No.

Date

HW-3

Q1) $x_i \in \mathbb{R}^p \sim N(\mu, \Sigma)$

a) $l_n(\mu, \Sigma) = -\frac{n}{2} Tr(\Sigma^{-1} S_n) - \frac{n}{2} \log|\Sigma| + C$

Multivariate Gaussian distribution, Pdf

$$f(x_i|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left[-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)\right]$$

x_is are iid, take product and then log

$$l_n(\mu, \Sigma) = \underbrace{-\frac{p}{2} \times n \log(2\pi)}_{\text{constant}} - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

$$l_n(\mu, \Sigma) = -\frac{n}{2}\log|\Sigma| - \frac{n}{2} Tr(\Sigma^{-1} S_n) + C$$

b) $f(X) = Tr(AX^{-1})$, $A, X \succeq 0$

$$Y = X + \Delta$$

$$f(Y) = Tr(A(X+\Delta)^{-1}) = Tr(AX^{-1}(I + X^{-1}\Delta)^{-1})$$

$$= Tr[AX^{-1}(I + X^{-1}\Delta)^{-1}]$$

$$= Tr[A(I - X^{-1}\Delta)X^{-1}]$$

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

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

$$\frac{df(x)}{dx} = -X^{-1}AX^{-1}$$

No.

Date

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

let $Z = X + \Delta$

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

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

$= \log |X| + \log |I + X^{-1/2} \Delta X^{-1/2}|$

$= \log |X| + \sum_{i=1}^{n} \log (1 + \lambda_i)$ λ_i - ith eigenvalue of $X^{-1/2} \Delta X^{-1/2}$

Since Δ is small $\Rightarrow$ λ_i are small

$\log(1+\lambda_i) \approx \lambda_i$

$\Rightarrow \log |Z| = \log |X| + \sum_{i=1}^{n} \lambda_i$

$= \log |X| + Tr(X^{-1/2} \Delta X^{-1/2})$

$= \log |X| + Tr(X^{-1} \Delta)$

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

$\frac{dg}{dX} = X^{-1}$

d) $\ln(\mu, \Sigma) = -\frac{n}{2} Tr(\Sigma^{-1} S_n) - \frac{n}{2} \log |\Sigma| + C$

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

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

$\frac{\partial \ln(\mu, \Sigma)}{\partial \mu} = \sum_{i=1}^{n} \Sigma^{-1} (x_i - \mu) = 0$

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial \Sigma} l_n(\mu, \Sigma) = -\frac{n}{2}\left(-\Sigma^{-1} S_n \Sigma^{-1}\right) - \frac{n}{2}\Sigma^{-1} = 0$$

$$\Rightarrow \Sigma^{-1} S_n \Sigma^{-1} = \Sigma^{-1}$$

$$\Rightarrow S_n \Sigma^{-1} = I$$

$$\Sigma = S_n$$

Q2 $y \sim N(\mu, I_p)$ $\quad \min_{\mu} \frac{1}{2}\|y-\mu\|_2^2 + \frac{\lambda}{2}\|\mu\|_2^2$

derivative $\Rightarrow \frac{1}{2} \times 2(y-\mu)(-1) + \frac{\lambda}{2} \times 2\mu = 0$

$$\mu(\lambda+1) = y \Rightarrow \mu_{ridge} = \frac{y}{\lambda+1}$$

$$\text{Bias} = \mu - \frac{\mu}{1+\lambda} = \frac{\lambda}{1+\lambda}\mu$$

$$\text{Variance} = \left(\frac{1}{1+\lambda}y - \frac{1}{1+\lambda}\mu\right)^T\left(\frac{y}{1+\lambda} - \frac{\mu}{1+\lambda}\right)$$

$$= \left(\frac{1}{1+\lambda}\right)^2 \times (y-\mu)^T(y-\mu) = \frac{p}{(1+\lambda)^2}$$

$\text{Bias}(\hat{\mu})^2 =$ $\cancel{\mu}$ $= 1$

$+ \text{Variance} = \text{Risk}$

No.

