

$$1) a) f(x|\theta) = \frac{x}{\theta^2} \exp\left\{-\frac{x^2}{2\theta^2}\right\}, x \geq 0$$

$$= \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}}$$

$$L(\theta|x) = L(\theta|x_1, \dots, x_n) = \ln \left(\prod_i f(x_i|\theta) \right)$$

$$= \sum_i \ln(f(x_i|\theta))$$

$$= \sum_i \ln \left(\frac{x_i}{\theta^2} \exp\left(-\frac{x_i^2}{2\theta^2}\right) \right)$$

$$= \sum_i \left(\ln x_i - \ln \theta^2 + \ln \exp\left(-\frac{x_i^2}{2\theta^2}\right) \right)$$

$$= \sum_i \left(\ln x_i - 2\ln \theta - \frac{x_i^2}{2\theta^2} \right)$$

$$= \sum_i \ln x_i - \sum_i 2\ln \theta - \sum_i \frac{x_i^2}{2\theta^2}$$

$$\frac{\partial L}{\partial \theta} = 0 - \sum_i \frac{2}{\theta} + \sum_i \frac{x_i^2}{\theta^3} = 0$$

$$= -N \frac{2}{\theta} + \frac{\sum_i x_i^2}{\theta^3} = 0$$

$$= \frac{\sum_i x_i^2}{\theta^3} = \frac{2N}{\theta}$$

$$= \frac{\theta \sum_i x_i^2}{\theta \times 2} = 2N$$

$$= \frac{\sum_i x_i^2}{2N} = \theta^2$$

$$= \boxed{\sqrt[2]{\frac{\sum_i x_i^2}{2N}} = \hat{\theta}}$$

$$1) b) f(x|\alpha, \beta, \theta) = \alpha \theta^{-\alpha} x^\beta \exp\left(-\left(\frac{x}{\theta}\right)^\beta\right), x \geq 0 \quad \alpha, \beta, \theta > 0$$

$$\mathcal{L}(\alpha, \beta, \theta | x) = \mathcal{L}(\alpha, \beta, \theta | x_1, \dots, x_n) = \ln\left(\prod_i f(x_i | \alpha, \beta, \theta)\right)$$

$$= \sum_i \ln(f(x_i | \alpha, \beta, \theta))$$

$$= \sum_i \ln\left(\alpha \theta^{-\alpha} x_i^\beta \exp\left(-\left(\frac{x_i}{\theta}\right)^\beta\right)\right)$$

$$= \sum_i \left(\ln \alpha - \alpha \beta \ln \theta + \beta \ln x_i - \left(\frac{x_i}{\theta}\right)^\beta \right)$$

$$= \sum_i \ln \alpha - \sum_i \alpha \beta \ln \theta + \sum_i \beta \ln x_i - \sum_i \left(\frac{x_i}{\theta}\right)^\beta$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0 - \frac{\alpha \beta}{\theta} + 0 - \frac{\beta x_i^{\beta-1}}{\theta^{\beta}} = 0$$

$$= \frac{\alpha \beta}{\theta} - \frac{\beta x_i^{\beta-1}}{\theta^{\beta}} = 0$$

$$= \cancel{\frac{\alpha \beta}{\theta}} = \cancel{\frac{\beta x_i^{\beta-1}}{\theta^{\beta}}} \Rightarrow \frac{\alpha}{\theta} = x_i^{\beta-1} \theta^{-\beta}$$

$$\Rightarrow \hat{\theta} = \beta^{-1} \sqrt{\frac{x_i^{\beta-1}}{\alpha}}$$

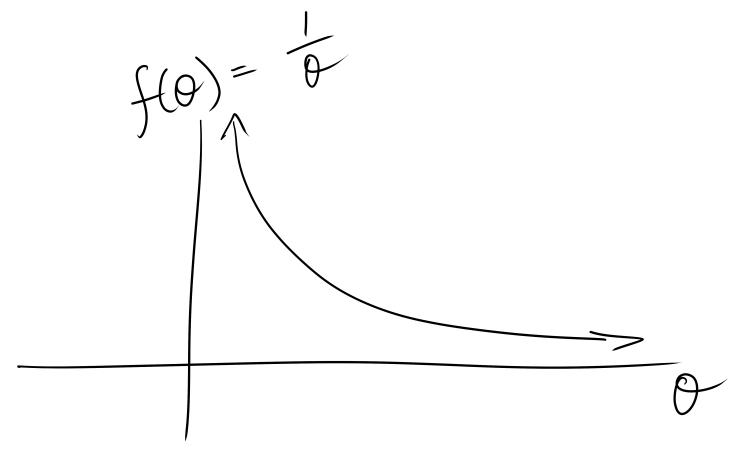
$$1) c) f(x|\theta) = \frac{1}{\theta}, 0 \leq x \leq \theta, \theta > 0$$

