HW1 :

3. (a)

If K is positive semi-definite, then its eigenvalue are all nonnegative :

K is $n\times n$ positive semi-definite if for $\forall\, x\in\mathbb{R}^n$, $x^TKx\geqslant 0$ holds.

K is a real symmetric matrix, $K=Q\Lambda Q^T$, where Q is an orthogonal matrix, Λ is a diagonal matrix with eigenvalues of K on the diagonal.

Let $y=Q^Tx$, $x^TKx=(Q^Tx)^T\Lambda(Q^Tx)=y^T\Lambda y$ and $y^T\Lambda y\geqslant 0$, to ensure $y^T\Lambda y\geqslant 0$ then the eigenvalues in Λ must be nonnegative. proof done.

If all eigenvalues of K are nonnegative, then K is positive semi-definite.

We have $K=Q\Lambda Q^T$, where Λ has nonnegative eigenvalues; $x^TKx=y^T\Lambda y$, $\because$ Λ's eigenvalues are nonnegative $\therefore$ $y^T\Lambda y\geqslant 0$ so for $\forall\, x\in\mathbb{R}^n$ $x^TKx\geqslant 0$ which shows that K is positive semidefinite.

3.(b)

The squared Euclidean distance between u_i, u_j is $\|u_i-u_j\|^2=(u_i-u_j)^T(u_i-u_j)=u_i^Tu_i-u_iu_j-u_j^Tu_i+u_j^Tu_j$ math $K_{ii}+K_{jj}-2K_{ij}$, we can get $d_{ij}=K_{ii}+K_{jj}-2K_{ij}$, $K_{ii}=u_i^Tu_i$, $u_i^Tu_j=K_{ij}$, $u_j^Tu_j=K_{jj}$, $\|u_i-u_j\|^2=K_{ii}+K_{jj}-2K_{ij}$.

3. (c)

From $H_a=I-ea^T$, the householder centering matrix, $H_a^2=(I-ea^T)(I-ea^T)=I-2ea^T+ea^Tea^T=I-ea^T$; $H_a^T=I-ae^T$

$B_a=-\frac{1}{2}H_aDH_a^T=-\frac{1}{2}(I-ea^T)D(I-ae^T)=-\frac{1}{2}D+\frac{1}{2}ea^TD+\frac{1}{2}Dae^T-\frac{1}{2}ea^TDae^T$, is symmetric; $x^TB_ax=-\frac{1}{2}x^TDx+\frac{1}{2}(a^TDx)^2-\frac{1}{2}(a^TDx)(x^Ta)$, is positive semi-definite.

3.(d)

We have $x^TAx\geqslant 0$, $x^TBx\geqslant 0$, and $x^T(A+B)=x^TAx+x^TBx\geqslant 0$ $\therefore$ $(A+B)$ is positive semi-definite.

$(A\circ B)_{ij}=A_{ij}\times B_{ij}$, Vectorize A and B, $vec(A)$, $vec(B)$ $A\circ B=vec(A)\cdot vec(B)^T$, $x^T(A\circ B)x\geqslant 0$

4.(a) If d is a distance function, $d(x,y)\geqslant 0$; $d(x,y)=0$ iff. $x=y$; $d(x,y)=d(y,x)$; $d(x,z)\leq d(x,y)+d(y,z)$

$d^2(x,y)=(d(x,y))^2$: ① $d^2(x,y)\geqslant 0$ ② $\because$ $d(x,y)=0$ $\therefore (d(x,y))^2=0$, $\because d^2(x,y)=0$ $\therefore$ $d(x,y)=0$ $\therefore$ $x=y$;

③ $d(x,y)=d(y,x)$, $d^2(x,y)=(d(x,y))^2=(d(y,x))^2=d^2(y,x)$

④ $d^2(x,z)=(x-z)^2$ if $x=0, y=1, z=2$ for example, $d^2(0,2)=4$, $d^2(0,1)=1$, $d^2(1,2)=1$

$\therefore$ $d^2(x,z)>d^2(x,y)+d^2(y,z)$

$\therefore$ d^2 is not a distance function.

(b) $\sqrt{d}$: ① $\sqrt{d(x,y)}\geqslant 0$ ② $\sqrt{d(x,y)}=0$ $d(x,y)=0$, $x=y$ ③ $d(x,y)=d(y,x)$, $\sqrt{d(x,y)}=\sqrt{d(y,x)}$

④ $\sqrt{d(x,y)}\leq\sqrt{d(x,y)+d(y,z)}$

$\therefore$ $\sqrt{d}$ is a distance function.

