
<td>0.2561</td>
<td>0.9877</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( p )</th>
<th>30</th>
<th>70</th>
<th>Normal</th>
<th>( (a,b) = (1,1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_{10} )</td>
<td>0.1702</td>
<td>0.1786</td>
<td>0.1787</td>
<td>0.1823</td>
</tr>
<tr>
<td>150</td>
<td>0.1769</td>
<td>0.1849</td>
<td>0.1869</td>
<td>0.1896</td>
</tr>
<tr>
<td>250</td>
<td>0.1790</td>
<td>0.1911</td>
<td>0.1889</td>
<td>0.1932</td>
</tr>
<tr>
<td>500</td>
<td>0.1801</td>
<td>0.2189</td>
<td>0.2029</td>
<td>0.2117</td>
</tr>
<tr>
<td>700</td>
<td>0.1799</td>
<td>0.2632</td>
<td>0.2174</td>
<td>0.2464</td>
</tr>
</tbody>
</table>

We conduct Quadratic Discriminant Analysis using the microarray data where expression levels for 2000 genes were measured on 22 normal and on 40 colon tumor tissues. Discriminant rule is

\[
\frac{N_1}{N_1 + 1}(x - \bar{x}_1)^T \Omega_p^{(1)} (x - \bar{x}_1) - \frac{N_2}{N_2 + 1}(x - \bar{x}_2)^T \Omega_p^{(2)} (x - \bar{x}_2) < 0 \Rightarrow x \in \Pi_1
\]

where \( \Omega_p^{LR} \), \( \Omega_p^{LW} \), \( \Omega_p^{ridge} \), \( \Omega_p^{linear} \), \( \Omega_p^{MP} \), \( \Omega_p^{diag} \) are used. Correct classification rates are evaluated by leave-one-out cross-validation.

References

### Table 2: Empirical Risks of $\Omega_p^{oracle}$, $\Omega_p^{LW}$, $\Omega_p^{LR}$, $\Omega_p^{ridge}$ and $\Omega_p^{linear}$ with $N = 50$ under Normal Distribution

<table>
<thead>
<tr>
<th>$(a, b)$</th>
<th>$p$</th>
<th>oracle</th>
<th>LW</th>
<th>LR</th>
<th>ridge</th>
<th>linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.1216</td>
<td>0.1354</td>
<td>0.1327</td>
<td>0.1343</td>
<td>0.1437</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.1335</td>
<td>0.1415</td>
<td>0.1416</td>
<td>0.1472</td>
<td>0.8494</td>
<td></td>
</tr>
<tr>
<td>(1.5,1.5)</td>
<td>150</td>
<td>0.1388</td>
<td>0.1468</td>
<td>0.1467</td>
<td>0.1534</td>
<td>0.8043</td>
</tr>
<tr>
<td>250</td>
<td>0.1405</td>
<td>0.1551</td>
<td>0.1487</td>
<td>0.1580</td>
<td>0.9087</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.1415</td>
<td>0.1887</td>
<td>0.1631</td>
<td>0.1813</td>
<td>0.9766</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.2122</td>
<td>0.2288</td>
<td>0.2253</td>
<td>0.2304</td>
<td>0.2542</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.2359</td>
<td>0.2441</td>
<td>0.2436</td>
<td>0.2463</td>
<td>0.8932</td>
<td></td>
</tr>
<tr>
<td>(0.5,0.5)</td>
<td>150</td>
<td>0.2444</td>
<td>0.2534</td>
<td>0.2565</td>
<td>0.2559</td>
<td>0.8279</td>
</tr>
<tr>
<td>250</td>
<td>0.2468</td>
<td>0.2627</td>
<td>0.2547</td>
<td>0.2619</td>
<td>0.9008</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.2480</td>
<td>0.2982</td>
<td>0.2629</td>
<td>0.2866</td>
<td>0.9738</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: Empirical Risks of $\Omega_p^{oracle}$, $\Omega_p^{LW}$, $\Omega_p^{LR}$, $\Omega_p^{ridge}$ and $\Omega_p^{linear}$ with $N = 50$ under Normal Distribution

<table>
<thead>
<tr>
<th>$(a, b)$</th>
<th>$p$</th>
<th>oracle</th>
<th>LW</th>
<th>LR</th>
<th>ridge</th>
<th>linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.0496</td>
<td>0.0585</td>
<td>0.0570</td>
<td>0.0693</td>
<td>0.0595</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.0536</td>
<td>0.0595</td>
<td>0.0604</td>
<td>0.0754</td>
<td>0.8357</td>
<td></td>
</tr>
<tr>
<td>(5,5)</td>
<td>150</td>
<td>0.0557</td>
<td>0.0624</td>
<td>0.0612</td>
<td>0.0757</td>
<td>0.7958</td>
</tr>
<tr>
<td>250</td>
<td>0.0563</td>
<td>0.0688</td>
<td>0.0653</td>
<td>0.0769</td>
<td>0.9089</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.0567</td>
<td>0.1072</td>
<td>0.0843</td>
<td>0.1015</td>
<td>0.9784</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.1123</td>
<td>0.1277</td>
<td>0.1257</td>
<td>0.1268</td>
<td>0.1421</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.1248</td>
<td>0.1340</td>
<td>0.1338</td>
<td>0.1376</td>
<td>0.8357</td>
<td></td>
</tr>
<tr>
<td>(2,5)</td>
<td>150</td>
<td>0.1323</td>
<td>0.1407</td>
<td>0.1416</td>
<td>0.1445</td>
<td>0.8042</td>
</tr>
<tr>
<td>250</td>
<td>0.1350</td>
<td>0.1487</td>
<td>0.1443</td>
<td>0.1507</td>
<td>0.9100</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.1364</td>
<td>0.1843</td>
<td>0.1589</td>
<td>0.1760</td>
<td>0.9769</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4: Correct Classification Rates in the Colon Cancer Dataset

<table>
<thead>
<tr>
<th>$p$</th>
<th>LW</th>
<th>LR</th>
<th>ridge</th>
<th>linear</th>
<th>MP</th>
<th>diag</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>67.7 %</td>
<td>87.1 %</td>
<td>71.0 %</td>
<td>83.9 %</td>
<td>38.7 %</td>
<td>86.5 %</td>
</tr>
<tr>
<td>250</td>
<td>65.2 %</td>
<td>87.1 %</td>
<td>83.9 %</td>
<td>87.1 %</td>
<td>38.7 %</td>
<td>83.9 %</td>
</tr>
<tr>
<td>500</td>
<td>61.3 %</td>
<td>87.1 %</td>
<td>72.6 %</td>
<td>83.9 %</td>
<td>41.9 %</td>
<td>87.1 %</td>
</tr>
<tr>
<td>900</td>
<td>66.1 %</td>
<td>87.1 %</td>
<td>61.3 %</td>
<td>87.1 %</td>
<td>43.6 %</td>
<td>87.1 %</td>
</tr>
</tbody>
</table>


Graduate School of Economics, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033
JAPAN
E-mail address: tsubasa.ito710@gmail.com

東京大学大学院経済学研究科 伊藤 翼