Homework 3

WU Jiamin 20659352

1. (a) $$l_n(\mu, \Sigma) = \log \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(X_i-\mu)^T \Sigma^{-1} (X_i-\mu)}$$

$$= -\frac{1}{2}\sum_{i=1}^{n} (X_i-\mu)^T \Sigma^{-1} (X_i-\mu) + \log \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}}$$

$$= -\frac{1}{2}\sum_{i=1}^{n} tr\left[(X_i-\mu)^T \Sigma^{-1} (X_i-\mu)\right] - \frac{n}{2}\log det(\Sigma) + C$$

$$= -\frac{1}{2}\sum_{i=1}^{n} tr\left[\Sigma^{-1} (X_i-\mu)^T (X_i-\mu)\right] - \frac{n}{2}\log det(\Sigma) + C$$

$$= -\frac{n}{2} tr(\Sigma^{-1} S_n) - \frac{n}{2}\log det(\Sigma) + C$$

where $S_n = \frac{1}{n}\sum_{i=1}^{n} (X_i-\mu)(X_i-\mu)^T$.

(b) $f(x) = \cancel{tr}\, (AX^{-1}) \quad A, X \succeq 0$

$$f(X+\Delta) = tr(A(X+\Delta)^{-1})$$

$$= tr(AX^{-1}(I+\Delta X^{-1})^{-1})$$

$$= tr(AX^{-1}(I-\Delta X^{-1}))$$

$$= tr(AX^{-1} - AX^{-1}\Delta X^{-1})$$

$$= f(X) - tr(AX^{-1}\Delta X^{-1})$$

$$= f(X) - tr(X^{-1}A' X^{-1}\Delta)$$

$$\frac{df(X)}{dX} = \frac{f(X+\Delta)-f(X)}{\Delta} = \frac{-tr(X^{-1}A'X^{-1}\Delta)}{\Delta}$$

$$= -X^{-1}AX^{-1}$$

(c) $g(X) = \log det(X)$

$$g(X+\Delta) = \log det(X+\Delta)$$

$$= \log det(X + X^{\frac{1}{2}}X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}}X^{\frac{1}{2}})$$

$$= \log det(X^{\frac{1}{2}}(I + X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}})X^{\frac{1}{2}})$$

Consider the eigenvalues of $X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}}$, suppose

$$X^{-\frac{1}{2}} \Delta X^{-\frac{1}{2}} = Q \Lambda Q^{-1}$$

$$g(X+\Delta) = \log\det(X^{\frac{1}{2}} Q (I+\Lambda) Q^{-1} X^{\frac{1}{2}})$$
$$= \log\det [X Q (I+\Lambda) Q^{-1}]$$
$$= \log\det(X) + \log\det(I+\Lambda)$$
$$= g(X) + tr(\Lambda)$$
$$= g(X) + tr(X^{-\frac{1}{2}} \Delta X^{-\frac{1}{2}})$$
$$= g(X) + tr(X^{-1}\Delta)$$

hence $\frac{dg(X)}{dX} = \frac{g(X+\Delta) - g(X)}{\Delta} = \frac{tr(X^{-1}\Delta)}{\Delta} = X^{-1}$

(d) The Maximum likelihood estimator of Σ is to find the $\frac{d(l_n(\mu,\Sigma))}{d\Sigma}$

$$\frac{d(l_n(\mu,\Sigma))}{d\Sigma} = \left(-\frac{n}{2} tr(\Sigma^{-1} S_n)\right)' - \left(\frac{n}{2}\log\det(\Sigma)\right)'$$
$$= \frac{n}{2}\Sigma^{-1} S_n \Sigma^{-1} - \frac{n}{2}\Sigma^{-1}$$
$$= \frac{n}{2}\Sigma^{-1}(S_n - \Sigma)\Sigma^{-1}$$
$$= 0$$

$\Rightarrow \quad S_n = \Sigma$

2. $y \sim N(\mu, I_p)$

(a) $\frac{1}{2}\|y-\mu\|_2^2 + \frac{\lambda}{2}\|\mu\|_2^2$

$$= \frac{1}{2}\|y-\mu\|_2^2 + \frac{\lambda}{2}\frac{\sum_{i=1}^{p} y_i^2}{\lambda+1}$$
$$= \frac{1}{2}\sum_{i=1}^{p}(y_i^2 - 2\mu y_i + \mu_i^2) \quad + \frac{\lambda}{2}\frac{\sum_{i=1}^{p} y_i^2}{\lambda+1}$$

