1.

山东财经大学实验报告

学院：　　　　班级：　　　　姓名：　　　　学号：　　　　年　　月

课程名称：

实验名称：

实验内容：

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

$L(\theta|X) = \prod_{t=1}^{N} p(x^t|\theta) = \left(\frac{1}{\sqrt{2\pi}\theta}\right)^n \exp\left(-\frac{(X_1-2)^2}{2\theta^2} - \frac{(X_2-2)^2}{2\theta^2} \cdots - \frac{(X_n-2)^2}{2\theta^2}\right)$

$\log L(\theta|X) = \log\left(\frac{1}{\sqrt{2\pi}\theta}\right)^n + \left(-\frac{(X_1-2)^2+(X_2-2)^2+\cdots(X_n-2)^2}{2\theta^2}\right)$

$\frac{\partial \log L(\theta|X)}{\partial \theta^2} = 0 \Rightarrow \hat{\theta}^2 = \frac{1}{n}\sum_{t=1}^{N}(X_t-2)^2$

(b) $p(X|\theta) = \frac{1}{\theta}\exp\left(-\frac{X}{\theta}\right)$

$L(\theta|X) = \prod_{t=1}^{N} p(x^t|\theta) = \frac{1}{\theta^N}\exp\left(-\frac{\sum_{t=1}^{N} X_t}{\theta}\right)$

$\log L(\theta|X) = -\log\theta^N - \frac{\sum_{t=1}^{N} X_t}{\theta}$

$\frac{\partial \log L(\theta|X)}{\partial \theta} = 0 \Rightarrow \hat{\theta} = \frac{1}{N}\sum_{t=1}^{N} X_t = \bar{X}$

(c) $p(X|\theta) = \frac{1}{2\theta^3}X^2\exp\left(-\frac{X}{\theta}\right)$

$L(\theta|X) = \prod_{t=1}^{N} p(x^t|\theta) = \left(\frac{1}{2\theta^3}\right)^N \prod_{t=1}^{N} X_t^2 \exp\left(-\frac{\sum_{t=1}^{N} X_t}{\theta}\right)$

$\log L(\theta|X) = N\log\left(\frac{1}{2\theta^3}\right) + \log\left(\prod_{t=1}^{N} X_t^2\right) + \left(-\frac{\sum_{t=1}^{N} X_t}{\theta}\right)$

$\frac{\partial \log L(\theta|X)}{\partial \theta} = 0 \Rightarrow \hat{\theta} = \frac{1}{3N}\sum_{t=1}^{N} X_t = \frac{1}{3}\bar{X}$

(d) $p(X|\theta) = \theta X^{\theta-1}$

$L(\theta|X) = \theta^N \prod_{t=1}^{N} X_t^{\theta-1}$

$\log L(\theta|X) = N\log\theta + (\theta-1)\log\prod_{t=1}^{N} X_t$

$\frac{\partial \log L(\theta|X)}{\partial \theta} = 0 \Rightarrow \hat{\theta} = \frac{N}{\log\prod_{t=1}^{N} X_t}$

$$\cancel{\frac{\partial \log L(\theta|X)}{\partial \theta} = -\frac{n}{\theta} \neq 0 \Rightarrow \hat{\theta} = 0}$$

(e) $p(x|\theta) = \begin{cases} \frac{1}{\theta} & 0 < x \le \theta \\ 0 & \text{o.w.} \end{cases}$

$L(\theta|X) = \prod_{i=1}^{n} \frac{1}{\theta} I(0 < X_i \le \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 \le X_i \le \theta)$

Since $\frac{1}{\theta^n}$ decreasing when θ increasing

We should get the smallest value of θ

Therefore, $\hat{\theta} = \max\{X_1, \cdots, X_n\}$

2.

(b)

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

$$L(\mu,\Sigma) = (2\pi)^{-\frac{dN}{2}} \cdot |\Sigma|^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\sum_{t=1}^{N}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]$$

