Machine Learning Homework 2

Hairui Yin

1.
(a) When $x > 0$, the marginal pdf is

$$
\begin{aligned}
f_X(x) &= \int_0^{+\infty} 6e^{-(2x+3y)} dy \\
&= 6e^{-2x}(-\frac{1}{3})e^{-3y}\Big|_0^{+\infty} \\
&= 6e^{-2x}(\frac{1}{3}) \\
&= 2e^{-2x}
\end{aligned}
$$

When $x \leq 0$, the marginal pdf is $f_X(x) = 0$.
Therefore, the marginal pdf f_X of random variable X is

$$
f_X(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}
$$

(b) With the marginal pdf f_X, its expectations i

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx \\
&= \int_0^{+\infty} 2xe^{-2x} dx \\
&= (-\frac{1}{2})2xe^{-2x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-2x} dx \\
&= -xe^{-2x}\Big|_0^{+\infty} - \frac{1}{2}e^{-2x}\Big|_0^{+\infty} \\
&= \frac{1}{2}
\end{aligned}
$$

Therefore, the expected value $\mathbb{E}[X] = \frac{1}{2}$.
(c) To determine whether X and Y are independent, we need to find if $f_{X,Y} = f_X \times f_Y, \forall x, y$.

The marginal pdf of Y when $y > 0$ is

$$\begin{aligned} f_Y(y) &= \int_0^{+\infty} 6e^{-(2x+3y)} dx \\ &= 3e^{-3y} \end{aligned}$$

(1) $x > 0, y > 0$

$$\begin{aligned} f_X \times f_Y &= 2e^{-2x} \times 3e^{-3y} \\ &= 6e^{-(2x+3y)} = f_{X,Y} \end{aligned}$$

(2) $x > 0, y \leq 0$ or $x \leq 0, y > 0$ or $x, y \leq 0$

$$\begin{aligned} f_X \times f_Y &= 0 \\ &= f_{X,Y} \end{aligned}$$

Since $f_{X,Y} = f_X \times f_Y$, X and Y are independent.

2. Denote the event that the tumor is benign B and the tumor is malignant M, then $P(B) = 0.9, P(M) = 0.1$.
Denote the cost as $C_{S_1 \to S_2}$, where $S_1, S_2 \in \{B, M\}$ (benign or malignant), and $S_1 \to S_2$ means the patient tumor is state S_1 but diagnose as S_2. Therefore, $C_{B\to B} = C_{M\to M} = 0, C_{B\to M} = 3, C_{M\to B} = 20$.
Denote the sample growth rate of tumor in the patient as $\mathbf{x}$.
And there are only two result that the doctor can diagnose, denote them as D_B, D_M (Diagnose as Benign, Diagnose as Malignant).
For the given $\mathbf{x}$, we need to find the posterior probabilities $P(B|\mathbf{x})$ and $P(M|\mathbf{x})$,

$$\begin{aligned} P(B|\mathbf{x}) &= \frac{P(\mathbf{x}|B)P(B)}{P(\mathbf{x})} \\ &= \frac{p_B(\mathbf{x})P(B)}{P(\mathbf{x})} \\ &= \frac{5e^{-5\mathbf{x}} \times 0.9}{P(\mathbf{x})} \\ &= \frac{4.5e^{-5\mathbf{x}}}{P(\mathbf{x})} \end{aligned}$$

$$\begin{aligned} P(M|\mathbf{x}) &= \frac{P(\mathbf{x}|M)P(M)}{P(\mathbf{x})} \\ &= \frac{p_M(\mathbf{x})P(M)}{P(\mathbf{x})} \\ &= \frac{2e^{-2\mathbf{x}} \times 0.1}{P(\mathbf{x})} \\ &= \frac{0.2e^{-2\mathbf{x}}}{P(\mathbf{x})} \end{aligned}$$

Cosidering the expected cost $R(D_B|\mathbf{x}), R(D_M|\mathbf{x})$,

$$\begin{aligned} R(D_B|\mathbf{x}) &= C_{B\to B}P(B|\mathbf{x}) + C_{M\to B}P(M|\mathbf{x}) \\ &= 0 + 20 \times \frac{0.2e^{-2\mathbf{x}}}{P(\mathbf{x})} \\ &= \frac{4e^{-2\mathbf{x}}}{P(\mathbf{x})} \end{aligned}$$

$$\begin{aligned} R(D_M|\mathbf{x}) &= C_{M\to M}P(M|\mathbf{x}) + C_{B\to M}P(B|\mathbf{x}) \\ &= 0 + 3 \times \frac{4.5e^{-5\mathbf{x}}}{P(\mathbf{x})} \\ &= \frac{13.5e^{-5\mathbf{x}}}{P(\mathbf{x})} \end{aligned}$$

Therefore, we can design the Bayes Classifier that can minimize overall risk as diagnosing

$$\begin{cases} \text{Benign :} & \text{When } R(D_B|\mathbf{x}) < R(D_M|\mathbf{x}) \\ \text{Malignant :} & \text{When } R(D_B|\mathbf{x}) \geq R(D_M|\mathbf{x}) \end{cases}$$

More specificly, the Bayes Classifier is designed as to diagnose

$$\begin{cases} \text{Benign :} & \text{When } \mathbf{x} < \ln\left(\frac{3}{2}\right) \\ \text{Malignant :} & \text{When } \mathbf{x} \geq \ln\left(\frac{3}{2}\right) \end{cases}$$

3.

