EECE 5644 Homework #1

Yuxin Lin

September 17, 2022

1. Let x be a real-valued random variable.

   (a) Prove that the variance of $x = \sigma^2 = E\left[(x-\mu)^2\right] = E\left[x^2\right] - \mu^2$.
   According to the definition, $\mu = E[x]$.
   Here we have $\sigma = E\left[(x-\mu)^2\right] = E[x^2 - 2\mu x + \mu^2] = E[x^2] - 2\mu E[x] + \mu^2 = E[x^2] - 2\mu^2 + \mu^2 = E[x^2] - \mu^2$.
   Therefore $\sigma^2 = E\left[(x-\mu)^2\right] = E\left[x^2\right] - \mu^2$.

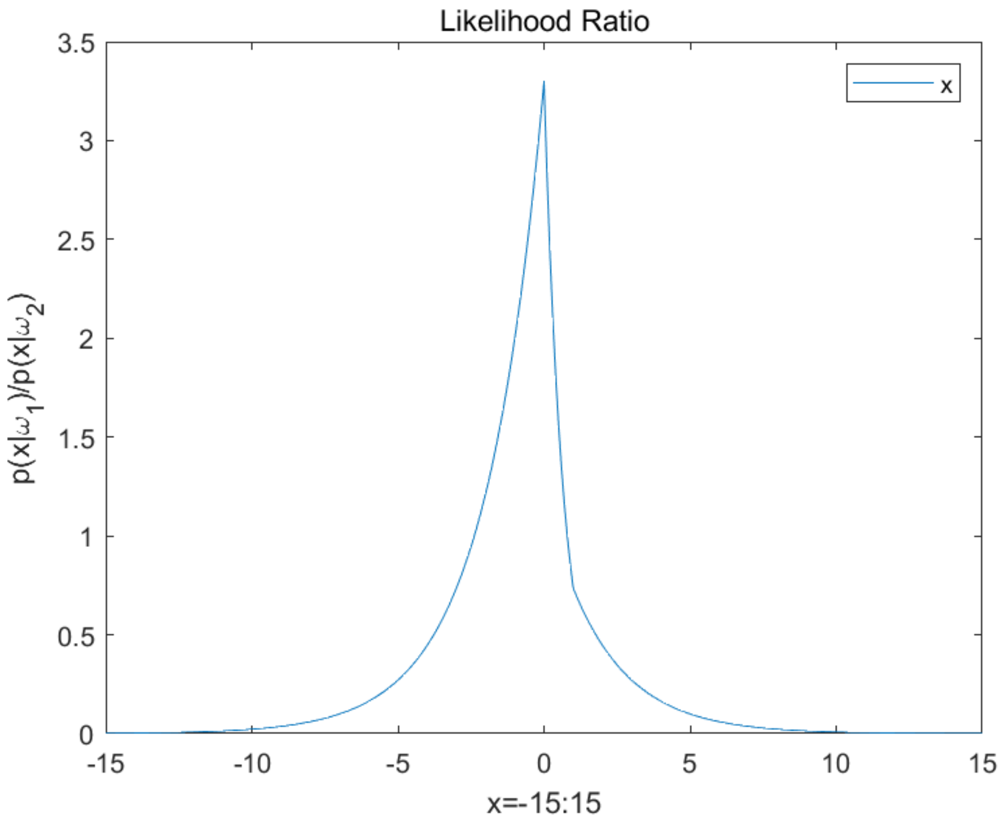
   (b) Let $\mathbf{x}$ be a real-valued random vector. Prove that the covariance matrix of $\mathbf{x} = \Sigma = E\left[\mathbf{x}\mathbf{x}^T\right] - \mu\mu^T$.
   According to the definition, $\Sigma = E[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T]$, where $\mu = E[\mathbf{x}]$.
   Here we have $\Sigma = E[(\mathbf{x}-\mu)(\mathbf{x}^T-\mu^T)] = E[\mathbf{x}\mathbf{x}^T - \mu\mathbf{x}^T - \mathbf{x}\mu^T + \mu\mu^T]$.
   Since μ is a constant vector, we can derive that $\Sigma = E[\mathbf{x}\mathbf{x}^T] - \mu E[\mathbf{x}^T] - E[\mathbf{x}]\mu^T + \mu\mu^T = E[\mathbf{x}\mathbf{x}^T] - \mu E[\mathbf{x}]^T - E[\mathbf{x}]\mu^T + \mu\mu^T = E[\mathbf{x}\mathbf{x}^T] - \mu\mu^T$.
   Therefore $\Sigma = E\left[\mathbf{x}\mathbf{x}^T\right] - \mu\mu^T$.