Date

$$R_{B}\lambda = \frac{P}{(1+\lambda)^2} + \frac{\lambda^2}{(1+\lambda)^2} \|\mu\|^2$$

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

$\partial f(\mu) = (\mu - y) + \lambda \text{ sigmoid}(\mu)$

$\text{sign}(x) > 0$ if $x > 0$

$\text{sign}(x) < 0$ if $x < 0$

if $\mu_i > 0$

$\mu_i - y_i + \lambda = 0 \Rightarrow \mu_i = y_i - \lambda$

if $y_i > \lambda_i$

if $\mu_i < 0$

$\Rightarrow \mu_i - y_i - \lambda = 0 \Rightarrow \underline{\mu_i = y_i + \lambda}$ if $y_i + \lambda < 0$

$y_i < -\lambda$

if $\mu_i = 0$

$\Rightarrow y_i \in [-\lambda, \lambda]$

$$\mu_i = \left\{ \begin{array}{ll} y_i - \lambda, & \text{if } y_i > \lambda \\ 0, & \text{if } -\lambda \le y_i \le \lambda \\ y_i + \lambda & \text{if } y_i < -\lambda \end{array} \right\}$$

$\mu_i^{soft} = \text{sign}(y_i)(|y_i| - \lambda)_+$

$\hat{\mu}^{soft} = y + g(y)$ $\quad g(y) = \begin{cases} -\lambda, & y_i > \lambda \\ -y & -\lambda < y_i < \lambda \\ \lambda & y_i < -\lambda \end{cases}$

$R(\mu) = E_\mu (P + 2\nabla^T g(y) + \|g(y)\|^2)$

$\nabla^T g(y) = \sum_{i=1}^{P} \frac{\partial g(y)}{\partial y_i}$

No.

Date

$$\frac{\partial g(y)}{\partial y_i} = -I\{|y_i| \le \lambda\}$$

$$R(\hat{\mu}_{soft}) = E_\mu\left(p - 2\sum_{i=1}^{p} I\{|y_i| \le \lambda\} + \sum_{i=1}^{p} \min(\lambda^2, y_i^2)\right)$$

minimize w.r.t λ to get λ_{SURE}

Univariate case, $y = \mu + z \sim N(\mu, 1)$ from Lemma

$$r(\lambda, \mu) \le r(\lambda, 0) + \min(\mu^2, 1+\lambda^2)$$

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

$$r(\lambda, \mu) \le \frac{1}{p} + (2\log p + 1)\min(\mu^2, 1)$$

Lemma 2.9

$$R(\hat{\mu}_{soft}) \le 1 + \sum_{i=1}^{p} \min(\mu_i^2, 1+\lambda^2)$$

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

If μ is sparse then $R(\hat{\mu}_{soft}, \mu) < R(\hat{\mu}_{MLE}, \mu)$

c) $\min_{\mu} \|y - \mu\|^2 + \lambda^2\|\mu\|_0 = \min_{\mu} \sum_{i=1}^{n} \left((y_i - \mu_i)^2 + \lambda^2 \mathbb{1}\{\mu_i \ne 0\}\right)$

$\min_{\mu_i} (y_i - \mu_i)^2 + \lambda^2 \mathbb{1}(\mu_i \ne 0)$

if $\mu_i = 0 \Rightarrow$ cost $= y_i^2$

if $y_i^2 \le \lambda^2$, then set $\mu_i = 0$

if $\mu_i \ne 0 \Rightarrow$ min cost is λ^2 when $\mu_i = y_i$

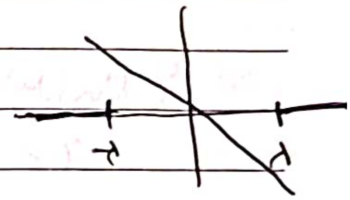
c) $\mu_i = \begin{cases} y_i & \text{if } y_i^2 \ge \lambda^2 \\ 0 & \text{if } y_i^2 \le \lambda^2 \end{cases}$

$\mu_i = \begin{cases} y_i, & \text{if } y_i \ge \lambda \\ 0 & \text{if } -\lambda \le y_i \le \lambda \\ y_i & \text{if } y_i < -\lambda \end{cases}$

