Homework 3

2021年3月7日 20:32

1. (a)

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

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

$(tr(AB) = tr(BA))$

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

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

(b)

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

$$= tr\left(A \cdot \left[(I+\Delta X^{-1})X\right]^{-1}\right).$$

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

by first order approximation

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

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

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

Thus $\frac{df}{dX} = -X^{-1}AX^{-1}$

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

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

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

Assum that $X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}} = U\Lambda U^{-1}$

where Λ is diagonal. U is orthogonal.

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

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

$$= g(X) + \operatorname{tr}\Lambda$$

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

1 d)

$$\frac{d\,\ell(\mu, \Sigma)}{d\Sigma} = \frac{n}{2}\Sigma^{-1} S_n \Sigma^{-1} - \frac{n}{2}\Sigma^{-1}$$

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

2.

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

$$\frac{d\ell}{d\mu} = \mu - y + \lambda\mu = 0.$$

$$\Rightarrow \hat{\mu} = \frac{1}{\lambda+1} y.$$

$$\mathbb{E}\|\mu - \hat{\mu}\|^2 = \mathbb{E}\left\|\mu - \frac{1}{\lambda+1} y\right\|_2^2$$

$$= \mathbb{E}\left[\mu^T\mu - \frac{2}{\lambda+1}\mu^T y + \left(\frac{1}{\lambda+1}\right)^2 y^T y\right]$$

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

$$= \mu^T\mu - \frac{[illegible]}{\lambda+1}\mu^T\mu + \frac{[illegible]}{(\lambda+1)}\left(p + \mu^T\mu\right).$$

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

(b)

$$l = \frac{1}{2}\|y-\mu\|^2 + \lambda\|\mu\|_1$$

$$\frac{\partial l}{\partial \mu_i} = \begin{cases} \mu_i - y + \lambda & \mu_i > 0. \\ \mu_i - y - \lambda & \mu_i < 0. \end{cases}$$

Thus. if $\mu_i > y$. $\hat{\mu}_i = y - \lambda$.

$-y < \mu_i \le y$ $\quad \mu_i = 0$.

$\mu_i < -y$ $\quad \hat{\mu}_i = y_i + \lambda$.

Thus. $\hat{\mu}_i = \mathrm{sign}(y_i)\left(|y_i| - \lambda\right)_+$.

By taking $\lambda = \sqrt{2\log p}$

$$\mathbb{E}\|\hat{\mu} - \mu\|^2$$

$$= \sum^{p} \mathbb{E}\left(\hat{\mu}_i - \mu_i\right)^2$$

$$= \sum_{i=1}^{p} \mathbb{E} (\hat{\mu}_i - \mu_i)^2$$

$$= \sum_{i=1}^{p} \int_{\lambda}^{\infty} (y_i - \lambda - \mu_i)^2 \, d\Phi(y_i - \mu_i)$$

$$+ \int_{-\lambda}^{\lambda} \mu_i^2 \, d\Phi(y_i - \mu_i)$$

$$+ \int_{-\infty}^{-\lambda} (y_i + \lambda - \mu_i) \, d\Phi(y_i - \mu_i).$$

Φ is the CDF of Gaussian distribution.

$$\text{risk} = \int_{\lambda-\mu_i}^{\infty} (x-\lambda)^2 \, d\phi(x) + \int_{\lambda-\mu_i}^{\lambda-\mu_i} \mu_i^2 \, d\phi(x) + \int_{-\infty}^{-\lambda-\mu_i} (x+\lambda) \, d\phi(x)$$

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

(c)

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

$$= \sum_{i=1}^{p} \left((y_i - \mu_i)^2 + \lambda^2 \cdot I(\mu_i \neq 0) \right)$$

$\Pi \quad \hat{\mu}_i = \mu_i \, I(|\mu_i| \quad \lambda)$

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

$$\hat{\mu} = y \cdot \left(1 - I(|y_i| \leq \lambda)\right).$$

$$g(y)_i = \begin{cases} 0 & |y_i| > \lambda \\ 1 & |y_i| \leq \lambda. \end{cases}$$

Thus $g(y)$ is differentiable

except. $\{ y \mid y_i = \lambda, \forall i = 1, \dots, p \}$.

$g(y)$ is weakly differentiable.

(d)

$$\mathbb{E} \| \hat{\mu} - \mu \|^2$$

$$= \mathbb{E} \left\| y - \frac{\alpha}{\|y\|} y - \mu \right\|^2$$

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

By SURE Lemma.

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

Thus

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

$$= U_\alpha(y)$$

when $p > 2$.

$$U_\alpha(y) < p.$$

Thus, risk is smaller than MLE.

(e) [odd], [monotone], [unbounded].

3.

By definition.

$$R(\hat{\mu}_C, \mu) = \sigma^2 \operatorname{tr}(C^T C) + \|(I-C)\mu\|^2$$

(a) if C is not symmetric.

Let $C^* = I - \left((I-C)^T(I-C)\right)^{\frac{1}{2}} = I - |I-C|$

Then.

$$(I - C^*)^2 = (I - C)^2$$

$$\mathrm{tr}\left((I - C^*)^T (I - C)\right)$$

$$= n - 2\,\mathrm{tr}(I - C^*) + \mathrm{tr}\left(|I - C|^T (I - C)\right).$$

$$= \mathrm{tr}(C^T C) + 2\,\mathrm{tr}\left[I - C - |I - C|\right].$$

$$< \mathrm{tr}(C^T C).$$

Thus not admissible.

(b)

Let $C = U \Lambda U^T$.

$$\Lambda = (\rho_1, \rho_2, \cdots, \rho_p).$$

$$R(\hat{\mu}_C, \mu) = \sum_{p=1}^{n} \left(\sigma^2 \rho_i^2 + (1 - \rho_i)^2 \mu_i^2 \right).$$

$$\geq \sum_{p=1}^{n} \left(\sigma^2 g(\rho_i) + (1 - g(\rho_i))^2 \mu_i^2 \right).$$

$$g(\rho_i) = \begin{cases} 0 & \rho_i < 0 \\ \rho_i & \rho_i \leq 1 \\ 1 & \rho_i > 1. \end{cases}$$

If and only if $\rho_i \in [0, 1]$.

$\therefore C = C_{x}..$

(c) If $\rho_i = \rho_j = \rho_k = 1$.

Since J-S estimater is better than MLE when $p > 2$. thus.

for dimension i, j, k. we can use J-S estimator to estimate.