Do derivative w.r.t. μ, we get

$$\hat{\mu}_i^{ridge} = \frac{1}{1+\lambda} y_i$$

$MSE = E\|\mu - \hat{\mu}\|^2$

$$= \sum_{i=1}^{p} E\left(\frac{\lambda}{1+\lambda}\mu_i - \frac{1}{1+\lambda}\sigma_i^2\right)$$

$$= \frac{\lambda^2}{(1+\lambda)^2}\|\mu\|_2^2 + \frac{p}{(1+\lambda)^2}$$

(b) $\frac{1}{2}\|y-\mu\|_2^2 + \lambda\|\mu\|_1$

$$= \frac{1}{2}\sum_{i=1}^{p}(\mu_i^2 - 2y_i\mu_i + y_i^2) + \sum_{i=1}^{p}\lambda|\mu_i|$$

$$= \frac{1}{2}\sum_{i=1}^{p}(\mu_i^2 - 2y_i\mu_i + 2\lambda|\mu_i|) + \frac{1}{2}\sum_{i=1}^{p} y_i^2$$

Take derivative w.r.t. μ_i

if $\mu_i > 0$ $\quad \mu_i = y_i - \lambda > 0$ $\quad$ so $y_i > \lambda$

if $\mu_i < 0$ $\quad \mu_i = y_i + \lambda < 0$ $\quad$ so $y_i < -\lambda$

So $\hat{\mu}_i^{soft} = \mu_{soft}(y_i;\lambda) := \mathrm{sign}(y_i)(|y_i| - \lambda)_+$

if $\lambda = \sqrt{2\log p}$

$E\|\hat{\mu}^{soft}(y) - \mu\|^2$

$$= \int_{\lambda}^{\infty}(y_i - \lambda - \mu_i)^2 dG(y_i - \mu_i) + \int_{-\infty}^{-\lambda}(y_i + \lambda - \mu_i)^2 dG(y_i - \mu_i)$$

$$+ \int_{-\lambda}^{\lambda}\mu^2 dG(y_i - \mu_i)$$

Set $u = y_i - \mu_i$

$$\Rightarrow \int_{\lambda-\mu_i}^{\infty}(u-\lambda)^2 dG(u) + \int_{-\infty}^{-\lambda-\mu_i}(u+\lambda)^2 dG(u) + \int_{-\lambda}^{\lambda}\mu^2 dG(u)$$

$$\Rightarrow \frac{\partial E\|\hat{\mu}^{soft}(y) - \mu\|^2}{\partial \mu_i} \le 2\mu_i$$

$$E\|\hat{\mu}^{soft}(y) - \mu\|^2 \le e^{-\frac{\lambda^2}{2}} \le 1 + \lambda^2$$

$$\le 1 + (2\log p + 1)\sum_{i=1}^{p}\min(\mu_i^2, 1)$$

$$\le MLE$$

$$\Rightarrow \sum_{i=1}^{p}\min(\mu_i^2, 1) < \frac{p-1}{2\log p + 1}$$

(c) $\|y-\mu\|_2^2 + \lambda^2\|\mu\|_0$

$$= \sum_{i=1}^{p}(y_i^2 - 2\mu_i y_i + \mu_i^2 + \lambda^2 I_{\{\mu_i \ne 0\}})$$

Take derivative w.r.t. μ_i, we get

$$\sum_{i=1}^{p} (2\mu_i - 2y_i) = 0$$

$$\mu_i = y_i \quad \text{for } \mu_i \neq 0$$

$$\min_{\mu} \|y - \mu\|_2^2 + \lambda^2 \|\mu\|_0 \geq \min(y_i^2, \lambda^2)$$

So $\hat{\mu}_i^{hard} = \mu_{hard}(y_i; \lambda) := y_i \, I(|y_i| > \lambda)$

$$= (1 - g(y))y$$

if $|y_i| > \lambda$, $g(y) = 0$

$|y_i| \leq \lambda$, $g(y) = 1$

Take $\forall\, \varphi \in C_c^{\infty}(\mathbb{R})$, then

$$\int_{\mathbb{R}} \varphi'(y)\, g(y)\, dy = \int_{-\infty}^{\lambda} \varphi'(y)\, dy$$

is not weakly differentiable from the lecture notes.