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

$\mu_{hard} = y + g(y)$

$\Rightarrow g(y) = \begin{cases} 0 & \text{if } y > \lambda \\ -y & \text{if } -\lambda < y < \lambda \\ 0 & \text{if } y < -\lambda \end{cases}$



g(y) ~~not~~ change suddenly &
not weakly differentiable

(d) $\hat{\mu}^{JS}(y) = \left(1 - \frac{\alpha}{\|y\|^2}\right) y$

$$E\|\hat{\mu}^{JS}(y) - \mu\|^2 = E\left\|y - \frac{\alpha y}{\|y\|^2} - \mu\right\|^2$$

$$= E\left(\|y-\mu\|^2 + 2(y-\mu)^T \frac{\alpha y}{\|y\|^2} + \frac{\alpha^2 \|y\|^2}{\|y\|^4}\right)$$

$$= p + 2\left(\frac{\alpha (y-\mu)^T y}{\|y\|^2}\right) + E\left(\frac{\alpha^2}{\|y\|^2}\right)$$

$$= p + E_\mu\left(\frac{2\alpha\|y\|^2 - 2\alpha \mu^T y + \alpha^2}{\|y\|^2}\right)$$

$$= E_y\, 2\alpha\left(\|y\|^2 - \mu^T y\right)$$

No.

Date

$$E_y\, 2\alpha\left(y^T y - \mu^T y\right) = 2\alpha\,(y - \mu^T)\,y$$

$$= E_y\, 2\alpha\left(\|y-\mu\|^2 - \|\mu\|^2 - \mu^T y\right)\Big]$$

$$= 2\alpha(P-2)$$

$$E\|\mu^{JS}(y) - \mu\|^2 = E\left[P - \frac{2\alpha(P-2) - \alpha^2)}{\|y\|^2}\right]$$

$$M_\alpha(y) = P - \frac{(2\alpha(P-2) - \alpha^2)}{\|y\|^2}$$

$$\frac{\partial M_\alpha(\mu)}{\partial \alpha} = \frac{-(2(P-2) - 2\alpha)}{\|y\|^2} = 0$$

$$\Rightarrow (P-2) - \alpha = 0 \qquad \alpha = P-2$$

$$R^{(\hat{\mu}_{MLE})}_{MLE} = E\|\mu_{MLE} - \mu\|^2$$

$$= E\|y-\mu\|^2 = P$$

$$R_{JS} = P - E\frac{(2\alpha(P-2) - \alpha^2)}{\|y\|^2}$$

$$\text{for } P > 2 \Rightarrow E\left(\frac{2\alpha(P-2) - \alpha^2}{\|y\|^2}\right) > 0$$

$$R_{JS} < P = R_{MLE}$$

$$\underline{R_{JS} < R_{MLE}}$$

No.

Date

(e) Shrinkage rule →

$\theta_\lambda(|t|) \le |t|$

$\theta_\lambda(-t) = \theta_\lambda(t)$

$\theta_\lambda(t) \le \theta_\lambda(t')$ for $t \le t'$

$\lim_{t\to\infty} \theta_\lambda(t) = \infty$

3. $y \sim N(\mu, \sigma^2 I_p)$, $\hat{\mu}_c(y) = Cy$

a) For symmetricity → $A = A^T$

Let D be s.t $I - D = |I - C|$ → D is symmetric

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

= Variance + bias$^2(\hat{\mu})$

For linear estimators, $Var(\hat{\mu}) = Tr(Cov\,\hat{\mu}) = Tr(\sigma^2 C C^T)$

$Var(\hat{\mu}) = \sigma^2 Tr(C^T C)$

Bias $= E\hat{\mu} - \mu = (C - I)\mu$

$$MSE = \sigma^2 Tr(C^T C) + ((C - I)\mu)^2$$

MSE of $\hat{\mu}_D$ is better than μ_C if C is not symmetric

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

Bias^2 is same

Variance, $Tr(D^T D) = Tr\,I - 2Tr(I - D) + Tr[(I - D)^T (I - D)]$

$Tr(D^T D) \le Tr(C^T C)$ iff. $Tr(I - D) = Tr|I - C| \ge Tr(I - C)$

It occurs only if C is not symmetric in which case MSE increases of C

No.