HW2 :

2.(a) let $\beta = \alpha u$, $x = \beta + \epsilon$. $x \sim N(0, \sigma^2 I_p + \lambda u u^T)$, $\Sigma = \sigma^2 I_p + \lambda u u^T$, $\alpha \sim N(0, \lambda_0)$,

$x_i \in N(0, \sigma^2 I_p + \lambda u u^T)$, $S_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T = \frac{1}{n} vec(X) vec(X)^T$, $y = \Sigma^{-\frac{1}{2}} x \sim N(0, I_p)$

for y, $T_n = \frac{1}{n} vec(y) vec(y)^T$, introduce $y = \Sigma^{-\frac{1}{2}} x$ to the T_n , $T_n = \frac{1}{n} (\Sigma^{-\frac{1}{2}} X)(\Sigma^{-\frac{1}{2}} X)^T = \Sigma^{-\frac{1}{2}} S_n \Sigma^{-\frac{1}{2}}$

then $S_n = \Sigma^{\frac{1}{2}} T_n \Sigma^{\frac{1}{2}}$, $\because$ $S_n \cdot$ eigenvector $= \lambda \cdot$ eigenvector , $\Sigma^{\frac{1}{2}} T_n \Sigma^{\frac{1}{2}} \cdot$ eigenvector $= \lambda \cdot$ eigenvector

$\Rightarrow \Sigma^{\frac{1}{2}} T_n \Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}} T_n \Sigma \Sigma^{-\frac{1}{2}} \Rightarrow T_n \Sigma (\Sigma^{-\frac{1}{2}}$ eigenvector$) = \lambda (\underline{\Sigma^{-\frac{1}{2}} \text{ eigenvector}})$ as the new eigenvector for T_n with Σ.

$SNR > \sqrt{\gamma}$, $\lambda = (1+\lambda_0)(1+\frac{\gamma}{\lambda_0})$, $\lambda_{max}(S_n) = \begin{cases} (1+\gamma)^2 = 6 , & \sigma x^2 \le \sqrt{\gamma} \\ (1+\sigma x^2)(1+\frac{\gamma}{\sigma x^2}) , & \sigma x^2 > \sqrt{\gamma} \end{cases}$

2.(b) From the max eigenvalue, if $\lambda_{max}(S_n) = 6$, $SNR \geq \sqrt{\gamma}$, if $\lambda_{max}(S_n) = (1+\sigma x^2)(1+\frac{\gamma}{\sigma x^2})$, $SNR \geq \sqrt{\gamma}$

2.(c) Calculate $|u^T v|^2 = \frac{1+R-\frac{\gamma}{R}-\frac{\gamma}{R^2}}{1+R+\gamma+\frac{\gamma}{R}} = \frac{1-\frac{\gamma}{R^2}}{1+\frac{\gamma}{R}}$, where $\gamma = \frac{p}{n}$, $R = SNR = \frac{\lambda_0}{\sigma^2}$

HW3.

1.(a) n random $x_1 \cdots x_n \in \mathbb{R}^p$ from a multivariate normal distribution, the joint likelihood function is
$L(\mu, \Sigma \mid x_1 \cdots x_n) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \exp(-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu))$

$\xrightarrow{\log}$ $\ln L(\mu, \Sigma \mid x_1 \cdots x_n) = \sum_{i=1}^{n} \ln f(x_i \mid \mu, \Sigma) = \sum_{i=1}^{n} (-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log\det(\Sigma) - \frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu))$
where $-\frac{n}{2}\log\det(\Sigma)$; $\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) = Tr(\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T)$,
introduce $S_n = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$, $Tr(\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T) = Tr(\Sigma^{-1}S_n)$
Finally, the log-likelihood function $= -\frac{n}{2}Tr(\Sigma^{-1}S_n) - \frac{n}{2}\log(\det(\Sigma)) +$ Constant, here the constant doesn't contain μ and Σ.