$$\begin{aligned} L(\theta|x) &= L(\theta|x_1, \dots, x_n) = \ln \left(\prod_i^{\theta} f(x_i|\theta) \right) \\ &= \sum_i^{\theta} \left(\ln \left(\frac{1}{\theta} \right) \right) \\ &= \sum_i^{\theta} (\ln 1 - \ln \theta) \\ &= \sum_i^{\theta} \ln 1 - \sum_i^{\theta} \ln \theta \end{aligned}$$

$$\frac{\partial L}{\partial \theta} = \theta - \sum_i^{\theta} \left(\frac{1}{\theta} \right)$$

$$\Rightarrow -\sum_i^{\theta} \left(\frac{1}{\theta} \right)$$

↑
 always < 0
 \therefore The higher the θ ,
 the lower



\therefore monotonically decreasing b/c the larger the θ
 for $\theta > 0$, the more minimized it is
 \therefore the function is more maximized the smaller

θ is, for $\theta > 0$

$$\therefore \hat{\theta} = \max(x_i | i \in \mathbb{R}^+)$$

2) a) $P(C_1) = P(C_2) = 0.5$ classify $x=2.5$

$$P(C_1 | x=2.5) = \frac{P(C_1) P(x=2.5 | C_1)}{P(x=2.5 | C_1) P(C_1) + P(x=2.5 | C_2) P(C_2)}$$
$$= \frac{0.5 * \frac{1}{6}}{\frac{1}{6} * 0.5 + \frac{1}{4}(2.5-1) * 0.5}$$
$$= \boxed{\frac{4}{13}}$$

$$P(C_2 | x=2.5) = \frac{P(C_2) P(x=2.5 | C_2)}{P(C_1) P(x=2.5 | C_1) + P(C_2) P(x=2.5 | C_2)}$$
$$= \frac{0.5 * \frac{1}{4}(2.5-1)}{0.5 * \frac{1}{6} + 0.5 * \frac{1}{4}(2.5-1)}$$
$$= \boxed{\frac{9}{13}}$$

$$P(C_2 | x=2.5) = \frac{9}{13} > P(C_1 | x=2.5) = \frac{4}{13} \quad \therefore$$

$x=2.5$ belongs to C_2

$$2) b) P(C_1) = 0.7, P(C_2) = 0.3, x = 4$$

$$\begin{aligned}
 P(C_1 | x=4) &= \frac{P(C_1) P(x=4 | C_1)}{P(C_1) P(x=4 | C_1) + P(C_2) P(x=4 | C_2)} \\
 &= \frac{0.7 * \frac{1}{6}}{\frac{0.7 * \frac{1}{6} + 0.3 * \frac{1}{4}(5-4)}{14}} = \frac{0.7 * \frac{1}{6}}{0.7 * \frac{1}{6} + 0.3 * \frac{1}{4}}
 \end{aligned}$$

$\boxed{\frac{14}{23}}$

$$\begin{aligned}
 P(C_2 | x=4) &= \frac{P(C_2) P(x=4 | C_2)}{P(C_1) P(x=4 | C_1) + P(C_2) P(x=4 | C_2)} \\
 &= \frac{0.3 * \frac{1}{4}(5-4)}{\frac{0.7 * \frac{1}{6} + 0.3 * \frac{1}{4}(5-4)}{14}}
 \end{aligned}$$

$\boxed{\frac{9}{23}}$

\therefore when $x=4$, it's classified for C_1

$$2) c) \phi(x) = (|x-3|) - \alpha, 0 \leq \alpha \leq 2$$

$$x = \begin{cases} C_2 & \text{IFF } \phi(x) < 0 \text{ or } 3-\alpha < x < 3+\alpha \\ C_1 & \text{ow} \end{cases}$$