Date

b) Eigen values are also symmetric

$C = U \Lambda U^T$. Let $x = u^T \mu$

$z = u^T y \sim N(x, \sigma^2 I_p)$

Now,

$u^T u = I$

$$E\|Cy - \mu\|^2 = E\|U \Lambda u^T y - \mu\|^2$$

$$= E\|\Lambda z - x\|^2$$

$$r(\hat{\mu}_c, \mu) = \sum_{i=1}^{p} \sigma^2 \lambda_i^2 + (1-\lambda_i)^2 x^2$$

$0 < \lambda_i^2 < 1$ for better MSE

c) $E\|Cy - \mu\|^2 = E\|\Lambda z - x\|^2$

~~Let $\lambda_i < 1 = \lambda d$ for $3 \le d \le i$ Let $x^d = (x_1 \ldots x_d)$~~

Assume that $\lambda_1 = \lambda_2 \ldots \lambda_K(C) = 1$ for $K > 2$

C is symmetric, $C = PDP^{-1}$ - D - diagonal matrix

P - eigenvector matrix

$C^2 = PD^2P^{-1}$

$\lambda_i(C^2) = 1$ for $i = 1, 2 \ldots K$

$Tr(C^2) = Tr(PD^2P^{-1}) = Tr(D^2)$

trace of matrix = sum of eigenvalues

$$Tr(D^2) = \sum_{i=1}^{k} \lambda_i^2 = K$$

$Tr(C^2)$ is also sum of squares of eigenvalues of C

or $\lambda_i < 1 \Rightarrow$ cannot exceed K

Contradiction

No.

Date

4. If $P=1$, $R(\mu^{JS}, \mu) = P - E_\mu \frac{(P-2)^2}{\|Y\|^2}$

$\|Y\|^2$ follows non-central chi squared distribution

$$E\left(\frac{1}{\chi_{P}^{2}}\right) = \frac{1}{P-2} \rightarrow P=1, 2$$

For $P=1 \Rightarrow R(\hat{\mu_{JS}}, \mu) = 1 - \frac{1}{P-2} = 1 - \frac{1}{(-1)} = 2 > 1 = \text{MLE}$

For $P=2$ $R(\hat{\mu_{JS}}, \mu) = 2 \Rightarrow R(\mu_{MLE}, \mu)$

$$R(\mu^{JS}, \mu) = P - E\frac{(P-2)^2}{\|Y\|^2}$$

$$E\left(\frac{1}{\|Y\|^2}\right) = E\left(\frac{1}{P+2N-2}\right) \geq \frac{1}{P+2EN-2} \quad \text{(Jensen)}$$

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

$$R(\mu^{JS}, \mu) \leq P - \frac{(P-2)^2}{P-2+\|\mu\|^2} = 2 + \frac{(P-2)^2\|\mu\|^2}{P-2+\|\mu\|^2}$$

5 a) Marginal distribution

$$P(x) = \int P(x/\theta)\, P(\theta)\, d\theta = \int N(x/\theta, 1) N(\theta/M, A)\, d\theta$$

Product of 2 normal distributions

$$P(x) = N(x \mid M, A+1)$$

ii) Posterior distribution: $P(\theta/x) = \frac{P(\theta)P(x/\theta)}{P(x)}$

$$= \frac{N(x/\theta, 1)\, N(\theta \mid M, A)}{N(x/M, A+1)}$$

No.

Date

~~simplify~~ $\rightarrow$ ~~$P(\theta/x) = N(\theta/M + B(x-M), B)$~~

$$B = \frac{A}{A+1}$$

$$P(\theta/x) = \frac{\exp -\frac{1}{2}(x-\theta)^2 \exp\left(-\frac{1}{2}\frac{(\theta-M)^2}{A}\right)}{\frac{1}{\sqrt{2\pi(A+1)}} \exp\left(-\frac{1}{2}\frac{(x-M)^2}{A+1}\right)}$$

$$= \frac{1}{\sqrt{2\pi B}} \exp\left(-\frac{1}{2}\frac{(\theta-M-B(x-M))^2}{B}\right)$$