2. Suppose two equally probable one-dimensional densities are of the form $p(x \mid \omega_i) \propto e^{-|x-a_i|/b_i}$ for $i = 1, 2$ and $b > 0$.

   (a) Write an analytic expression for each density, that is, normalize each function for arbitrary a_i, and positive b_i.
   Since $p(x \mid \omega_i) \propto e^{-|x-a_i|/b_i}$, we assume that $p(x \mid \omega_i) = k_i e^{-|x-a_i|/b_i}$ for $i = 1, 2$.
   Here we have $\int_{-\infty}^{+\infty}(k_i e^{-|x-a_i|/b_i})dx = 1$.
   So that $\int_{-\infty}^{+\infty}(k_i e^{-|x-a_i|/b_i})dx = k_i \int_{a_i}^{+\infty}(e^{(-x+a_i)/b_i})dx + k_i \int_{-\infty}^{a_i}(e^{(x-a_i)/b_i})dx$
   $= -k_i b_i e^{(-x+a_i)/b_i}\Big|_{a_i}^{+\infty} + k_i b_i e^{(x-a_i)/b_i}\Big|_{-\infty}^{a_i} = 2k_i b_i = 1$, and then we have $k_i = \frac{1}{2b_i}$.
   To sum up, the analytic expression for the two densities is $p(x \mid \omega_i) = \frac{1}{2b_i} e^{\frac{-|x-a_i|}{b_i}}$ for $i = 1, 2$, with arbitrary a_i and positive b_i.

   (b) Calculate the likelihood ratio $p(x \mid \omega_1)/p(x \mid \omega_2)$ as a function of your four variables.
   According to (a), we have:
   $p(x \mid \omega_1)/p(x \mid \omega_2) = \frac{(1/2b_1)e^{-|x-a_1|/b_1}}{(1/2b_2)e^{-|x-a_2|/b_2}} = \frac{b_2}{b_1} e^{-\frac{|x-a_1|}{b_1} + \frac{|x-a_2|}{b_2}}$.

   (c) Plot a graph (using MATLAB) of the likelihood ratio for the case $a_1 = 0, b_1 = 1, a_2 = 1$ and $b_2 = 2$. Make sure the plots are correctly labeled (axis, titles, legend, etc) and that the fonts are legible when printed.
   For this case, $p(x \mid \omega_1) = \frac{1}{2} e^{-|x|}$ and $p(x \mid \omega_2) = \frac{1}{4} e^{-\frac{|x-1|}{2}}$.
   Therefore we have $p(x \mid \omega_1)/p(x \mid \omega_2) = 2e^{-|x| + \frac{|x-1|}{2}}$. The corresponding graph plotted with MATLAB is shown below:



3. Consider a two-class problem, with classes $c1$ and $c2$ where $p(c1) = p(c2) = 0.5$. There is a one-dimensional feature variable x. Assume that the x data for class one is uniformly distributed between a and b, and the x data for class two is uniformly distributed between r and t. Assume that $a < r < b < t$. Derive a general expression for the Bayes error rate for this problem. (Hint: a sketch may help you think about the solution.)

