SUBHODIP SAHA
5485317
HW2
saha0073@umn.edu

$\theta > 0$

1/a $p(x|\theta) = \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{x^2}{2\theta^2}\right)$

$L(\theta|x) = \log\left(\prod_{i=1}^{N} p(x|\theta)\right)$

$= \sum_{i=1}^{N} \log p(x|\theta)$

$= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{N} \log\left(\frac{1}{\theta} \exp(-x^2/2\theta^2)\right)$

$\frac{\partial L(\theta|x)}{\partial\theta} = \frac{1}{\sqrt{2\pi}} \frac{\partial}{\partial\theta} \sum_{i=1}^{N} \log\left(\frac{1}{\theta} \exp(-x^2/2\theta^2)\right)$

$= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{N} \frac{\partial}{\partial\theta} \left(\log\frac{1}{\theta} - x^2/2\theta^2\right)$

$= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{N} \left[-\frac{1}{\theta} - \frac{1}{2}(-2)\, x^2/\theta^3\right]$

$= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{N} \left(-\frac{1}{\theta_{ML}} + x^2/\theta_{ML}^3\right) = 0$ for MLE estimation

$\frac{N}{\theta_{ML}} = \sum_{i=1}^{N} x^2/\theta_{ML}^3$

$N\theta_{ML}^2 = \sum_{i=1}^{N} x^2$

$\boxed{\theta_{ML}^2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2} \to \theta_{ML} = \left(\frac{1}{N} \sum_{i=1}^{N} x_i^2\right)^{1/2}$

$\underline{b}$ $P(x|\theta) = \frac{1}{\theta} \exp(-x/\theta)$

$L(\theta|x) = \sum_{i=1}^{N} \log \prod_{i=1}^{N} \left(\frac{1}{\theta} \exp(-x/\theta)\right)$

$\Rightarrow \frac{\partial}{\partial\theta} L(\theta|x) = \sum_{i=1}^{N} \frac{\partial}{\partial\theta} \log\left(\frac{1}{\theta} \exp(-x/\theta)\right)$

$\Rightarrow \frac{\partial}{\partial\theta} L(\theta|x) = \sum_{i=1}^{N} \left(-1/\theta + x/\theta^2\right) = 0$ — for MLE estim

$\Rightarrow \sum_{i=1}^{N} +1/\theta_{ML} = \sum_{i=1}^{N} x_i/\theta_{ML}^2$

$\Rightarrow \frac{N}{\theta_{ML}} = \frac{1}{\theta_{ML}^2} \sum_{i=1}^{N} x_i$

$\Rightarrow \boxed{\theta_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i}$

$\underline{c}$ $P(x|\theta) = \theta x^{\theta-1}$

$L(x|\theta) = \cancel{\prod_{i=1}^{N} \log(\theta x_i^{\theta-1})}$ $\log \prod_{i=1}^{N} (\theta x_i^{\theta-1})$

$= \sum_{i=1}^{N} \log(\theta x_i^{\theta-1})$

$\frac{\partial L(x|\theta)}{\partial\theta} = \sum_{i=1}^{N} \frac{\partial}{\partial\theta} \log(\theta x_i^{\theta-1})$

$= \sum_{i=1}^{N} \frac{\partial}{\partial\theta} \left(\log\theta + (\theta-1)\log x_i\right)$

1'

$$\frac{\partial L(x|\theta)}{\partial \theta} = \sum_{i=1}^{N} \left(\frac{1}{\theta} + \log x_i \right) = 0 \rightarrow \text{for MLE}$$

$$\frac{N}{\theta_{ML}} = -\sum_{i=1}^{N} \log x_i$$

$$\rightarrow \boxed{\theta_{ML} = -\frac{N}{\sum_{i=1}^{N} \log x_i}}$$

$$d \quad P(x|\theta) = \frac{1}{\theta}$$

$$L(x|\theta) = \log \sum_{i=1}^{N} P(x|\theta)$$

$$= \log \int_{0}^{\theta} P(x|\theta)\, dx$$

$$= \log \int_{0}^{\theta} \frac{1}{\theta}\, dx$$

$$\frac{\partial L(x|\theta)}{\partial \theta} =$$

Here we can change the ~~[illegible]~~ summation to $\int dx$

$$L(x|\theta) = \log \prod_{i=1}^{N} P(x|\theta)$$

$$= \sum_{i=1}^{N} \log P(x|\theta)$$

$$= \int_{0}^{\theta} \log P(x|\theta)\, dx$$

we can ~~[illegible]~~ change the $\sum_{i=1}^{N} \rightarrow \int_{0}^{\theta} dx$

