CSCI 5521 – Introduction to Machine Learning

Homework – 2

Mourya Karan Reddy Baddam – 5564234

1) a)

$$P(x/\theta) = \frac{1}{\sqrt{2\pi}\ \theta} \cdot \exp\left(-\frac{x^2}{2\theta^2}\right), \quad \theta > 0.$$

$$\Rightarrow \prod_{i=1}^{n} P(x_i/\theta) = \frac{1}{(2\pi)^{n/2} \cdot \theta^n} \cdot \exp\left(\frac{\sum_{i=1}^{n} -x_i^2}{2\theta^2}\right) = P(x/\theta)$$

log likelihood :-

$$l(\theta) = -\frac{n}{2}\log 2\pi - n\log\theta - \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2}$$

$\underset{\theta}{\operatorname{argmax}}\ l(\theta)$:- $\frac{\partial l(\theta)}{\partial \theta} = 0$

$$\Rightarrow \frac{-n}{\theta} + \frac{\sum_{i=1}^{n} x_i^2}{\theta^3} = 0$$

$$\Rightarrow \theta^2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

$\therefore$ ML estimate $\Rightarrow \theta = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$

b) $P(x/\theta) = \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right), \quad 0 \le x < \infty, \ \theta > 0.$

$$\prod_{i=1}^{n} p(x_i/\theta) = \frac{1}{\theta^n} \cdot \exp\left(\frac{\sum_{i=1}^{n} -x}{\theta}\right)$$

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

$\arg\max_\theta l(\theta) \Rightarrow \frac{\partial l(\theta)}{\partial \theta} = 0$

$$\Rightarrow \frac{-n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0$$

$$\Rightarrow \theta = \frac{\sum_{i=1}^{n} x_i}{n}$$ is the ML Estimate.

c) $p(x/\theta) = \theta x^{\theta-1}$

$$\prod_{i=1}^{n} p(x_i/\theta) = \theta^n \cdot \left(\prod_{i=1}^{n} x_i\right)^{\theta-1}$$

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

$\arg\max_\theta l(\theta) \Rightarrow \frac{\partial l(\theta)}{\partial \theta} = 0$

$$\Rightarrow \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i = 0$$

$$\Rightarrow \theta = \frac{-n}{\sum_{i=1}^{n} \log x_i}$$

d) $p(x/\theta) = \frac{1}{\theta}, \ 0 \le x \le \theta, \ \theta > 0$

$\prod_{i=1}^{n} p(x_i/\theta) = \frac{1}{\theta^n} \Rightarrow$ This is a monotonically decreasing function with θ. For likelihood to be maximum, θ should be minimum. Hence minimum value of $\theta = \max\{x_i\}$ since $\theta \geq x$

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

a) $$\prod_{i=1}^{n} p(x_i|\mu,\Sigma) = \frac{1}{(2\pi)^{dn/2}|\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right]$$

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

$\underline{\text{argmax}_{\mu,\Sigma}\ l(\mu,\Sigma)}$:-

$$\frac{\partial l(\mu,\Sigma)}{\partial \mu} = 0 \Rightarrow -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1} = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - n\hat{\mu} = 0$$

$$\Rightarrow \underline{\underline{\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}}}.$$

$\cancel{\frac{\partial l(\mu,\Sigma)}{\partial \Sigma}}$ ✱

$$l(\mu,\Sigma) = -\frac{dn}{2}\log 2\pi + \frac{n}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

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

$$\Rightarrow \underline{\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T}$$

using: $\frac{\partial}{\partial A}\log|A| = A^{-T}$; $\Sigma^T = \Sigma$; $tr[ABC] = tr[BAC] = tr[CBA]$ and

$x^T A x = tr[x^T A x] = tr[x x^T A]$.

$\frac{\partial}{\partial A} tr[AB] = B^T$.