   Let $R1$ denote $p(c1|x) > p(c2|x)$ (decide $c = c1$), and $R2$ denote $p(c1|x) \leq p(c2|x)$ (decide $c = c2$).
   Let the Bayes error rate for this problem be $p(error)$. Here we have $p(error) = p(R1, c2) + p(R2, c1)$.
   Let x_1 denote $x \in [a, r)$, x_2 denote $x \in [r, b)$, and x_3 denote $x \in [b, t]$.
   According to the distribution assumptions of x, we have:
   $p(c1|x_1) = 1$, $p(c2|x_1) = 0$, $p(c1|x_3) = 0$, $p(c2|x_3) = 1$,
   $p(c1|x_2) = \frac{p(x_2|c1)p(c1)}{p(x_2)} = \frac{0.5(b-r)/(b-a)}{p(x_2)}$, $p(c2|x_2) = \frac{p(x_2|c2)p(c2)}{p(x_2)} = \frac{0.5(b-r)/(t-r)}{p(x_2)}$.
   When x_1, $R1$ is always true. When x_3, $R2$ is always true.
   Consider the event x_2. We can see when $t - r > b - a$, $R1$ is true, otherwise $R2$ is true.
   Therefore we have:
   $p(R1, c2) = p(R1, c2|x_1)p(x_1) + p(R1, c2|x_2)p(x_2) + p(R1, c2|x_3)p(x_3)$
   $= p(R1, c2|x_2)p(x_2) = p(p(c1|x_2) > p(c2|x_2), c2|x_2)p(x_2) = \begin{cases} 0.5(b-r)/(t-r), & t-r > b-a \\ 0, & otherwise \end{cases}$,
   $p(R2, c1) = p(R2, c1|x_1)p(x_1) + p(R2, c1|x_2)p(x_2) + p(R2, c1|x_3)p(x_3)$
   $= p(R2, c1|x_2)p(x_2) = p(p(c1|x_2) \leq p(c2|x_2), c1|x_2)p(x_2) = \begin{cases} 0, & t-r > b-a \\ 0.5(b-r)/(b-a), & otherwise \end{cases}$.
   In conclusion, $p(error) = p(R1, c2) + p(R2, c1) = \begin{cases} 0.5(b-r)/(t-r), & t-r > b-a \\ 0.5(b-r)/(b-a), & otherwise \end{cases}$.

4. Consider a two-class, one-dimensional problem where $P(\omega_1) = P(\omega_2)$ and $p(x \mid \omega_i) \sim N\left(\mu_i, \sigma_i^2\right)$. Let $\mu_1 = 0, \sigma_1^2 = 1, \mu_2 = \mu$, and $\sigma_2^2 = \sigma^2$.

   (a) Derive a general expression for the location of the Bayes optimal decision boundary as a function of μ and σ^2.

   To make decision, we need to compare $p(\omega_1|x) = \frac{p(x|\omega_1)p(\omega_1)}{p(x)}$ and $p(\omega_2|x) = \frac{p(x|\omega_2)p(\omega_2)}{p(x)}$. Since $p(\omega_1) = p(\omega_2)$, we can simplify it to be the comparison between $p(x|\omega_1)$ and $p(x|\omega_2)$. The optimal decision boundary is where $p(x|\omega_1) = p(x|\omega_2)$.

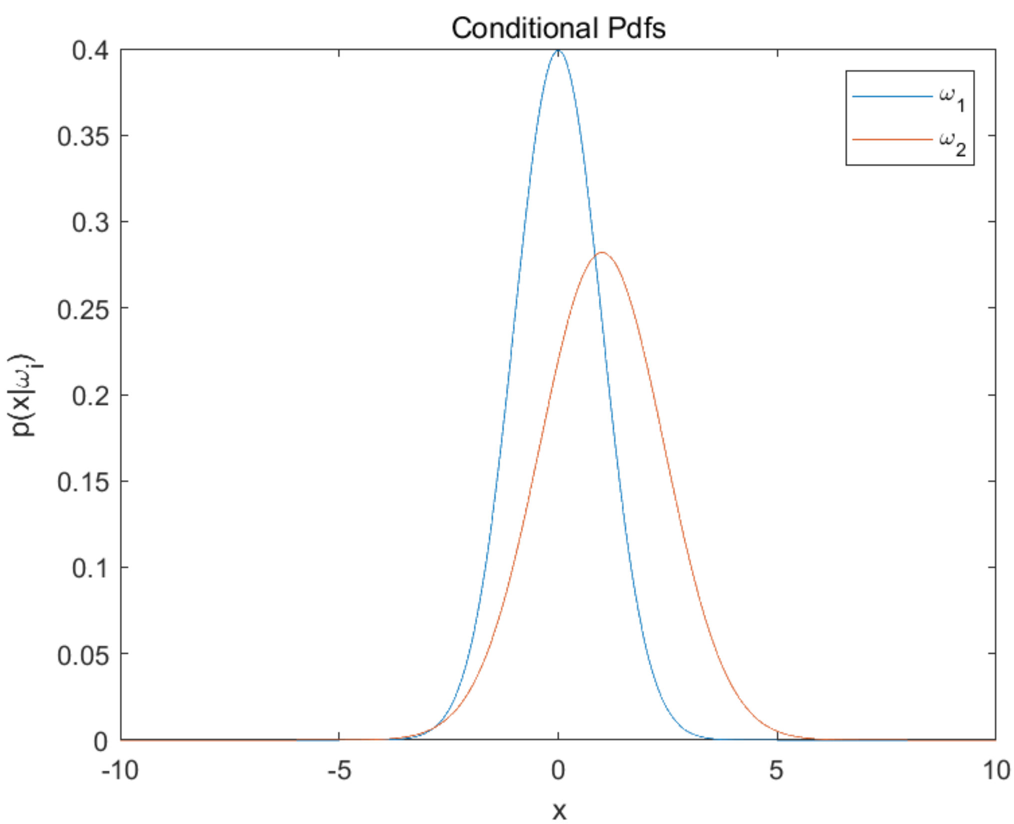
   Here we have $p(x|\omega_1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, and $p(x|\omega_2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Then we get:

