1 Problem 1

a)
We need to show

$$P(error) \leq \int \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} \times p(x)dx$$

and according to the equation from question, we know,

$$\int P(error|x) \times p(x)dx \leq \int \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} \times p(x)dx$$

By Fundamental theorem of calculus, we just need to show the equality,

$$P(error|x) \leq \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}}$$

we know,

$$P(error|x) = min[P(w_1|x), P(w_2|x)]$$

to show the equality:

$$min[P(w_1|x), P(w_2|x)] \leq \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}}$$

$$min[P(w_1|x), P(w_2|x)] \times \left(\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}\right) \leq 2$$

$$min\left[1 + \frac{P(w_1|x)}{P(w_2|x)}, 1 + \frac{P(w_2|x)}{P(w_1|x)}\right] \leq 2$$

$$min\left[\frac{P(w_1|x)}{P(w_2|x)}, \frac{P(w_2|x)}{P(w_1|x)}\right] \leq 1$$

we know the decision,

$$= \begin{cases} w_1, & \text{if } P(w_1|x) \geq P(w_2|x) \\ w_2, & \text{if } P(w_2|x) \geq P(w_1|x) \end{cases}$$

if decided w_1,

$$min\left[\frac{P(w_1|x)}{P(w_2|x)}, \frac{P(w_2|x)}{P(w_1|x)}\right] = \frac{P(w_2|x)}{P(w_1|x)} < 1$$

if decided w_2,

$$min\left[\frac{P(w_1|x)}{P(w_2|x)}, \frac{P(w_2|x)}{P(w_1|x)}\right] = \frac{P(w_1|x)}{P(w_2|x)} < 1$$

when $P(w_1|x) = P(w_2|x)$,

$$min\left[\frac{P(w_1|x)}{P(w_2|x)}, \frac{P(w_2|x)}{P(w_1|x)}\right] = 1$$

Proved.

b)
we know, $P(w_1|x) = \frac{P(x|w_1)P(w_1)}{P(x)}$, $P(w_2|x) = \frac{P(x|w_2)P(w_2)}{P(x)}$, and the equality from a) we rewrite it as below:

$$P(error) \leq \int \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} \times p(x)dx$$

$$P(error) \leq \int \frac{2P(w_1|x)P(w_2|x)}{P(w_1|x) + P(w_2|x)} p(x)dx$$

then we plug th,,e equations of $P(w_1|x)$ and $P(w_2|x)$ into the inequality :

$$P(error) \leq \int \frac{2P(x|w_1)P(w_1)P(x|w_2)P(w_2)}{P(x|w_1)P(w_1) + P(x|w_2)P(w_2)} dx$$

plug the values of $P(x|w_1)$ and $P(x|w_2)$ into the inequality:

$$\begin{aligned} P(error) &\leq \int \frac{2\pi^{-1}P(w_1)P(w_2)}{P(w_1)(1+(x+\mu)^2) + P(w_2)(1+(x+\mu)^2)} dx \\ &\leq \int \frac{2\pi^{-1}P(w_1)P(w_2)}{[P(w_1)+P(w_2)+(P(w_1)+P(w_2))\mu)^2] + [P(w_1)+P(w_2)]x^2 + [P(w_1)-P(w_2)] \times 2\mu x} dx \\ &(ps : P(w_1) + P(w_2) = 1) \\ &\leq \int \frac{2\pi^{-1}P(w_1)P(w_2)}{(1+\mu^2) + x^2 + (P(w_1)-P(w_2)) \times 2\mu x} dx \end{aligned}$$

Lemma from question,

$$\int \frac{1}{ax^2+bx+c} dx = \frac{2\pi}{\sqrt{4ac-b^2}}$$

, when $b^2 \leq 4ac$, in our case, $4\mu^2(P(w_1) - P(w_2)) - 4(1+\mu^2) < 0$, because $|P(w_1) - P(w_2)| \leq 1$
So,

$$\begin{aligned} P(error) &\leq 2\pi^{-1}P(w_1)P(w_2) \times \frac{2\pi}{\sqrt{4(1+\mu^2) - 4\mu^2(P(w_1)-P(w_2))}} \\ &\leq \frac{4P(w_1)P(w_2)}{2\sqrt{1+\mu^2-\mu^2[(P(w_1)+P(w_2))^2 - 4P(w_1)P(w_2)]}} \\ &(ps : P(w_1) + P(w_2) = 1) \\ &\leq \frac{2P(w_1)P(w_2)}{\sqrt{1+4\mu^2P(w_1)P(w_2)}} \end{aligned}$$