(a) With information in the question, we know that $P(GL) = 0.1, P(NL) = 0.9$. Consider the minimum error-rate classifier, which is given by

$$g_i(\mathbf{x}) = \log\left(p(\mathbf{x}|w_i)\right) + \log\left(P(w_i)\right), \quad i \in \{GL, NL\}$$

we need to find a larger $g_i(\mathbf{x})$ given $\mathbf{x}$ to minimize the probability of error. Since $\Sigma = \sigma^2 I_{2\times 2}$, the linear discriminant function is

$$\begin{aligned} g_{GL}(x) &= \frac{1}{\sigma^2}\mu_{GL}^T\mathbf{X} - \frac{1}{2\sigma^2}||\mu_{GL}||^2 + \log\left(P(GL)\right) \\ &= \frac{1}{100}\begin{bmatrix}10 & 10\end{bmatrix}\begin{bmatrix}X_1 \\ X_2\end{bmatrix} - \frac{1}{2\times 100} \times 200 + \log\left(\frac{1}{10}\right) \\ &= 0.1X_1 + 0.1X_2 - 1 + \log\left(\frac{1}{10}\right) \end{aligned}$$

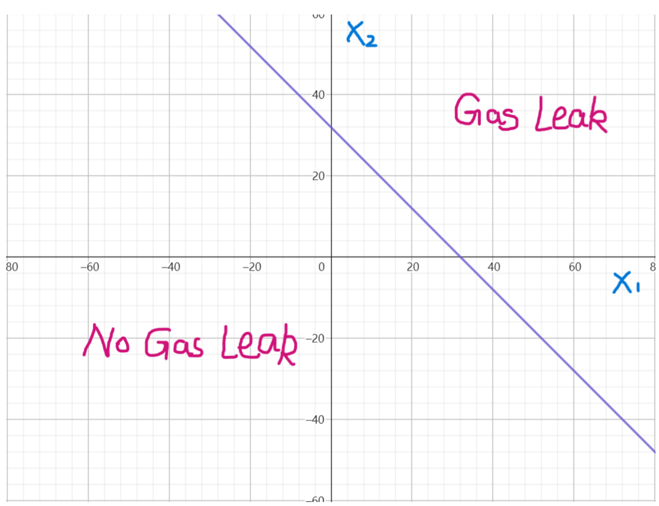
$$\begin{aligned} g_{NL}(x) &= \frac{1}{\sigma^2}\mu_{NL}^T\mathbf{X} - \frac{1}{2\sigma^2}||\mu_{NL}||^2 + \log\left(P(NL)\right) \\ &= \frac{1}{100}\begin{bmatrix}0 & 0\end{bmatrix}\begin{bmatrix}X_1 \\ X_2\end{bmatrix} - \frac{1}{2\times 100} \times 0 + \log\left(\frac{9}{10}\right) \\ &= \log\left(\frac{9}{10}\right) \end{aligned}$$

When there is a gas leak, $g_{GL}(x) > g_{NL}(x)$,

$$0.1X_1 + 0.1X_2 - 1 + \log\left(\frac{1}{10}\right) > \log\left(\frac{9}{10}\right)$$
$$X_1 + X_2 > 10\log 9 + 10$$

Therefore, we decide

$$\begin{cases} \text{Gas Leak} & X_1 + X_2 > \log 9 + 10 \\ \text{No Gas Leak} & X_1 + X_2 \le \log 9 + 10 \end{cases}$$



(b) Similarly to (a), but since $\Sigma = \begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix}$, the discriminant funciton is different, given by

$$g_i(\mathbf{x}) = (\Sigma^{-1}\mu_i)^T\mathbf{x} - \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \log\left(P(w_i)\right)$$

Therefore,

$$\begin{aligned}
g_{GL}(x) &= (\begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 10 \end{bmatrix})^T \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 10 \\ 10 \end{bmatrix}^T \begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 10 \end{bmatrix} + \log\frac{1}{10} \\
&= \begin{bmatrix} \frac{1}{75} & -\frac{1}{150} \\ -\frac{1}{150} & \frac{1}{75} \end{bmatrix} \begin{bmatrix} 10 \\ 10 \end{bmatrix})^T \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 10 \\ 10 \end{bmatrix}^T \begin{bmatrix} \frac{1}{75} & -\frac{1}{150} \\ -\frac{1}{150} & \frac{1}{75} \end{bmatrix} \begin{bmatrix} 10 \\ 10 \end{bmatrix} + \log\frac{1}{10} \\
&= \frac{1}{15}X_1 + \frac{1}{15}X_2 - \frac{2}{3} + \log\frac{1}{10}
\end{aligned}$$

$$\begin{aligned}
g_{NL}(x) &= (\begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix})^T \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \log\frac{9}{10} \\
&= \log\frac{9}{10}
\end{aligned}$$

When there is a gas leak, $g_{GL}(x) > g_{NL}(x)$,

$$\frac{1}{15}X_1 + \frac{1}{15}X_2 - \frac{2}{3} + \log\frac{1}{10} > \log\frac{9}{10}$$

$$X_1 + X_2 > 15\log 9 + 10$$

Therefore, the decision region is defined as

$$\begin{cases} \text{Gas Leak} & X_1 + X_2 > 15\log 9 + 10 \\ \text{No Gas Leak} & X_1 + X_2 \leq 15\log 9 + 10 \end{cases}$$