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

$$\frac{\partial \log L(\mu,\Sigma)}{\partial \mu} = 0 \Rightarrow \frac{\partial\left[-2\sum_{t=1}^{N}(x^t)^T\Sigma^{-1}\mu\right]}{\partial \mu} + \frac{\partial(N\mu^T\Sigma^{-1}\mu)}{\partial \mu} = 0$$

$$-2\sum_{t=1}^{N}\Sigma^{-1}x^t + 2N\Sigma^{-1}\mu = 0$$

$$\hat{\mu} = \frac{1}{N}\sum_{t=1}^{N}x^t = \bar{x}$$

$$\frac{\partial \log L(\mu,\Sigma)}{\partial \Sigma} = 0 \Rightarrow \frac{\partial\left[\log|\Sigma| + \frac{1}{N}\sum_{t=1}^{N}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]}{\partial \Sigma} = 0$$

$$d\log L(\mu,\Sigma) = tr(\Sigma^{-1}d\Sigma) + tr\left(-\Sigma^{-1}\frac{1}{N}\sum_{t=1}^{N}(x^t-\mu)(x^t-\mu)^T\Sigma^{-1}d\Sigma\right)$$

$$d\log L(\mu,\Sigma) = tr\left(\left(\Sigma^{-1}-\Sigma^{-1}\frac{1}{N}\sum_{t=1}^{N}(x^t-\mu)(x^t-\mu)^T\Sigma^{-1}\right)d\Sigma\right)$$

$$\frac{\partial \log L(\mu,\Sigma)}{\partial \Sigma} = \left(\Sigma^{-1}-\Sigma^{-1}\frac{1}{N}\sum_{t=1}^{N}(x^t-\mu)(x^t-\mu)^T\Sigma^{-1}\right)^T = 0$$

$$\hat{\Sigma} = \frac{1}{N}\sum_{t=1}^{N}(x^t-\mu)(x^t-\mu)^T$$

since $\hat{\mu} = \bar{x}$

$$\hat{\Sigma} = \frac{1}{N}\sum_{t=1}^{N}(x^t-\bar{x})(x^t-\bar{x})^T$$

(b) $E[\hat{\mu}_n] = E[\frac{1}{N}\sum_{t=1}^{N} x^t] = \frac{1}{N}\sum_t E[x^t] = \frac{N\mu}{N} = \mu$

Therefore, $\hat{\mu}_n$ is unbiased.

(c) $E(\hat{\Sigma}_n) = E[\frac{1}{N}\sum_{t=1}^{N} (x^t - \bar{x})(x^t - \bar{x})^T]$

$$= \frac{1}{N} E[\sum_{t=1}^{N} (x^t - \bar{x})(x^t - \bar{x})^T]$$

$$= \frac{1}{N}\sum_{t=1}^{N} E[x_t x_t^T] - N E[\mu\mu^T]$$

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

Therefore, $\hat{\Sigma}_n$ is ~~unbia~~ biased estimate.

3.

Error rates for MultiGaussClassify with full covariance matrix on Boston50						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.45544554455445544	0.504950495049505	0.1089108910891089	0.40594059405940597	0.17821782178217827	0.3306930693069307	0.15749733722889644

Error rates for MultiGaussClassify with full covariance matrix on Boston25						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.16831683168316836	0.8217821782178218	0.0	0.3366336633663366	0.4356435643564357	0.3524752475247525	0.2776940914577515

Error rates for MultiGaussClassify with full covariance matrix on Digits						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.8941504178272981	0.9025069637883009	0.9080779944289694	0.8969359331476323	0.9025069637883009	0.9008356545961002	0.0048567119148085576

Error rates for MultiGaussdiagClassify with diagonal covariance matrix on Boston50						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.42574257425742578	0.504950495049505	0.8910891089108911	0.40594059405940597	0.17821782178217827	0.4811881188118813	0.2319960652611108

Error rates for MultiGaussdiagClassify with diagonal covariance matrix on Boston25						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.85148514851485155	0.17821782178217827	0.0	0.3366336633663366	0.5643564356435644	0.38613861386138615	0.29768903719550316

Error rates for MultiGaussClassify with diagonal covariance matrix on Digits						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.89972144846796 65	0.8997214484679665	0.9025069637883009	0.8969359331476323	0.8969359331476323	0.8991643454038997	0.0020844887948601384

Error rates for LogisticRegression with Boston50						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.1386138613861386 3	0.13861386138613863	0.09900990099009899	0.27722772277227725	0.09900990099009899	0.1504950495049505	0.065795040498361

Error rates for LogisticRegression with Boston25						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.1584158415841584 5	0.2178217821782178	0.03960396039603964	0.12871287128712872	0.13861386138613863	0.13663366336633664	0.05752807543102552 6

Error rates for LogisticRegression with Digits						
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	SD
0.06963788300835 65	0.11977715877437323	0.058495821727019504	0.06128133704735372	0.0	0.06183844011142059	0.03808727495295199