Jared Yu
Data Mining

Module 3 Assignment

1. Let the conditional densities for a two-category one-dimensional problem be given by the Cauchy distribution

$$p(x|\omega_i) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x - a_i}{b}\right)^2}, \qquad i = 1,2.$$

   a. If $P(\omega_1) = P(\omega_2)$, show that $P(\omega_1|x) = P(\omega_2|x)$ if $x = (1/2)(a_1 + a_2)$. Sketch $P(\omega_1|x)$ for the case $a_1 = 3$, $a_2 = 2$, $b = 5$. How does $P(\omega_1|x)$ behave as $x \to -\infty$? as $x \to \infty$?

Ans:

The first goal is to show that $P(\omega_1|x) = P(\omega_2|x)$ if $x = (1/2)(a_1 + a_2)$.

$$P(\omega_1|x) = P(\omega_2|x)$$

$$\frac{p(x|\omega_1)P(\omega_1)}{p(x)} = \frac{p(x|\omega_2)P(\omega_2)}{p(x)}$$

It is given also that $P(\omega_1) = P(\omega_2)$.

$$p(x|\omega_1) = p(x|\omega_2)$$

$$\frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x - a_1}{b}\right)^2} = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x - a_2}{b}\right)^2}$$

$$\left(\frac{x - a_1}{b}\right)^2 = \left(\frac{x - a_2}{b}\right)^2$$

$$(x - a_1)^2 = (x - a_2)^2$$

$$(x - a_1)^2 - (x - a_2)^2 = 0$$

$$[(x - a_1) - (x - a_2)][(x - a_1) + (x - a_2)] = 0$$

$$(a_2 - a_1)[2x - a_1 - a_2] = 0$$

$$2x - a_1 - a_2 = 0$$

$$x = \frac{a_1 + a_2}{2}$$

So, it has been shown that when $x = \frac{a_1+a_2}{2}$, then $P(\omega_1|x) = P(\omega_2|x)$. This works when $a_1 \neq a_2$. ■

The next goal is to sketch $P(\omega_1|x)$ for the case $a_1 = 3$, $a_2 = 2$, $b = 5$. The formula for $P(\omega_1|x)$ is as follows:

$$\frac{p(x|\omega_1)P(\omega_1)}{p(x)} = \frac{\left[\frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x - a_1}{b}\right)^2}\right](0.5)}{\sum_{k=1}^{2} p(x|\omega_k)P(\omega_k)}$$

$$= \frac{\left[\frac{1}{\pi b} \cdot \frac{1}{1+\left(\frac{x-a_1}{b}\right)^2}\right](0.5)}{\left\{\left[\left[\frac{1}{\pi b} \cdot \frac{1}{1+\left(\frac{x-a_1}{b}\right)^2}\right]\right] * (0.5)\right\} + \left\{\left[\left[\frac{1}{\pi b} \cdot \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}\right]\right] * (0.5)\right\}}.$$

It is possible to simplify it further, but it was calculated in RStudio. Therefore, each of the components (i.e., likelihood, prior, and posterior) were calculated separately and combined into a single function. Then, the result was calculated for a series of x-values between $[-300, 300]$.

Below in Figure 1 is a plot $P(\omega_1|x)$ for the case $a_1 = 3$, $a_2 = 2$, $b = 5$. It is assumed that $P(\omega_1|x) = P(\omega_2|x)$. The range of x-values is between $[-300, 300]$. It can be seen that as $x \to \pm\infty$, $P(\omega_1|x)$ starts to approach 0.5.

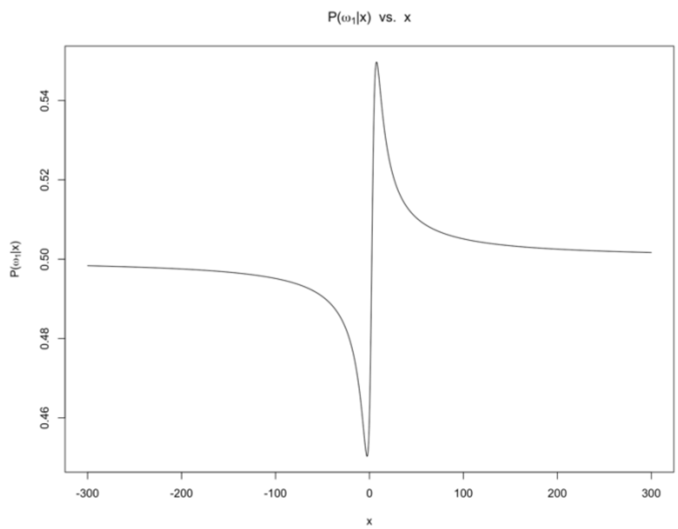

Figure 1 The above figure shows a plot of $P(\omega_1|x)$ against x for $x \in [-300, 300]$. It can be seen that the probability converges towards 0.5 as $x \to \pm\infty$.

b. Using the conditional densities in part a, and assuming equal *a priori* probabilities, show that the minimum probability of error is given by

$$P(\text{error}) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2 - a_1}{2b}\right|.$$

Sketch this as a function of $|(a_2 - a_1)/b|$.