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

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} x_i\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} E[x_i]\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu$$

$$= \frac{n\mu}{n} = \underline{\mu}.$$

$E[\hat{u}_n] = \mu$. Hence $\hat{u}_n$ is an un-biased estimate of μ.

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

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

$$= \frac{1}{n}\left[\sum_{i=1}^{n} E[x_i x_i^T - \mu\mu^T]\right] \quad ①.$$

$$= \frac{1}{n}\left(n\Sigma_x - n\Sigma_{\hat{\mu}}\right)$$ (since $\hat{\mu}$ is unbiased estimate of μ)

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

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

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

$$\Sigma_{\hat{\mu}} = \frac{1}{n^2}\sum_{i=1}^{n} E(x_i x_i^T - \mu\mu^T) = \frac{1}{n^2}\cdot n\Sigma_x = \frac{\Sigma_x}{n}.$$

↳ Analogous to $E[x^2-\mu^2] = \sigma^2$

Substituting in ①,

$$= \frac{1}{n}\left(n\Sigma_x - n\frac{\Sigma_x}{n}\right)$$

$$E[\hat{\Sigma}_n] = \frac{n-1}{n}\Sigma$$

$\therefore$ $\hat{\Sigma}_n$ is ~~not~~ a biased estimate of Σ

3.) a) $\lambda = 10$, $P(C_1/x_{test}) = 0.5$, $P(C_2/x_{test}) = 0.25$, $P(C_3/x_{test}) = 0.25$

$R(\text{Reject}/x_{test}) = \lambda = 10$

$R(\alpha_1/x_{test}) = 0.25 \times 10 + 0.25 \times 100 = 27.5$

$R(\alpha_2/x_{test}) = 0.5 \times 1 + 0.25 \times 100 = 25.5$

$R(\alpha_3/x_{test}) = 0.5 \times 1 + 0.25 \times 10 = 3$

Since C_3 has least risk, my predicted class is C_3.

b.) $\lambda = 5$, $P(C_1/x_{test}) = 0.4$, $P(C_2/x_{test}) = 0.5$, $P(C_3/x_{test}) = 0.1$

$R(\text{Reject}/x_{test}) = \lambda = 5$

$R(\alpha_1/x_{test}) = 0.5 \times 10 + 100 \times 0.1 = 15$

$R(\alpha_2/x_{test}) = 0.4 \times 1 + 100 \times 0.1 = 10.4$

$R(\alpha_3/x_{test}) = 0.4 \times 1 + 0.5 \times 10 = 5.4.$

I would predict "Reject" because other classes have higher risk.

Q4: Summary of Error Rates:

MultiGaussClassify with full covariance matrix on Boston50

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
18.63%	18.81%	19.8%	22.77%	24.75%	20.95%	2.41%

Summary: MultiGaussClassify with full covariance matrix on Boston75

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
21.57%	32.67%	27.72%	22.77%	20.79%	25.11%	4.49%

Summary: MultiGaussClassify with full covariance matrix on Digits

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
2.78%	3.06%	2.79%	5.29%	3.62%	3.51%	0.94%

Summary: MultiGaussClassify with diagonal covariance matrix on Boston50

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
20.59%	20.79%	18.81%	26.73%	28.71%	23.13%	3.87%

Summary: MultiGaussClassify with diagonal covariance matrix on Boston75

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
29.41%	36.63%	32.67%	26.73%	24.75%	30.04%	4.24%

Summary: MultiGaussClassify with diagonal covariance matrix on Digits

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
8.89%	10.28%	10.31%	11.7%	9.75%	10.18%	0.91%

Summary: LogisticRegression with Boston50

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
8.82%	13.86%	13.86%	11.88%	20.79%	13.84%	3.93%

Summary: LogisticRegression with Boston75

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
6.86%	14.85%	11.88%	6.93%	9.9%	10.09%	3.04%

Summary: LogisticRegression with Digits

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std Dev
2.5%	2.22%	2.23%	4.46%	3.06%	2.89%	0.84%