Homework 2

Arvind Renganathan
5595189
renga016@umn.edu

October 19, 2019

1

1.a()

P(x|θ) = $\frac{1}{\sqrt{2\pi}\theta} exp(\frac{-x^2}{2(\theta)^2}), \theta > 0$

$L(\theta|x) = logP(x|\theta) = \sum_{t=1}^{n} P(x_t|\theta)$

$L(\theta|x) = \sum_{t=1}^{n} log(\frac{1}{\sqrt{2\pi}\theta} exp(\frac{-(x_t)^2}{2(\theta)^2}))$

$= nlog\frac{1}{\sqrt{2\pi}\theta} - \frac{\sum_{t=1}^{n}(x_t)^2}{2(\theta)^2}$

$\frac{\partial L}{\partial \theta} = -n\sqrt{2\pi}\theta\frac{1}{\sqrt{2\pi}\theta^2} + \frac{\sum_{t=1}^{n}(x_t)^2}{\theta^3} = 0$

$\frac{-n}{\theta} + \frac{\sum_{t=1}^{n}(x_t)^2}{\theta^3} = 0$

$\frac{1}{\theta}(\frac{\sum_{t=1}^{n}(x_t)^2}{\theta^2} - n) = 0$

$\theta = \pm\sqrt{\frac{\sum_{t=1}^{n}(x_t)^2}{n}}$

$\theta > 0\, so\, \hat{\theta} = \sqrt{\frac{\sum_{t=1}^{n}(x_t)^2}{n}}$

$1.b()$

$P(x|\theta) = \frac{1}{\theta} exp(\frac{-x}{2\theta}), 0 \le x < \infty, \theta < \infty$

$P(x|\theta) = \prod_{t=1}^{n} P(x_t|\theta)$

$L(\theta|x) = logP(x|\theta) = \sum_{t=1}^{n} P(x_t|\theta)$

$L(\theta|x) = \sum_{t=1}^{n} log(\frac{1}{\theta} exp(\frac{-x_t}{\theta})$

$= nlog\frac{1}{\theta} - \frac{\sum_{t=1}^{n} x_t}{\theta}$

$\frac{\partial L}{\partial \theta} = -n\theta \frac{1}{\theta^2} + \frac{\sum_{t=1}^{n} x_t}{\theta^2} = 0$

$-n\theta + \sum_{t=1}^{n} x_t = 0$

$\hat{\theta} = \frac{\sum_{t=1}^{n} x_t}{n}$

$1.c()$

$P(x|\theta) = \theta x^{\theta-1}, 0 \leq x \leq 1, 0 < \theta < \infty$

$L(\theta|x) = logP(x|\theta) = \sum_{t=1}^{n} log(\theta(x_t)^{\theta-1}) = nlog\theta + (\theta - 1)\sum_{t=1}^{n} log(x_t)$

$\frac{\partial L}{\partial \theta} = \frac{n}{\theta} + \sum_{t=1}^{n} logx_t = 0$

$\frac{n}{\theta} = -\sum_{t=1}^{n} logx_t$

$\hat{\theta} = \frac{-n}{\sum_{t=1}^{n} logx_t}$

$1.d()$

$P(x|\theta) = \frac{1}{\theta}, 0 \leq x \leq \theta, \theta > 0$

$L(\theta|x) = \sum_{t=1}^{n} log\frac{1}{\theta} = nlog\frac{1}{\theta}$

$\frac{\partial L}{\partial \theta} = n\theta(\frac{-1}{\theta^2} = 0$

$\frac{1}{\theta} = 0$

$\hat{\theta} = \infty$

2

2. $\mathrm{x} = \{\mathrm{x}_1, ...x_n\}, x_i \in \mathbb{R}^d$

$mean\, \mu \in \mathbb{R}^d$

$Covariance\, \Sigma \in \mathbb{R}^{dxd}$

$P(x|\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} exp[\frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)]$

(a)

$For Maximum Likelihood$

$L(\mu,\Sigma|x) = \sum_{i=1}^{n} logP(x_i|\mu,\Sigma)$

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

$Partially\, differentiating$

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

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

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

$Which\, leads\, to$

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

$Partially\, differentiating$

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

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

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

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

(b)

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

$E[\hat{\mu}_n] = E[\frac{1}{n}\sum_{i=1}^{n} x_i]$

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

$\hat{\mu}_n is\, not\, a\, biased\, estimate\, of\, true\, mean\, \mu$

(c)

$let\, true\, Covariance = \Sigma$

$Estimated\, Covariance\, \hat{\Sigma} = \frac{1}{n}(\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})^T)$

$= \frac{1}{n}\sum_{i=1}^{n}(Ex_ix_i^T - Ex_i\bar{x}^T - E\bar{x}x_i^T + E\bar{x}\bar{x}^T) = \frac{1}{n}(n\Sigma + n\mu\mu^T - nE\bar{x}\bar{x}^T)$

$= \Sigma + \mu\mu^T - \mu\mu^T - \frac{\Sigma}{n}$

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

$So\, it\, is\, biased.$

3 SUMMARY OF RESULTS

Error rates For MultiGaussClassify with full covariance matrix on Boston50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.22	0.31	0.16	0.12	0.14	0.24	0.28	0.18	0.12	0.14	0.19	0.06

Error rates For MultiGaussClassify with full covariance matrix on Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.28	0.18	0.22	0.20	0.16	0.31	0.24	0.20	0.18	0.26	0.22	0.05

Error rates For MultiGaussClassify with full covariance matrix on DIGITS											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.02	0.02	0.05	0.03	0.02	0.02	0.03	0.03	0.01	0.01	0.02	0.01

Error rates For MultiGaussClassify with Diagonal covariance matrix on Boston50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.20	0.29	0.14	0.20	0.16	0.10	0.26	0.24	0.26	0.20	0.20	0.06

Error rates For MultiGaussClassify with Diagonal covariance matrix on Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.18	0.20	0.25	0.25	0.25	0.29	0.24	0.18	0.18	0.16	0.22	0.04

Error rates For MultiGaussClassify with Diagonal covariance matrix on Digits											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.00	0.03	0.05	0.03	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.006

Error rates For Logistic Regression with Boston50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.21	0.06	0.23	0.31	0.06	0.17	0.12	0.12	0.16	0.12	0.16	0.08

Error rates For Logistic Regression with Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.06	0.03	0.14	0.10	0.14	0.14	0.14	0.04	0.08	0.08	0.09	0.04

Error rates For Logistic Regression with Digits											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.06	0.14	0.04	0.02	0.03	0.05	0.09	0.10	0.05	0.09	0.07	0.04