Proved

c)

We know that $P(error) = \int P(error, x)dx$, marginalizing out a random variable from joint distribution is same as sum of all the probablities $P(error, x)$, so if there is no upper bound of error, we can just simply aggregate all the $P(error, x)$. But when x is a continuous variable, we need to consider how long of each bin width to sum them up, it's better to divide the whole interval as small as possible. For the low-dimensional data, it's computationally feasible to realize the calculation, however, for the high-dimensional, we have to not only consider the computation, but also need to "think of " the curse of high-dimensional, by the increasing the dimensions, the accuracy would decrease.

2 Problem 2

a)
As we know, when $P(w_1|x) = P(w_2|x)$, the error would reach the Maximum value, so:

$$P(w_1|x) = \frac{P(x|w_1)P(w_1)}{P(x)} = P(w_2|x) = \frac{P(x|w_2)P(w_2)}{P(x)}$$

$$\frac{P(w_1)}{P(w_2)} = \frac{P(x|w_2)}{P(x|w_1)} = \frac{\exp\left(\frac{-|x+\mu|}{\sigma}\right)}{\exp\left(\frac{-|x-\mu|}{\sigma}\right)}$$

then we made logarithums on both sides, we get:

$$\ln\left(\frac{P(w_1)}{P(w_2)}\right) = \frac{-|x+\mu|}{\sigma} + \frac{|x-\mu|}{\sigma}$$

Therefore, the Bayes decision Boundary is the set of points $x \in \mathbb{R}$, that satisfy above the equation, we could set a new set Z, $Z \subseteq \mathbb{R}$,

$$Z = \{x \in \mathbb{R} | |x-\mu| - |x+\mu| = \sigma[\ln P(w_1) - \ln P(w_2)]\}$$

b)
The optimal decision is always the first class w_1, it should satisfy: $P(w_1|x) > P(w_2|x)$, so we could get the relationship on below, and the process is same as question a).

$$\frac{P(x|w_1)P(w_1)}{P(x)} > \frac{P(x|w_2)P(w_2)}{P(x)}$$

$$\frac{P(w_1)}{P(w_2)} > \frac{P(x|w_2)}{P(x|w_1)} = \frac{\exp\left(\frac{-|x+\mu|}{\sigma}\right)}{\exp\left(\frac{-|x-\mu|}{\sigma}\right)}$$

$$\ln\left(\frac{P(w_1)}{P(w_2)}\right) > \frac{-|x+\mu|}{\sigma} + \frac{|x-\mu|}{\sigma}$$

we know that $||x| - |y|| \leq |x-y|$, therefore, $||x+\mu| - |x+\mu|| \leq |2\mu|$

$$\sigma \ln\left(\frac{P(w_1)}{P(w_2)}\right) > 2|\mu|$$

c) The similar process as a),b)

$$P(w_1|x) = \frac{P(x|w_1)P(w_1)}{P(x)} = P(w_2|x) = \frac{P(x|w_2)P(w_2)}{P(x)}$$

$$\frac{P(w_1)}{P(w_2)} = \frac{P(x|w_2)}{P(x|w_1)} = \frac{\exp\left(\frac{-(x+\mu)^2}{2\sigma^2}\right)}{\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}$$

$$\ln\left(\frac{P(w_1)}{P(w_2)}\right) = \frac{-(x+\mu)^2}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^2}$$
$$2\sigma^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) = -(x+\mu)^2 + (x-\mu)^2$$
$$2\sigma^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) = -4\mu x$$

If we want to always predict the first class w_1, so it should satisfy the inequality:

$$2\sigma^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) > -4\mu x$$

$$(ps : P(w_1|x) > P(w_2|x))$$

Then, we could deduce that

$$P(w_1) > P(w_2)$$
$$\frac{P(w_1)}{P(w_2)} > 1$$
$$ln\left(\frac{P(w_1)}{P(w_2)}\right) > 0$$
$$2\sigma^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) > 0$$

Only when $\mu = 0$, the inequality will hold.