Ans: Reference ([1])

From part a we have that $x = (1/2)(a_1 + a_2)$ is the decision boundary, since at that point $P(\omega_1|x) = P(\omega_2|x)$. Furthermore, we can state that we'd classify a point x^* as ω_1 if $P(\omega_1|x) \geq P(\omega_2|x)$. Using this inequality, it would also imply that for $a_2 \geq a_1$, we'd classify the point as belonging to ω_1 if $x \leq \frac{a_1+a_2}{2}$.

The reasoning is as follows:

$$P(\omega_1|x) \geq P(\omega_2|x)$$

$$\vdots$$
$$(a_2 - a_1)[2x - a_1 - a_2] \le 0$$
Without loss of generality, let us assume that $a_2 > a_1$ (we know $a_1 = a_2$ does not work from part a). This helps to prevent exploring too many cases associated with the inequality. From there we have:
$$\begin{cases} a_2 - a_1 \ge 0 \\ 2x - a_1 - a_2 \le 0 \end{cases} = \begin{cases} a_2 \ge a_1 \\ x \le \dfrac{a_1 + a_2}{2} \end{cases}$$

The average probability of error is as follows:
$$P(\text{error}) = \int_{-\infty}^{\infty} P(\text{error}, x)dx = \int_{-\infty}^{\infty} P(\text{error}|x)p(x)dx$$
$$= \int_{-\infty}^{\frac{a_1+a_2}{2}} P(\text{error}|x)p(x)dx + \int_{\frac{a_1+a_2}{2}}^{\infty} P(\text{error}|x)p(x)dx$$
$$= \int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{p(x|\omega_2)P(\omega_2)}{p(x)} p(x)dx + \int_{\frac{a_1+a_2}{2}}^{\infty} \frac{p(x|\omega_1)P(\omega_1)}{p(x)} p(x)dx$$

The reason is that probably of the error given the data is as follows,
$$P(\text{error}|x) = \begin{cases} p(\omega_1|x), & \text{if we decide } \omega_2 \\ p(\omega_2|x), & \text{if we decide } \omega_1 \end{cases}$$
$$= \begin{cases} \dfrac{p(x|\omega_1)P(\omega_1)}{p(x)}, & \text{if we decide } \omega_2 \\ \dfrac{p(x|\omega_2)P(\omega_2)}{p(x)}, & \text{if we decide } \omega_1. \end{cases}$$

$$\cdots = \int_{-\infty}^{\frac{a_1+a_2}{2}} p(x|\omega_2)P(\omega_2)dx + \int_{\frac{a_1+a_2}{2}}^{\infty} p(x|\omega_1)P(\omega_1)dx$$
$$= \frac{1}{2\pi b}\int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{1}{1 + \left(\frac{x-a_2}{b}\right)^2} dx + \frac{1}{2\pi b}\int_{\frac{a_1+a_2}{2}}^{\infty} \frac{1}{1 + \left(\frac{x-a_1}{b}\right)^2} dx$$
Then, do a change of variables $y = \frac{x-a_2}{b}$ and $z = \frac{x-a_1}{b}$.
$$\cdots = \frac{1}{2\pi}\left[\int_{-\infty}^{\frac{a_1-a_2}{2b}} \frac{1}{1+y^2} dy + \int_{\frac{a_2-a_1}{2b}}^{\infty} \frac{1}{1+z^2} dz\right]$$
$$= \frac{1}{2\pi}\left[\tan^{-1} y\Big|_{-\infty}^{\frac{a_1-a_2}{2b}} + \tan^{-1} z\Big|_{\frac{a_2-a_1}{2b}}^{\infty}\right]$$
$$= \frac{1}{2\pi}\left[\tan^{-1}\frac{a_1-a_2}{2b} + \frac{\pi}{2} + \frac{\pi}{2} - \tan^{-1}\frac{a_2-a_1}{2b}\right]$$
Also, since arctangent is an odd function, $\tan^{-1}(-x) = -\tan^{-1}(x)$. Here, we have that $a_1 > a_1$ and so $\frac{a_1-a_2}{2b}$ is the negative of $\frac{a_2-a_1}{2b}$. Therefore, $\tan^{-1}\frac{a_1-a_2}{2b} = -\tan^{-1}\frac{a_2-a_1}{2b}$.