$$\frac{\partial}{\partial \theta} L(x|\theta) = \frac{\partial}{\partial \theta} \int_0^\theta \log p(x|\theta)\, dx$$

$$= \frac{\partial}{\partial \theta} \int_0^\theta \log\left(\frac{1}{\theta}\right) dx$$

$$= -\frac{\partial}{\partial \theta} \int_0^\theta \log \theta \, dx$$

$$= -\frac{\partial}{\partial \theta}(\theta \log \theta)$$

$$= -\log \theta_{ML} - \frac{\theta}{\theta} = 0 \rightarrow \text{for MLE}$$

$$\Rightarrow -\log \theta_{ML} = 1$$

$$\Rightarrow \log \theta_{ML} = -1$$

$$\Rightarrow \boxed{\theta_{ML} = e^{-1}}$$

---

d) $L(\theta|x) = \frac{1}{\theta^n}$ is monotonically decreasing for $\theta > 0$, which indicates smallest value of θ maximizes $L(\theta|x)$.

As we also know $\theta \geq x$, so θ should be the largest of the set $\{x_i | i = 1, \ldots N\}$

$$\therefore \boxed{\hat{\theta}_{MLE} = \text{largest of } \{x_i | i = 1, \ldots N\}}$$

2) $p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$ $\underline{2°}$

$L(x|\mu,\Sigma) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)$

$\frac{\partial L}{\partial \mu} L(x|\mu,\Sigma) = \frac{\partial}{\partial\mu}\cancel{\left(\frac{N}{2}\log(2\pi)\right)}^{\to 0} - \frac{\partial}{\partial\mu}\cancel{\left(\frac{N}{2}\log|\Sigma|\right)}^{\to 0}$

$- \frac{1}{2}\frac{\partial}{\partial\mu}\left(\sum_{i=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right)$

$\frac{\partial}{\partial\mu_n}\left[(x_k-\mu_k)\Sigma^{-1}_{km}(x_m-\mu_m)\right]$

$(x_k-\mu_k)\Sigma^{-1}_{km}\delta_{mn}$

$(x_m-\mu_m) - (x_k-\mu_k)\Sigma^{-1}_{km}\delta_{mn}$

$= -\delta_{kn}\Sigma^{-1}_{km}(x_m-\mu_m)$

$-\Sigma^{-1}_{nm}(x_m-\mu_m) - (x_k-\mu_k)\Sigma^{-1}_{kn}$

$= -\Sigma^{-1}_{nm}(x_m-\mu_m) - (x_k-\mu_k)\Sigma^{-1}_{nk}$ → symmetric

$= -\Sigma^{-1}_{nm}(x_m-\mu_m)$

$\cancel{= -2\Sigma^{-1}_{nk}(x_k-\mu_k)}$

$= -2(x_k-\mu_k)\Sigma^{-1}_{nk}$

$\frac{\partial L}{\partial\mu_n} = \sum_{i=1}^{N}2(x_i-\mu)^T\Sigma^{-1} = 0$

$= \sum_{i=1}^{N}(x_i^T-\mu^T)\Sigma^{-1} = 0$

$\Rightarrow \sum_{i=1}^{N}x_i^T\Sigma^{-1} - \mu^T\Sigma^{-1} = 0$

$\Rightarrow \sum_{i=1}^{N}x_i^T\Sigma^{-1} = N\mu^T\Sigma^{-1}$

Scanned by CamScanner

$\boxed{\hat{\mu} = \frac{1}{N}\sum_{i}^{N} x_i}$

The ~~MLE~~ MLE of mean $\hat{\mu} = \frac{1}{N}\sum_{i}^{N} x_i$

$$L(x|\mu,\Sigma) = -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)$$

we need

$$\frac{\partial L}{\partial \Sigma^{-1}} = -\frac{\partial}{\partial \Sigma}\left(\frac{N}{2}\log|\Sigma|\right) - \frac{1}{2}\frac{\partial}{\partial \Sigma}\sum_{i=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)$$

we need Some identity to derive this.