   $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \rightarrow \ln e^{-\frac{x^2}{2}} = \ln \frac{1}{\sqrt{\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \rightarrow \frac{x^2}{2} = \frac{1}{2}\ln\sigma^2 + \frac{(x-\mu)^2}{2\sigma^2}$

   Therefore, the Bayes optimal decision boundary can be expressed as $\sigma^2 x^2 = \sigma^2 \ln\sigma^2 + (x-\mu)^2$. When $(\sigma^2 - 1)x^2 + 2\mu x - \mu^2 - \sigma^2\ln\sigma^2 < 0$, we decide ω_1, otherwise we decide ω_2.

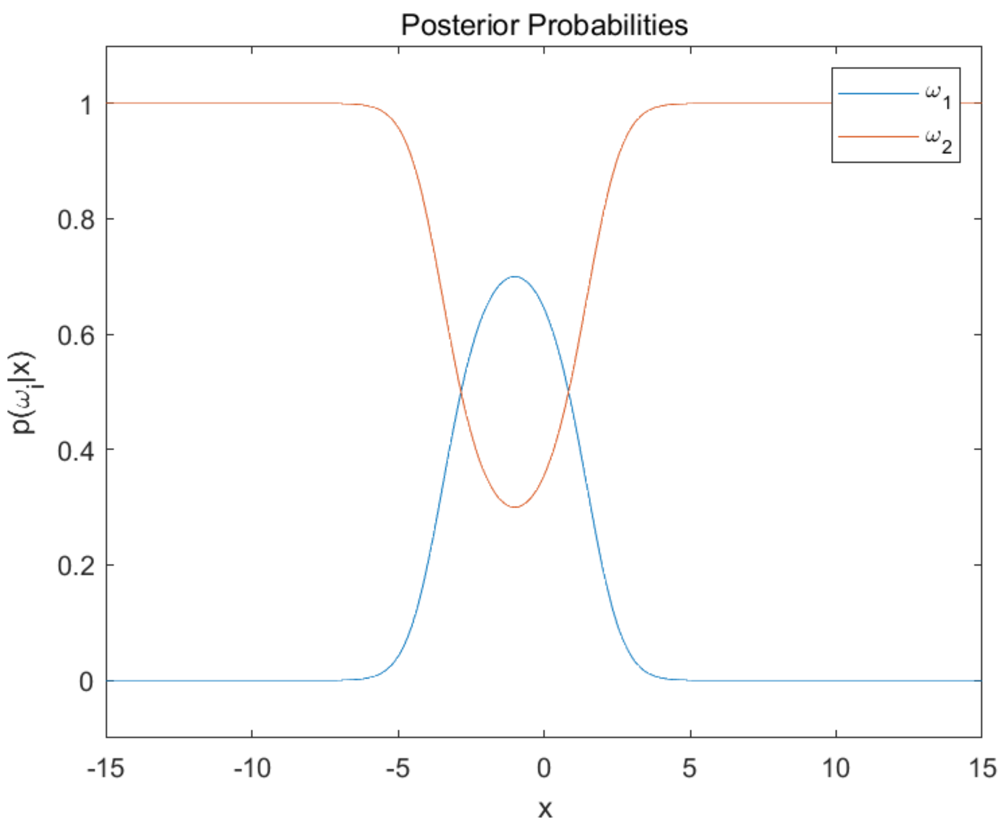
   (b) With $\mu = 1$ and $\sigma^2 = 2$, make two plots using MATLAB: one for the class conditional pdfs $p(x \mid \omega_i)$ and one for the posterior probabilities $p(\omega_i \mid x)$ with the location of the optimal decision regions. Make sure the plots are correctly labeled (axis, titles, legend, etc) and that the fonts are legible when printed.

