CSci 5521: Spring'20
Introduction To Machine Learning

Homework 2

(Due Friday, March 6, 11:59 pm)

Solutions:

1. Question 1

   (a)

$$
\begin{aligned}
p(x|\theta) &= \frac{1}{\sqrt{2\pi}\theta}\exp\left(-\frac{x^2}{2\theta^2}\right), \theta > 0 \\
L(\theta|x) &= \underset{\theta}{\text{maximize}} \prod_i p(x_i|\theta) \\
&\underset{\theta}{\text{maximize}} \sum_i \log p(x_i|\theta)
\end{aligned}
$$

Taking derivatives and equating to 0, we get,

$$
\begin{aligned}
\frac{\partial L(\theta|x)}{\partial \theta} &= \frac{-n}{\theta} + \frac{1}{\theta^3}\sum_i x_i^2 = 0 \\
\theta^2 &= \frac{n}{\sum_i x_i^2} \\
\theta &= \sqrt{\frac{\sum_i x_i^2}{n}}, \quad \theta > 0
\end{aligned}
$$

   (b)

$$
\begin{aligned}
p(x|\theta) &= \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right), 0 \leq x < \infty, \theta > 0 \\
L(\theta|x) &= \underset{\theta}{\text{maximize}} \prod_i p(x_i|\theta) \\
&\underset{\theta}{\text{maximize}} \sum_i \log p(x_i|\theta)
\end{aligned}
$$

Taking derivatives and equating to 0, we get,

$$
\begin{aligned}
\frac{\partial L(\theta|x)}{\partial \theta} &= \frac{-n}{\theta} + \frac{1}{\theta^2}\sum_i x_i = 0 \\
\theta = \frac{\sum_i x_i}{n}, &\quad 0 <= x < \infty, \theta > 0
\end{aligned}
$$

(c)

$$\begin{aligned}
p(x|\theta) &= \theta x^{\theta-1} \ ,0 \leq x \leq 1, 0 < \theta < \infty \\
L(\theta|x) &= \underset{\theta}{\text{maximize}} \ \prod_i p(x_i|\theta) \\
&= \underset{\theta}{\text{maximize}} \ \sum_i \log p(x_i|\theta) \\
&= \underset{\theta}{\text{maximize}} \ n \log \theta + \sum_i (\theta - 1) \log x_i
\end{aligned}$$

Taking derivatives and equating to 0, we get,

$$\begin{aligned}
\frac{\partial L(\theta|x)}{\partial \theta} &= \frac{n}{\theta} + (1) \sum_i x_i = 0 \\
\theta &= \frac{-n}{\sum_i \log x_i}
\end{aligned}$$

(d)

$$\begin{aligned}
p(x|\theta) &= \frac{1}{\theta} \ ,0 \leq x \leq \theta, \theta > 0 \\
L(\theta|x) &= \underset{\theta}{\text{maximize}} \ \prod_i p(x_i|\theta) \\
&= \underset{\theta}{\text{maximize}} \ \frac{1}{\theta^n}
\end{aligned}$$

Since we want to maximize likelihood, and since the likelihood is a decreasing function of theta, we choose the value of theta that will minimum, but, within the given constraints. Since $0 \leq x \leq \theta$,

$$\theta = \underset{i \in [1 \ \dots \ n]}{\text{maximum}} \quad x_i$$

2. Question 2

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

Log likelihood can be given by,

$$l(\mu, \Sigma|x) = -\frac{nd}{2} \log 2\pi - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_i (\mathbf{x}_i - \mu)^T \Sigma^{-1} (\mathbf{x}_i - \mu).$$

(a) i. Taking derivatives with μ and setting to 0, we get,

$$\frac{\partial L}{\partial \mu} = \sum_i (\mathbf{x}_i - \mu)^T \Sigma^{-1} = 0 \tag{1}$$

$$\hat{\mu} = \frac{\sum_i \mathbf{x}_i}{N} \tag{2}$$

ii. Refer to section 13.5 in this link - https://people.eecs.berkeley.edu/ jordan/courses/260-spring10/other-readings/chapter13.pdf

(b)

$$\begin{aligned}
E[\hat{\mu}] &= E\Big[\frac{\sum_i x_i}{n}\Big] \\
&= \frac{1}{n}\sum_i E\Big[x_i\Big] = \frac{1}{n}\sum_i \mu \\
&= \frac{n\mu}{n} = \mu
\end{aligned}$$

Thus, it is an unbiased estimate.