$$P(C_1) = P(C_2) = .5$$

$$\begin{aligned} P(\text{Miss}) &= P(C_1 | C_2) + P(C_2 | C_1) \\ &= 2 \left(\int_1^{3-\alpha} p(x | C_1) p(C_2) dx + \int_{3-\alpha}^3 p(x | C_2) p(C_1) dx \right) \quad * \text{ symmetry} \\ &= \cancel{2} \left(\cancel{\int_1^{3-\alpha} \frac{1}{6} dx} + \cancel{\int_{3-\alpha}^3 \frac{1}{4}(x-1) dx} \right) \\ &= \int_1^{3-\alpha} \frac{1}{6} dx + \int_{3-\alpha}^3 \frac{1}{4}(x-1) dx = \left[\frac{x}{6} \right]_1^{3-\alpha} + \left[\frac{x^2}{2} - x \right]_{3-\alpha}^3 \\ &= \boxed{-\frac{\alpha^2}{8} + \frac{\alpha}{2} + \frac{2-\alpha}{6}} \Rightarrow \frac{\partial(P(\text{Miss}))}{\partial \alpha} = \frac{1}{3} - \frac{\alpha}{4} = 0 \\ &\Rightarrow -\frac{\alpha}{4} = -\frac{1}{3} \\ &\Rightarrow -\alpha = -\frac{4}{3} \\ &\Rightarrow \alpha = \frac{4}{3} \\ &= -\frac{(4/3)^2}{8} + \frac{(4/3)}{2} + 2 - \frac{(4/3)}{6} \end{aligned}$$

$$P(\text{Miss}) = \boxed{\frac{5}{9}}$$

3.

The discriminant in each case was:

- S_1 and S_2 learned independently:

- $-(1/2) * \log(\text{np.linalg.det}(\text{self.S}[c])) - (1/2) * \text{np.dot}(\text{xtest}[i] - \text{self.mean}[c], \text{np.linalg.inv}(\text{self.S}[c])).\text{dot}(\text{xtest}[i] - \text{self.mean}[c]) + \log(\text{self.p}[c])$

- Taken from *Bayes_Parametric_Multivariate_II.pdf* page 12

- $g_i(x) = -\frac{1}{2} \log |S_i| - \frac{1}{2}(x - m_i)^T S_i^{-1}(x - m_i) + \log P_{est.}(C_i)$

- Classified based on a weighted distance from the mean using the covariance matrix S .

- $S_1 = S_2$

- $-(1/2) * \log(\text{np.linalg.det}(\text{self.S})) - (1/2) * \text{np.dot}(\text{xtest}[i] - \text{self.mean}[c], \text{np.linalg.inv}(\text{self.S})).\text{dot}(\text{xtest}[i] - \text{self.mean}[c]) + \log(\text{self.p}[c])$

- Taken from *Bayes_Parametric_Multivariate_II.pdf* page 12

- $g_i(x) = -\frac{1}{2} \log |S| - \frac{1}{2}(x - m_i)^T S^{-1}(x - m_i) + \log P_{est.}(C_i)$

- Classified based similarly to the previous assumption, but there's no indexing the covariance matrix S because it's shared between classes, which leads to slightly different confusion matrices.

- $S_1 = S_2$, Diagonal Matrix

- $\text{tempTerm} = \text{np.sum}(\text{np.power}((\text{xtest}[i] - \text{self.mean}[c]) / \text{np.sqrt}(\text{self.S}), 2))$
 $\text{tempList.append}(-(1/2) * (\text{tempTerm}) + \log(\text{self.p}[c]))$

- Taken from *Bayes_Parametric_Multivariate_II.pdf* page 17

- $g_i(x) = -\frac{1}{2} \sum_{j=1}^d \left(\frac{x_j - m_{ij}}{s_j} \right)^2 + \log P_{est.}(C_i)$

- Classified based on weighted Euclidean distance to the nearest mean as the diagonals are variances. This discriminant takes the sum of the normalized distances from the means by some standard deviation squared -- which is different from the first two assumptions where they take a weighted distance from the mean instead.

Confusion matrices:

Confusion Matrix for Gaussian Discriminant with class-dependent covariance

```
[[14  5]
 [16 65]]
```

Confusion Matrix for Gaussian Discriminant with class-independent covariance

```
[[24  5]
 [ 6 65]]
```

Confusion Matrix for Gaussian Discriminant with diagonal covariance

```
[[26 11]
 [ 4 59]]
```

Looking at the results of the Gaussian Discriminants, one thing to note is that the 2nd dimension of the first two matrices are the same, with values of 5 and 65. However, their first dimension differs. The reason for such a difference in the first dimension while being the same in the 2nd is because of the first term is due to having different sets of parameters, S and m at index i . What's more is that the covariances stored in S from the `fit()` function contributed to such a difference.

What makes the third confusion matrix different from both of the former confusion matrices is that the discriminant function is vastly different compared to the first two, which share the same. In this discriminant function, the first term instead takes the distance a data point is from the mean and normalizes it by some standard deviation, squared, and then summed. This of course leads to a matrix that is different from the others.