5521 hw 2

note. $\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdots x_n$

$\mathcal{X} = \{x_1, \dots, x_n\}$

1 (a) $L(\theta) = \prod_{i=1}^{n} P(X = x_i) = \left(\frac{1}{\sqrt{2\pi}\,\theta}\right)^n \exp\left(-\frac{1}{2\theta^2} \sum_{i=1}^{n} X_i^2\right)$

$\log L(\theta) = n \log\left(\frac{1}{\sqrt{2\pi}\,\theta}\right) - \frac{1}{2\theta^2} \sum_{i=1}^{n} X_i^2 = -n \log(\sqrt{2\pi}\,\theta) - \frac{1}{2\theta^2} \sum_{i=1}^{n} X_i^2$

$\frac{\partial \log L(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^3} \sum_{i=1}^{n} X_i^2$

$\frac{\partial \log L(\theta)}{\partial \theta} = 0 \iff \frac{1}{\theta^3} \sum_{i=1}^{n} X_i^2 = \frac{n}{\theta} \iff \theta^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} \iff \theta = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}$

(b) $L(\theta) = \prod_{i=1}^{n} P(X_i = x_i) = \left(\frac{1}{\theta}\right)^n \exp\left(-\frac{\sum_{i=1}^{n} X_i}{\theta}\right)$

$\log L(\theta) = -n \log \theta - \frac{1}{\theta} \sum_{i=1}^{n} X_i$

$\frac{\partial \log L(\theta)}{\partial \theta} = 0 \iff -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} X_i = 0 \iff \theta = \frac{\sum_{i=1}^{n} X_i}{n}$

(c) $L(\theta) = \prod_{i=1}^{n} P(X_i = x_i) = \theta^n \left[\prod_{i=1}^{n} x_i\right]^{\theta - 1}$

$\log L(\theta) = n \log \theta + (\theta - 1) \log \prod_{i=1}^{n} x_i$

$\frac{\partial \log L(\theta)}{\partial \theta} = 0 \iff \frac{n}{\theta} + \log \prod_{i=1}^{n} x_i = 0 \iff \theta = -\frac{n}{\log \prod_{i=1}^{n} x_i}$

(d) $L(\theta) = \prod_{i=1}^{n} P(X_i = x_i) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}[x_i] = \begin{cases} \frac{1}{\theta^n}, & 0 \le X_i \le \theta \\ 0, & \text{otherwise} \end{cases}$

i. $L = \frac{1}{\theta^n}$ if $\theta \ge X_n$

ii. $L = 0$ if $\theta < X_n$

L is maximized at $\theta = X_n$ since $\frac{1}{\theta^n}$ is decreasing as θ increases

it should be the smallest value of θ such that $\theta \ge X_i$,

so $\hat{\theta} = \max(X_1, \dots, X_n)$

2. $L(\mu, \Sigma) = \prod_{i=1}^{n} P(X_i = x_i) = (2\pi)^{-\frac{dn}{2}} \cdot (\Sigma)^{-n/2} \cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)\right]$

(a) $\log L(\mu, \Sigma) = -\frac{dn}{2}\log(2\pi) - \frac{n}{2}\log\Sigma - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$ $\circledast$

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

$$\frac{\partial \log L(\mu,\Sigma)}{\partial \Sigma} = 0 \Leftrightarrow -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-2}(x_i-\mu) = 0 \Leftrightarrow \Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$$

(b) $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, $E[\hat{\mu}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n}E(x_i) = \frac{1}{n}\cdot n\cdot \mu = \mu$

where $\mu = E(x_i)$ for $i = 1, 2, \ldots, n$, the mean of the estimate of the mean and the true mean converge, so $\hat{\mu}_n$ is a unbiased estimate of the true mean μ.

(c) $\hat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$, $E[\hat{\Sigma}_n] = \frac{1}{n}E\left[\sum_{i=1}^{n} x_i x_i^T - 2\sum_{i=1}^{n} x_i \mu + \sum_{i=1}^{n}\mu^2\right]$

$= \frac{1}{n}E\left[\sum_{i=1}^{n} x_i x_i^T - n\mu^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} x_i x_i^T\right] - E(\mu^2) = E(x x^T) - E(\mu^2)$ $\oplus$

$\Sigma_x = E(x x^T) - (E(x))^2$ $\quad \Sigma_\mu = E(\mu^2) - (E(\mu))^2$

$(=)$ $= \Sigma_x + (E(x))^2 - (\Sigma_\mu + (E(\mu))^2)$

$= \Sigma_x - \Sigma_\mu$

$= \Sigma_x - \frac{1}{N^2}\cdot N \cdot \Sigma_x$

$= \frac{N-1}{N}\Sigma_x \neq \Sigma_x$

the covariance of the estimate and the true covariance don't converge, so $\hat{\Sigma}_n$ is a biased estimate of the true covariance Σ_n

Q3

Summary

MultiGaussClassify with Boston50

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.326732673267	0.29702970297	0.0990099009901	0.346534653465	0.117647058824	0.237390797903	0.106713622914

MultiGaussClassify with Boston75

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.386138613861	0.138613861386	0.39603960396	0.445544554455	0.0392156862745	0.281110463988	0.161305024613

MultiGaussClassify with Digits

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.161559888579	0.194986072423	0.177777777778	0.105849582173	0.147222222222	0.157479108635	0.0303521755008

LogisticRegression with Boston50

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.128712871287	0.108910891089	0.0891089108911	0.277227722772	0.107843137255	0.142360706659	0.0685878735016

LogisticRegression with Boston75

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.0891089108911	0.128712871287	0.138613861386	0.108910891089	0.0490196078431	0.102873228499	0.0318471419286

LogisticRegression with Digits

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.0696378830084	0.111420612813	0.0444444444444	0.041782729805	0.0111111111111	0.0556793562365	0.033489091051