HW3. Name: LEE Jooran, SID: 20538819

Q1. (a) what we need to prove.

$$l_n(\mu, \Sigma) = -\frac{n}{2}\operatorname{trace}(\Sigma^{-1}S_n) - \frac{n}{2}\log\det(\Sigma) + C.$$

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

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

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

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

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

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

(b) If $f(X) = \operatorname{trace}(AX^{-1})$ with $A, X \succeq 0$,

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

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

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

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

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

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

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

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

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

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

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

considering the eigenvalues of $X^{-\frac{1}{2}} \Delta X^{-\frac{1}{2}} = Q \Lambda Q^{-1}$

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

$$= \log\det\left(X Q (I+\Lambda) Q^{-1}\right)$$

$$= \log\det(X) + \log\det(I+\Lambda)$$

$$= g(X) + tr(\Lambda) = g(X) + tr\left(X^{-\frac{1}{2}} \Delta X^{-\frac{1}{2}}\right)$$

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

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

(d) $\hat{\Sigma} = \dfrac{d(\ln(M,\Sigma))}{d\Sigma}$ .. where $\hat{\Sigma}$ = maximum likelihood estimator of Σ.

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

$$= \frac{n}{2} \Sigma^{-1} S_n \Sigma^{-1} - \frac{n}{2} \Sigma^{-1}$$

$$= \frac{n}{2} \Sigma^{-1} (S_n - \Sigma) \Sigma^{-1} = 0$$

$$\therefore S_n = \Sigma$$

Q2.

(a) $y \sim N(\mu, I_p)$

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

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

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

if we defferentiate the function w.r.t μ, $\hat{\mu}_i^{ridge} = \frac{1}{1+\lambda} y_i$

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

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

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

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

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

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

Taking derivative w.r.t μ_i,

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

if $\lambda = \sqrt{2\log p}$, $E\|\hat{\mu}^{soft}(y) - \mu\|^2$

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

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

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

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

$$\therefore \quad \frac{\partial E\|\hat{\mu}^{soft}(y) - \mu\|^2}{\partial \mu_i} \leq 2\mu_i.$$

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

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

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

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

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

Taking derivative w.r.t μ_i, we get : $\sum_{i=1}^{p} (2\mu_i - 2y_i) = 0$

$\therefore \mu_i = y_i$ for $\mu_i \neq 0$.

$\therefore \min_{\mu} \|y-\mu\|_2^2 + \lambda^2 \|\mu_0\| \geq \min(y_i^2, \lambda^2)$

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

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

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

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

$\therefore$ Taking $\forall$ $y \in C_c^\infty(\mathbb{R})$, then $\int_{\mathbb{R}} \varphi'(y) g(y) dy$

$= \int_{-\infty}^{\lambda} y'(y) dy$

it's not weakly differentiate if we look up the lecture material.

Q3. (a) Setting D s.t $I-D = |I-C|$, by using the Variance-bias decomposition,

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

C and D have same bias, therefore,

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

and If C is symmetric, then

$$tr(I-D) = tr(I-C) > tr(I-C) \text{ and}$$

$$tr\, DD^T < tr\, C^TC.$$

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

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

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

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

$$= \sum_i r(P_i, \eta_i)$$

If $\begin{cases} P_i > 1 \rightarrow P_i = 1 \\ P_i < 0 \rightarrow P_i = 0 \end{cases}$ $\therefore P_i \in [0, 1]$

Q4. $p=1, \quad R(\hat{\mu}^{JS}, \mu) = 1 - E_\mu \cdot \frac{1}{\|Y\|^2} < 1 = p = R(\hat{\mu}^{MLE}, \mu)$

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

For $p=1$, risk for JS estimator < MLE

For $p=2$, risk for JS estimator = MLE

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

$$= p - (p-2)^2 E\left(\frac{1}{\|Y\|^2}\right)$$

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

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

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

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

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

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

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