d)
$$E\|\hat{\mu}^{JS}(y) - \mu\|^2$$

$$= E\|(1 - \frac{\alpha}{\|y\|^2})y - \mu\|^2$$

$$= \sum_{i=1}^{p} E(y_i - \frac{\alpha}{\|y\|^2} y_i - \mu_i)^2$$

$$= \sum_{i=1}^{p} E[(y_i - \mu_i) - \frac{\alpha}{\|y\|^2} y_i)^2$$

$$= \sum_{i=1}^{p} E(\sigma_i - \frac{\alpha}{\|y\|^2} y_i)^2$$

$$= \sum_{i=1}^{p} E(\sigma_i^2 + (\frac{\alpha}{\|y\|^2} y_i)^2 - 2\sigma_i \frac{\alpha}{\|y\|^2} y_i)$$

$$= p + E\frac{\alpha^2}{\|y\|^2} - 2E\alpha \frac{p\mu}{\|y\|^2}$$

$$= E[p + \frac{\alpha^2}{\|y\|^{2}} - 2\alpha \sum_{i=1}^{p} \frac{\|y\|^2 - 2y_i^2}{\|y\|^4}]$$

$$= E[p + \frac{\alpha^2 - 2\alpha(p-2)}{\|y\|^2}]$$

$$= E U_\alpha(y)$$

Take derivative w.r.t. α, we get

$$2\alpha - 2(p-2) = 0 \Rightarrow \alpha^* = p-2$$

$$U_\alpha(y) = p + \frac{(p-2)^2 - 2(p-2)^2}{\|y\|^2}$$

$$= p - \frac{(p-2)^2}{\|y\|^2}$$

if $p>2$ $\quad U_\alpha(y) < p \Rightarrow E(U_\alpha(y)) < p = \text{risk}(MLE)$

(e) Shrinkage rule, odd rule, Monotone, unbounded are shrinkage rules.

3. (a) Set D s.t. $I - D = |I-C|$

then we use the variance-bias decomposition.

$$(I-D)^T(I-D) = |I-C|^2 = (I-C)^T(I-C)$$

C and D have the same bias, then

$$tr(D^TD) = tr I - 2tr(I-D) + tr(I-D)^T(I-D)$$

if C is symmetric,

$$tr(I-D) = tr|I-C| > tr(I-C)$$

and $tr D^TD < tr C^TC$

(b) If C is symmetric, $C = U\Lambda U^T$ where U is orthogonal and $\Lambda = diag(l_i)$

Let $\eta = U^T\mu$, $x = U^Ty \sim N(\eta, \sigma^2 I_p)$

$E\|Cy - \mu\|^2 = E\|\Lambda - \eta\|^2$, then

$$r(\hat{\mu}_c, \mu) = r(\hat{\eta}_\Lambda, \eta) = \sum_i \sigma^2 l_i^2 + (1-l_i)^2\eta_i^2 = \sum_i r(l_i, \eta_i)$$

if $l_i > 1$ we may take $l_i = 1$

if $l_i < 0$ $\quad \cdots\cdots \quad l_i = 0$

so $l_i \in [0,1]$

(c) Suppose $f_1 = \cdots = f_d = 1 > f_i$ for $i > d \geq 3$

$x^d = (x_1, \cdots x_d)$

$r(\hat{\eta}, \eta) = r(\hat{\eta}^{JS}, \eta^d) + \sum_{i>d} r(f_i \eta_i) < r(\lambda, \eta)$

So $\hat{\eta}$ dominates $\hat{\eta}_\lambda$ and $\hat{\mu}_c$

Hence $\hat{\mu}_c$ is admissible only if (a), (b), (c).

4. $p=1$ $R(\hat{\mu}^{JS}, \mu) = 1 - E_\mu \frac{1}{\|Y\|^2} < 1 = p = R(\hat{\mu}^{MLE}, \mu)$

$p=2$ $R(\hat{\mu}^{JS}, \mu) = 2 - 0 = 2 = p = R(\hat{\mu}^{MLE}, \mu)$

For $p=1$ risk for James Stein Estimator is lower than MLE

$p=2$ risk for James Stein Estimator is equal to MLE

$$R(\hat{\mu}^{JS}, \mu) = p - E_\mu \frac{(p-2)^2}{\|Y\|^2} = p - (p-2)^2 E\left(\frac{1}{\|Y\|^2}\right)$$

$\|Y\|^2 \sim \chi^2(\|\mu\|^2, p)$

$\stackrel{d}{=} \chi^2(0, p+2N)$ where $N \sim \text{Poisson}\left(\frac{\|\mu\|^2}{2}\right)$

$$E\left(\frac{1}{\|Y\|^2}\right) = E_N E_Y \left[\frac{1}{\|Y\|^2} \Big| N\right]$$

$$= E \frac{1}{p+2N-2}$$

$$\leq \frac{1}{p+2EN-2}$$

$$= \frac{1}{p+\|\mu\|^2-2}$$

$$R(\hat{\mu}^{JS}, \mu) = p - (p-2)^2 E\left(\frac{1}{\|Y\|^2}\right)$$

$$\leq p - (p-2)^2 \frac{1}{p+\|\mu\|^2-2}$$

$$= 2 + \frac{(p-2)\|\mu\|^2}{p-2+\|\mu\|^2}$$