$\to$ we need $\text{tr}[x^T A x] = \text{tr}[x x^T A]$

$\underbrace{x^T A x}_{\text{Scalar}} = \text{tr}$

$$\frac{\partial}{\partial a_{ij}}\text{tr}[AB] = \frac{\partial}{\partial a_{ij}}\sum_{k,l} a_{kl} b_{lk} = \delta_{ik}\delta_{jl} b_{lk} = b_{ji} = B^T$$

$$\frac{\partial}{\partial A}\text{tr}[BA] = B^T$$

$$\frac{\partial}{\partial A}(x^T A x) = \frac{\partial}{\partial A}\text{tr}(x x^T A) = (x x^T)^T = x x^T \quad \ldots (i)$$

$$\frac{\partial}{\partial A}\log|A| = \frac{1}{|A|}\frac{\partial}{\partial a_{ij}}|A|$$

$$A^{-1} = \frac{1}{|A|}\tilde{A} \to \text{Matrix of cofactors.}$$

$\frac{\partial}{\partial a_{ij}}$ ~~|A|~~ $=$

2'

Now we need to prove,

$$\frac{\partial}{\partial a_{ij}} |A| = \tilde{A} \quad \ldots (ii)$$

$$|A| = \sum_{j} (-1)^{i+j} a_{ij} \underbrace{M_{ij}}_{\text{minor}}$$

$$\frac{\partial |A|}{\partial a_{ij}} = \sum_{j} (-1)^{i+j} m_{ij} = \tilde{A}$$

~~L(Σ)~~

$$L(x|\mu, \Sigma) = \frac{N}{2} \log|\Sigma^{-1}| - \frac{1}{2} \sum_{n} \text{tr} \left[(x_n - \mu)(x_n - \mu)^T \Sigma^{-1} \right]$$

$$\frac{\partial L}{\partial \Sigma^{-1}} = \frac{N}{2} \Sigma_{ML} - \frac{1}{2} \sum_{i=1}^{n} (x_n - \mu)(x_n - \mu)^T = 0$$

$$\boxed{\hat{\Sigma}_{ML} = \frac{1}{N} \sum_{i=1}^{N} (x_n - \mu)(x_n - \mu)^T}$$

b/

$$E[\hat{\mu}_n] = E\left[\frac{\sum_{i=1}^{N} x_i}{N} \right] = \frac{1}{N} \sum_{i=1}^{N} E(x_i) = \frac{\sum_{i=1}^{N} \mu}{N} = \frac{\cancel{N} \mu}{\cancel{N}} = \mu$$

$$\boxed{E[\hat{\mu}_n] = \mu}$$

For a particular sample $\hat{\mu}_n$ might be different from μ, ~~But~~ but their average will get close to μ as the # of such sample increases.

So, $\hat{\mu}_n$ is an unbiased estimator.

c $E[\hat{\Sigma}]$

$$\Sigma = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} = \frac{\sum_{i=1}^{N} x_i^2 + \mu^2 - 2\mu \sum_{i=1}^{N} x_i}{N}$$

$$= \frac{\sum_i x_i^2 - N\mu^2}{N}$$

$$E[\Sigma] = \frac{\sum_i^N E(x_i^2) - N\, E[\mu^2]}{N}$$

Let's derive for single variable, we can generalize that for multi variable case (as dimensionality should not change variance)

$$E[s^2] = \frac{\sum_{i=1}^{N} E[x_i^2] - N.E[\mu^2]}{N}$$

$$Var[X] = E[X^2] - E[X]^2$$

$$E[X^2] = Var[X] + E[X]^2$$

$$E[(x_i)^2] = \sigma^2 + \mu^2 \quad \& \quad E[\mu^2] = \sigma^2/N + \mu^2$$

$$E[s^2] = \frac{N(\sigma^2+\mu^2) - N(\sigma^2/N + \mu^2)}{N} = \left(\frac{N-1}{N}\right)\sigma^2$$

for multivariable,

$$\boxed{E[\hat{\Sigma}_{ML}] = \left(\frac{N-1}{N}\right)\hat{\Sigma}}$$

As, $E[\hat{\Sigma}_{ML}] \neq \hat{\Sigma}$ ⇝ it is an biased estimator.

so, even if sample size increase $\hat{\Sigma}_{ML}$ will not approach true variance, so biased estimator.