$$\cdots = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{a_2 - a_1}{2b}$$

Going back to $(a_2 - a_1)[2x - a_1 - a_2] \leq 0$, if $a_1 > a_2$, then the rule for classifying ω_1 would be lead to $x \geq \frac{a_1+a_2}{2}$. The result of this is that $P(\text{error}) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{a_1-a_2}{2b}$. Thus, it follows then that,

$$P(\text{error}) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2 - a_1}{2b}\right|. \blacksquare$$

The sketch of this $P(\text{error})$ as a function of $|(a_2 - a_1)/b|$ can be seen below.

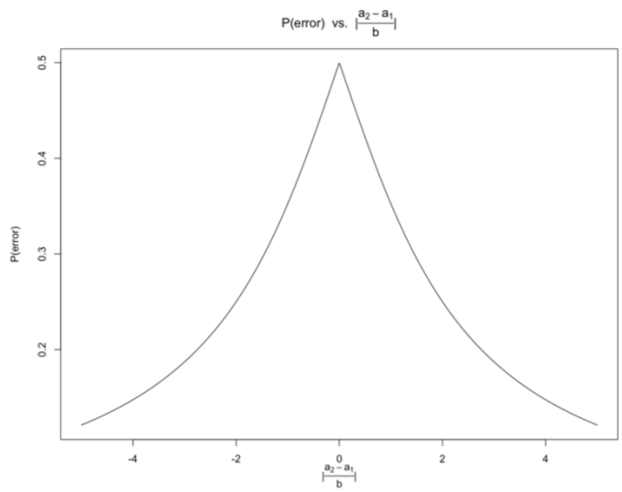

Figure 2 The above plot shows $P(error)$ as a function of $[(a_2 - a_1)/b]$.

2. The Poisson distribution for discrete k, $k = 0, 1, 2, \cdots$ and real parameter λ is

$$P(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

   a. Find the mean of k.

<u>Ans:</u> Reference ([2])

$$\mu = E(K) = \sum_{k=0}^{\infty} kp(k|\lambda) = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = 0 + \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!}$$

Next, consider the change of variables, $x = k - 1$.

$$\cdots = \sum_{(x+1)=1}^{\infty} (x+1)e^{-\lambda}\frac{\lambda^{x+1}}{(x+1)!} = \lambda\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = \lambda\sum_{x=0}^{\infty} p(x|\lambda) = \lambda$$

In the end, the expected value was shown to be λ multiplied by the sum over all possible values of an alternate Poisson probability mass function ($p(x|\lambda)$) which equates to 1. $\blacksquare$

   b. Find the variance of k.

<u>Ans:</u> Reference ([3])

$$\sigma^2 = Var(K) = E(K^2) - E(K)^2 = E[K(K-1) + K] - E(K)^2$$
$$= E[K(K-1)] + E(K) - E(K)^2$$

First, let us solve for $E[K(K-1)]$:

$$E[K(K-1)] = \sum_{k=0}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = 0 + 0 + \sum_{k=2}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-2)!}$$

Next, consider the change of variables, $x = k - 2$.

$$\cdots = \sum_{(x+2)=2}^{\infty} e^{-\lambda}\frac{\lambda^{(x+2)}}{[(x+2)-2]!} = \lambda^2 \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = \lambda^2 \sum_{x=0}^{\infty} p(x|\lambda) = \lambda^2$$

In the end, the expected value was shown to be λ^2 multiplied by the sum over all possible values of an alternate Poisson probability mass function ($p(x|\lambda)$) which equates to 1. Therefore, $Var(K) = E[K(K-1)] + E(K) - E(K)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$, since it has been shown already that $E(K) = \lambda$. ■

c. Find the mode of k.

Ans: (Reference: [4])

There are two possibilities, whether λ is an integer or a fraction. In the case of λ being an integer, the probability mass function (pmf) has a bimodal distribution and therefore two distinct modes. In the case of λ being a fraction, then the pmf has a unimodal distribution and therefore only one mode.

First, let us consider the case when λ is an integer. Let us look then at the pmf itself,

$$P(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

The mode is the value of k that maximizes the probability mass of this pmf. Looking at the pmf, it is apparent that the "$e^{-\lambda}$" term remains constant and does not vary depending on k. It will first be noted though that the values of k that maximize $\frac{\lambda^k}{k!}$ are λ and $\lambda - 1$. When k is equal to either of these values, they are equivalent. This will be shown as follows:

$$\frac{\lambda^{\lambda-1}}{(\lambda-1)!} = \frac{\lambda^{\lambda}\lambda^{-1}}{(\lambda-1)!} = \frac{\lambda^{\lambda}}{\lambda(\lambda-1)!} = \frac{\lambda^{\lambda}}{\lambda!}.$$

To show that this is the maximum probability for $P(k|\lambda)$, we will consider two alternate values of k, $\lambda + 1$ and $\lambda - 2$.