1.(b) Approximate $(X+\Delta)^{-1} \approx X^{-1} - X^{-1}\Delta X^{-1}$, then $f(X) = tr(AX^{-1}) \Rightarrow f(X+\Delta) = tr(A(X+\Delta)^{-1})$
$tr(A \cdot (X+\Delta)^{-1}) \approx tr(A \cdot (X^{-1} - X^{-1}\Delta X^{-1})) = tr(AX^{-1} - AX^{-1}\Delta X^{-1}) = tr(AX^{-1}) - tr(AX^{-1}\Delta X^{-1}) = f(X) - tr(AX^{-1}\Delta X^{-1})$

1.(c) $\because \frac{d(\det(X))}{dX} = \det(X)X^{-1}$, $\frac{d\log\det(X)}{dX} = X^{-1}$, $\therefore g(X+\Delta) \approx g(X) + tr(X^{-1}\Delta)$

1.(d) We have $l(\mu,\Sigma) = -\frac{n}{2}tr(\Sigma^{-1}S_n) - \frac{n}{2}\log\det(\Sigma) + C$, $\frac{d\,tr(\Sigma^{-1}S_n)}{d\Sigma^{-1}} = -S_n$, $\frac{d\log\det(\Sigma)}{d\Sigma} = \Sigma^{-1}$,
$\frac{dl(\mu,\Sigma)}{d\Sigma} = -\frac{n}{2}(-\Sigma^{-1}S_n + \Sigma^{-1}) = 0$, $\Sigma^{-1}S_n = \Sigma^{-1}$ the $\Sigma = S_n$ $\therefore \hat{\Sigma}^{MLE} = S_n$

2.(a) $\|y-\mu\|_2^2 = (y-\mu)'(y-\mu) = y'y - 2y'\mu + \mu'\mu$, $\|\mu\|_2^2 = \mu'\mu$, $\therefore \|y-\mu\|_2^2 + \lambda\|\mu\|_2^2 = y'y - 2y'\mu + \mu'\mu + \lambda\mu'\mu$
$= y'y - 2y'\mu + (1+\lambda)\mu'\mu$, the gradient of the objective function $= -2y + 2(1+\lambda)\mu$ $\therefore -2y + 2(1+\lambda)\mu = 0$ $\mu = \frac{1}{1+\lambda}y$.

2.(b) function : $L(\mu) = \frac{1}{2}\|y-\mu\|_2^2 + \lambda\|\mu\|_1$, $\frac{dL(\mu)}{d\mu} = \mu y + \lambda\,\mathrm{sign}(\mu) = 0 \Rightarrow \mu = \begin{cases} y-\lambda, & y>\lambda \\ 0, & -\lambda \le y \le \lambda \\ y+\lambda, & y<-\lambda \end{cases}$

2.(c) function : $L'(\mu) = \|y-\mu\|_2^2 + \lambda^2\|\mu\|_0$, if $\mu=0$ $\|\mu\|_0 = 0$ $L'(\mu) = y^2+\lambda^2$, if $\mu \ne 0$, $\|\mu\|_0 = 1$, $L'(\mu) = (y-\mu)^2 + \lambda^2$
$\hat{\mu} = y_i I(|y_i| > \lambda)$, if $\hat{\mu} = (1-g(y))y \Rightarrow g(y) = I(|y| \le \lambda)$ which is not absolutely continuous, $\therefore g(y)$ is not weakly differentiable.

2.(d) $\hat{\mu}^{JS}(y) - \mu = (1 - \frac{\alpha}{\|y\|^2})y - \mu$, $\|\hat{\mu}^{JS}(y) - \mu\|^2 = \|y-\mu\|^2 - 2\frac{\alpha}{\|y\|^2}y'(y-\mu) + \alpha^2$, take the expectation of both side,
$E\|\hat{\mu}^{JS}(y) - \mu\|^2 = p - \frac{2\alpha(p-2)}{\|\mu\|^2} + \frac{\alpha^2 p}{\|\mu\|^2}$, $\frac{d(p - \frac{2\alpha(p-2)}{\|\mu\|^2} + \frac{\alpha^2 p}{\|\mu\|^2})}{d\alpha} = 0$, $\alpha^* = \arg\min_\alpha U_\alpha(y) = p-2$, $\|y\|^2 \sim \chi^2(\|\mu\|^2, p)$ $N \sim$ Possion $(\frac{1}{2}\|\mu\|^2)$
$E\|y\|^{-2} = E(\frac{1}{p+2N-2}) \ge \frac{1}{p+\|\mu\|^2-2}$ $\therefore$ Risk$(\mu) < p - \frac{(p-2)^2}{p-2+\|\mu\|^2} < p$

2.(e). Above four estimates are shrinkage rules.