3/a 30

	C_1	C_2	C_3	Reject
C_1	0	1	1	λ
C_2	10	0	10	λ
C_3	100	100	0	λ

$\lambda = 10$

Risk of reject $R(\alpha_{K+1}|x) = \sum_{k=1}^{K} \lambda P(C_k|x)$

$= \lambda P(C_1|x) + \lambda P(C_2|x) + \lambda P(C_3|x)$

$= \lambda(0.5 + 0.25 + 0.25) = \lambda = 10$

Risk of choosing C_i

$R(\alpha_i|x) = \sum_{k=1}^{K} \lambda_{ik} P(C_k|x)$

$R(C_1|x) = \lambda_{11} P(C_1|x) + \lambda_{12} P(C_2|x) + \lambda_{13} P(C_3|x)$

$= 0 + 10 \times 0.25 + 100 \times 0.25 = 2.5 + 25 = 27.5$

$R(C_2|x) = \lambda_{21} P(C_1|x) + \lambda_{22} P(C_2|x) + \lambda_{23} P(C_3|x)$

$= 1 \times 0.5 + 0 + 100 \times 0.25 = 0.5 + 25 = 25.5$

$R(C_3|x) = \lambda_{31} P(C_1|x) + \lambda_{32} P(C_2|x) + 0$

$= 1 \times 0.5 + 10 \times 0.25 + 0 = 0.5 + 2.5 = 3.$

So, $\boxed{\text{Choose class 3}}$

as, $\underset{3}{R(C_3|x)} < \underset{27.5}{R(C_1|x)}$, $\underset{3}{R(C_3|x)} < \underset{25.5}{R(C_2|x)}$

$\underset{3}{R(C_3|x)} < \underset{10}{R(\alpha_{K+1}|x)}$

b $\lambda = 5$

$R(\alpha_{K+1}|x) = \lambda(0.4 + 0.5 + 0.1) = \lambda = 5$

$R(C_1|x) = \lambda_{12} P(C_2|x) + \lambda_{13} P(C_3|x)$

$= 10 \times 0.5 + 100 \times 0.1 = 5 + 10 = 15$

$R(C_2|x) = \lambda_{21} P(C_1|x) + \lambda_{23} P(C_3|x)$

$= 1 \times 0.4 + 100 \times 0.1 = 0.4 + 10 = 10.4$

$R(C_3|x) = \lambda_{31} P(C_1|x) + \lambda_{32} P(C_2|x)$

$= 1 \times 0.4 + 10 \times 0.5 = 0.4 + 5 = 5.4$

$\boxed{\text{choose reject}}$

$\underset{5}{R(\alpha_{K+1}|x)} < \underset{15}{R(C_1|x)}$, $\underset{5}{R(\alpha_{K+1}|x)} < \underset{10.4}{R(C_2|x)}$

$\underset{5}{R(\alpha_{K+1}|x)} < \underset{5.4}{R(C_3|x)}$.

Problem 3

1. Error rates for MGC with full cov matrix on Boston 50

F1	F2	F3	F4	F5	Mean	SD
0.2475	0.2079	0.2079	0.1881	0.1782	0.2059	0.0237

2. Error rates for MGC with full cov matrix on Boston 75

F1	F2	F3	F4	F5	Mean	SD
0.1980	0.3069	0.2673	0.2970	0.2376	0.2613	0.0398

3. Error rates for MGC with full cov matrix on Digits

F1	F2	F3	F4	F5	Mean	SD
0.0362	0.0389	0.0835	0.0445	0.0389	0.0484	0.0177

4. Error rates for MGC with diagonal cov matrix on Boston50

F1	F2	F3	F4	F5	Mean	SD
0.2277	0.1980	0.1584	0.2277	0.3168	0.2257	0.0521

5. Error rates for MGC with diagonal cov matrix on Boston75

F1	F2	F3	F4	F5	Mean	SD
0.2871	0.4158	0.2673	0.2277	0.2871	0.2970	0.0632

6. Error rates for MGC with diagonal cov matrix on Digits

F1	F2	F3	F4	F5	Mean	SD
0.1281	0.0947	0.1086	0.1058	0.1197	0.1114	0.0115

7. Error rates for LogisticRegression with Boston50

F1	F2	F3	F4	F5	Mean	SD
0.1386	0.1287	0.1485	0.1089	0.1485	0.1346	0.0148

8. Error rates for LogisticRegression with Boston75

F1	F2	F3	F4	F5	Mean	SD
0.0990	0.1584	0.1683	0.0792	0.0297	0.1069	0.0514

9. Error rates for LogisticRegression with Digits

F1	F2	F3	F4	F5	Mean	SD
0.0334	0.0362	0.0696	0.0222	0.0389	0.0401	0.0158