(c)

$$\begin{aligned}
E[\hat{\Sigma_x}] &= E[\frac{1}{N}\sum(x_i - \hat{\mu})(x_i - \hat{\mu})^T] \\
&= E(\frac{1}{N}\sum(x_i x_i^T - 2x_i\hat{\mu}^T + \hat{\mu}\hat{\mu}^T)) \\
&= E(xx^T) - E(\hat{\mu}\hat{\mu}^T) \\
&= (\Sigma_x + \mu\mu^T) - (\Sigma_{\hat{(x)}} + \mu\mu^T) \\
&= \Sigma_x - \Sigma_{\hat{X}} \\
&= \Sigma_x - \Sigma_{\frac{1}{n}\Sigma_X} \\
&= \Sigma_x - \frac{1}{N}\Sigma_x \\
&= \frac{n-1}{n}\Sigma_x
\end{aligned}$$

Thus, it is a biased estimate.

3. Question 3

(a) $P(C_1|X_{test}) = 0.5$
$P(C_1|X_{test}) = 0.25$
$P(C_1|X_{test}) = 0.25$
$R(\alpha_1|X) = 0 * 0.5 + 10 * 0.25 + 100 * 0.25 = 27.5$
$R(\alpha_2|X) = 1 * 0.5 + 0 * 0.25 + 100 * 0.25 = 25.5$
$R(\alpha_3|X) = 1 * 0.5 + 10 * 0.25 + 0 * 0.25 = 3$
$R(\alpha_4|X) = \lambda * 0.5 + \lambda * 0.25 + \lambda * 0.25 = \lambda = 10$
Note α_4 refers to reject.
So we choose C3

(b) $P(C_1|X_{test}) = 0.4$
$P(C_1|X_{test}) = 0.5$
$P(C_1|X_{test}) = 0.1$
$R(\alpha_1|X) = 0 * 0.4 + 10 * 0.5 + 100 * 0.1 = 15$
$R(\alpha_2|X) = 1 * 0.4 + 0 * 0.5 + 100 * 0.1 = 10.4$
$R(\alpha_3|X) = 1 * 0.4 + 10 * 0.5 + 0 * 0.1 = 5.4$
$R(\alpha_4|X) = \lambda * 0.4 + \lambda * 0.5 + \lambda * 0.1 = \lambda = 5$
So we choose Reject

Rubric:

1. Question 1 For all sub-parts, 1 point each for the listing out the correct likelihood, 4 points for the detailed steps in deriving the solution.
2. Question 2
   (a) 5 point each for correct MLE of μ and Σj.
   (b) 5 points for correct steps leading to solution.
   (c) 5 points for correct steps leading to solution.
3. Question 3
   (a) 1 point for each correct $R(\alpha_i|X)$, 1 point for correct choice
   (b) 1 point for each correct $R(\alpha_i|X)$, 1 point for correct choice
4. Question 4

   There are 9 (method/dataset pairs) to be tested in total. You will automatically get 5 points if attempted any of it. For each correct pair you will get another 5 points.
   As for each pair includes MultiGaussClassify, the points are given as follows:
   i. correct implemented estimates of μ: 1 point
   ii. correct implemented estimates of cov: 1 point
   iii. correct implemented estimates of priors: 1 point
   iv. correct predict() function: 1 point
   v. table of summary: 1 point (1 points deducted if you report accuracy rather than error rate)

   There are also some possible cases that causes points deduction:
   i. Forget to import (1 points for each missing import line)