Case 1: $\lambda + 1$

$$\frac{\lambda^{\lambda+1}}{(\lambda+1)!} = \frac{\lambda^{\lambda}\lambda}{(\lambda+1)\cdot\lambda\cdot(\lambda-1)!} = \frac{\lambda^{\lambda}}{(\lambda+1)\cdot(\lambda-1)!}$$

In the above situation, the denominator is obviously larger than in the situation before with $\frac{\lambda^{\lambda}}{\lambda!}$, therefore it is a smaller number overall. This would be the case for any integer $\lambda + 1$ or larger.

Case 2: $\lambda - 2$

$$\frac{\lambda^{\lambda-2}}{(\lambda-2)!} = \frac{\lambda^{\lambda}\lambda^{-2}}{(\lambda-2)!} = \frac{\lambda^{\lambda}}{\lambda^2(\lambda-2)!}$$

Again, like with $\lambda + 1$, the denominator is larger than before with $\frac{\lambda^{\lambda}}{\lambda!}$, therefore it is also a smaller overall value. This would apply for any integer smaller than $\lambda - 1$. Therefore, it has been shown that in the case of λ being an integer, the mode is both λ and $\lambda - 1$.

Next, let us consider the alternate situation when λ is a fraction. The Poisson distribution is a discrete density function and so it can only take on values that are integers. Therefore, the mode of it must also be an integer. So, although λ itself is possibly a fraction, the mode must be an integer.

We will begin by looking at the pmf again carefully,

$$P(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Once again, the "$e^{-\lambda}$" term does not depend on k, and so it is not directly important for finding the k value that maximizes the probability for a given (arbitrary) λ. So, again we focus on the term $\frac{\lambda^k}{k!}$. In the numerator, it shows "λ^k" which is always increasing as k increases. However, the denominator "$k!$" is also increasing as k increases. The goal then is to find the integer k at which this $\frac{\lambda^k}{k!}$ is maximum for a given (fractional) λ.

Let us consider two cases, when $\lambda < 1$ and $\lambda > 1$. We first begin with the former.

Case: $\lambda < 1$
Again, let's take a look at the pmf,

$$P(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Similar to before, we are looking for the k that maximizes the pmf for some arbitrary given λ. In the same way, we can ignore $e^{-\lambda}$, since the choice of k does not impact this value. Looking at the numerator, "λ^k", since λ must be a positive fraction less than 1, then as k increases it is strictly decreasing. In fact, for some given arbitrary $\lambda < 1$, the largest value of the pmf is at $k = 0$, which evaluates to $e^{-\lambda}$. All other values of k start to decrease the overall value, since $\frac{\lambda^k}{k!}$ will itself be a fraction less than 1. In this case, when $\lambda < 1$, the mode is $\lfloor\lambda\rfloor = 0$, where $\lfloor\lambda\rfloor$ is the floor of λ.

Case: $\lambda > 1$
The last case is when λ is a fraction greater than 1. To understand, we can think about some example as follows (*Note: Let $P(k|\lambda)$ be the equivalent of $P(K = k)$*):

$$P(K = 0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-\lambda}$$

$$P(K = 1) = e^{-\lambda}\frac{\lambda^1}{1!} = \frac{\lambda}{1}e^{-\lambda}$$

$$P(K = 2) = e^{-\lambda}\frac{\lambda^2}{2!} = \frac{\lambda^2}{2 \times 1}e^{-\lambda}$$

$$\vdots$$

$$P(K = \lfloor\lambda\rfloor - 1) = e^{-\lambda}\frac{\lambda^{\lfloor\lambda\rfloor - 1}}{(\lfloor\lambda\rfloor - 1)!} = \frac{\lambda^{\lfloor\lambda\rfloor}}{(\lfloor\lambda\rfloor - 1) \times (\lfloor\lambda\rfloor - 2) \times \cdots \times 2 \times 1}e^{-\lambda}$$

$$P(K = \lfloor\lambda\rfloor) = e^{-\lambda}\frac{\lambda^{\lfloor\lambda\rfloor}}{\lfloor\lambda\rfloor!} = \frac{\lambda^{\lfloor\lambda\rfloor}}{\lfloor\lambda\rfloor \times (\lfloor\lambda\rfloor - 1) \times \cdots \times 2 \times 1}e^{-\lambda}$$

$$P(K = \lfloor\lambda\rfloor + 1) = e^{-\lambda} \frac{\lambda^{\lfloor\lambda\rfloor+1}}{(\lfloor\lambda\rfloor + 1)!} = \frac{\lambda^{\lfloor\lambda\rfloor+1}}{(\lfloor\lambda\rfloor + 1) \times \lfloor\lambda\rfloor \times \cdots \times 2 \times 1} e^{-\lambda}$$