   For this case, the class conditional pdfs are $p(x|\omega_1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, and $p(x|\omega_2) = \frac{1}{\sqrt{4\pi}}e^{-\frac{(x-1)^2}{4}}$. The corresponding graphs plotted with MATLAB is shown below:



   The posterior probabilities can be expressed as $p(\omega_1|x) = \frac{p(x|\omega_1)p(\omega_1)}{p(x)}$, and $p(\omega_2|x) = \frac{p(x|\omega_2)p(\omega_2)}{p(x)}$.

Since $p(x) = p(x|\omega_1)p(\omega_1) + p(x|\omega_2)p(\omega_2)$, and $p(\omega_1) = p(\omega_2)$, we can derive that $p(\omega_1|x) = \frac{p(x|\omega_1)}{p(x|\omega_1)+p(x|\omega_2)}$, and $p(\omega_2|x) = \frac{p(x|\omega_2)}{p(x|\omega_1)+p(x|\omega_2)}$.
Therefore we have:
$p(\omega_1|x) = \frac{\sqrt{2}e^{-x^2/2}}{\sqrt{2}e^{-x^2/2}+e^{-(x-1)^2/4}}$, and $p(\omega_2|x) = \frac{e^{-(x-1)^2/4}}{\sqrt{2}e^{-x^2/2}+e^{-(x-1)^2/4}}$.
The corresponding graphs plotted with MATLAB is shown below:



(c) Estimate the Bayes error rate p_e.

Let A_1 denote the area where we decide ω_1, and A_2 denote the area where we decide ω_2. We have $p_e = \sum_i p(\omega_i)p(x \notin A_i|\omega_i)$, and $p(\omega_1) = p(\omega_2) = 1/2 = 0.5$.
Then we get:

$$p_e = p(\omega_1)\int_{A_2} p(x|\omega_1)dx + p(\omega_2)\int_{A_1} p(x|\omega_2)dx$$

where A_i can be determined according to the derivation in (a).

i. Consider $\sigma^2 = 1$ and $\mu = 0$: $p_e = 0$.

ii. Consider $\sigma^2 = 1$ and $\mu \neq 0$: when $2\mu x - \mu^2 < 0 \rightarrow x < \frac{\mu}{2}$, we decide ω_1, otherwise we decide ω_2. Therefore, $p_e = 0.5\int_{\mu/2}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + 0.5\int_{-\infty}^{\mu/2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$.

iii. Consider $\sigma^2 < 1$: let x_1 and x_2 be the two solutions of $\sigma^2 x^2 = \sigma^2 \ln \sigma^2 + (x-\mu)^2$, where $x_1 < x_2$. When $x < x_1$ or $x > x_2$, we decide ω_1, otherwise we decide ω_2. Therefore, $p_e = 0.5\int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + 0.5\int_{-\infty}^{x_1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + 0.5\int_{x_2}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$.

iv. Consider $\sigma^2 > 1$: let x_1 and x_2 be the two solutions of $\sigma^2 x^2 = \sigma^2 \ln \sigma^2 + (x-\mu)^2$, where $x_1 < x_2$. When $x_1 < x < x_2$, we decide ω_1, otherwise we decide ω_2. Therefore, $p_e = 0.5 \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + 0.5 \int_{-\infty}^{x_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + 0.5 \int_{x_2}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$

(d) Comment on the case where $\mu = 0$, and σ^2 is much greater than 1. Describe a practical example of a pattern classification problem where such a situation might arise.

When $\mu = 0$, and σ^2 is much greater than 1, the distribution of samples in ω_2 would be much more scattered than in ω_1, which means that samples are more likely to be decided as ω_2 when x tends to be extremely large and small. However, when x is around 0, they are more likely to be decided as ω_1.

Consider two regions on earth. One of them is a plain region (ω_1), and the other is with complex topography (ω_2). Suppose that the mean altitudes of the cities on the two regions are both 0. Here we can arise a classification problem that satisfy the requirements: given the altitude of a city, to which region is it more likely to belong?

It's obvious that there are more cities on ω_1 with altitudes around 0 than on ω_2, but mountains and canyons are more likely to be found on ω_2. Hence, when the altitude is around 0, we tend to consider it to be part of ω_1, otherwise we would say it's part of ω_2.