1. The 1-nearest neighbour for each training data point is itself, so the training error is always zero and test error is 36%, higher than the test error by logistic regression. We prefer to use the logistic regression.

2. Let $Y = 0$ and $Y = 1$ denote Classes 1 and 2 respectively. Let $g_1(x) = \frac{1}{4}I(0 \le x \le 4)$ and $g_2(x) = \frac{1}{3}I(-2 \le x \le 1)$ be the density function for $X|Y = 0$ and $X|Y = 1$, respectively. For $x \in [0, 1]$, since $\frac{g_1(x)P(Y=0)}{g_1(x)P(Y=0)+g_2(x)P(Y=1)} = \frac{1/4\times1/3}{1/4\times1/3+1/3\times2/3} = \frac{3}{11}$, we always classify x to Class 2. The corresponding Bayesian error rate is $\varepsilon(x) = 3/11$. For $x \notin [0, 1]$ since $g_1(x)$ and $g_2(x)$ do not overlap, then $\varepsilon(x) = 0$.

$$\begin{aligned}
E[\varepsilon(X)] &= E\left[\frac{3}{11}I(X \in [0,1])\right] = \frac{3}{11}P(X \in [0,1]) \\
&= \frac{3}{11}\{P(X \in [0,1]|Y = 0)P(Y = 0) + P(X \in [0,1]|Y = 1)P(Y = 1)\} \\
&= \frac{3}{11}\left(\frac{1}{4}\times\frac{1}{3}+\frac{1}{3}\times\frac{2}{3}\right) = \frac{1}{12}.
\end{aligned}$$

3. Note $g_j(\mathbf{x}) = (2\pi)^{-1/2}|\boldsymbol{\Sigma}_j|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\right\}, j = 1, 2$ where $\mathbf{x} = (x_1, x_2)^T$. Then the Bayes decision rule is given by

$$\begin{cases} \text{Class 1} & \text{if } \pi_1 g_1(\mathbf{x}) > \pi_2 g_2(\mathbf{x}), \\ \text{Class 2} & \text{otherwise} \end{cases}$$

Solve for $\pi_1 g_1(\mathbf{x}) > \pi_2 g_2(\mathbf{x})$, i.e. $\log\pi_1 + \log(g_1(\mathbf{x})) > \log\pi_2 + \log(g_2(\mathbf{x}))$ and some calculations show that the Bayes decision rule is given by

$$\begin{cases} \text{Class 1} & \text{if } \frac{1}{2}x_1 - \frac{3}{4} + \log 2 > 0 \\ \text{Class 2} & \text{otherwise} \end{cases}$$

Or one can use discriminant score for LDA, since the covariance matrices for two classes are the same.

4. Let $g_1(x), g_2(x), g_3(x)$ be the density function for $X|Y = 1$, $X|Y = 2$, $X|Y = 3$, respectively. $P(Y = i|X = x) = \frac{g_i(x)P(Y=i)}{f(x)}, i = 1, 2, 3$. We only need to compare $\pi_i g_i(x_0)$'s

$$\pi_1 g_1(x_0) = \frac{2}{5}\times\frac{1}{2\pi}\exp\left[-\frac{1}{2}(0.3-0, 0.3-0)(0.3-0, 0.3-0)^T\right] = 0.0582;$$

$$\pi_2 g_2(x_0) = \frac{2}{5} \times \frac{1}{2\pi} \exp\left[-\frac{1}{2}(0.3-1, 0.3-1)(0.3-1, 0.3-1)^T\right] = 0.0390;$$

$$\begin{aligned}\pi_3 g_3(x_0) &= \frac{1}{5} \times \frac{1}{4\pi} \exp\left[-\frac{1}{2}(0.3-0.5, 0.3-0.5)(0.3-0.5, 0.3-0.5)^T\right] \\ & \frac{1}{5} \times \frac{1}{4\pi} \exp\left[-\frac{1}{2}(0.3+0.5, 0.3-0.5)(0.3+0.5, 0.3-0.5)^T\right] = 0.0266\end{aligned}$$

Since $\pi_1 g_1(x_0)$ is the largest, we classify $x_0 = (0.3, 0.3)^T$ to be from Class 1.

5. See the solution to problem 5 of HW 2 generated by R Markdown (provided by Cheng Chen, TA for ST443).