At each step, for the next $P(K = k + 1)$, the difference between it and $P(K = k)$ is that the next equation is multiplied by $\frac{\lambda}{k}$. This $\frac{\lambda}{k}$ term is always positive fraction but increases and then later decreases. When $k < \lambda$, then $\frac{\lambda}{k}$ is greater than 1. When $k > \lambda$, then $\frac{\lambda}{k}$ is less than 1. Therefore, up until $P(K = \lfloor\lambda\rfloor + 1)$, the output for $P(k|\lambda)$ will constantly be increasing by a factor larger than 1. However, at $P(K = \lfloor\lambda\rfloor)$, the output will maximize, because at $P(K = \lfloor\lambda\rfloor + 1)$ and beyond, it will continually multiple by a $\frac{\lambda}{k}$ term that is less than 1. Therefore, the mode of the Poisson distribution when λ is a fraction is $\lfloor\lambda\rfloor$.

To summarize, if λ is an integer, then the mode is both λ and $\lambda - 1$. If λ is a non-integer fraction, then the mode is $\lfloor\lambda\rfloor$. ■

d. Assume two categories C_1 and C_2, equally probable *a priori*, distributed with Poisson distributions and $\lambda_1 > \lambda_2$. What is the Bayes classification decision?

Ans:
The Bayes decision rule is as follows,

$$\text{Decide} \begin{cases} C_1, & \text{if } p(k|\lambda_1)P(\lambda_1) > p(k|\lambda_2)P(\lambda_2) \\ C_2, & \text{otherwise.} \end{cases}$$

Here, it is assumed that $P(\lambda_1) = P(\lambda_2)$, therefore we only need to consider if $p(k|\lambda_1) > p(k|\lambda_2)$. (*Note: Here, $p(k|\lambda_1)$ and $p(k|\lambda_2)$ are the likelihood functions rather than the probability mass functions.*)

The likelihood function for the Poisson distribution with parameter λ_j is as follows:

$$L = p(k|\lambda_j) = e^{-\lambda_j} \frac{\lambda_j^k}{k!}.$$

Therefore, the decision rule can be rewritten as follows,

$$\text{Decide} \begin{cases} C_1, & \text{if } p(k|\lambda_1) > p(k|\lambda_2) \\ C_2, & \text{otherwise,} \end{cases}$$

where

$$p(k|\lambda_1) = e^{-\lambda_1} \frac{\lambda_1^k}{k!} \text{ and } p(k|\lambda_2) = e^{-\lambda_2} \frac{\lambda_2^k}{k!}.$$

e. What is the Bayes error rate?

Ans:
We choose to classify an example to C_1 if

$$P(\lambda_1|k) \geq P(\lambda_2|k).$$

Then let the following be shown:

$$\frac{p(k|\lambda_1)P(\lambda_1)}{p(k)} \geq \frac{p(k|\lambda_2)P(\lambda_2)}{p(k)}$$

$$p(k|\lambda_1) \geq p(k|\lambda_2)$$

$$e^{-\lambda_1} \frac{\lambda_1^k}{k!} \geq e^{-\lambda_2} \frac{\lambda_2^k}{k!}$$

$$e^{-\lambda_1}\lambda_1^k \geq e^{-\lambda_2}\lambda_2^k$$
$$\frac{e^{-\lambda_1}}{e^{-\lambda_2}} \geq \frac{\lambda_2^k}{\lambda_1^k}$$
$$e^{-\lambda_1+\lambda_2} \geq \left(\frac{\lambda_2}{\lambda_1}\right)^k$$
$$-\lambda_1 + \lambda_2 \geq k \ln\frac{\lambda_2}{\lambda_1}$$

Note: $\ln\frac{\lambda_2}{\lambda_1} < 0$ *since* $\lambda_1 > \lambda_2$.

$$k \geq \frac{-\lambda_1 + \lambda_2}{\ln\frac{\lambda_2}{\lambda_1}}$$
$$k \geq \frac{\lambda_2 - \lambda_1}{\ln\lambda_2 - \ln\lambda_1}$$

$$P(\text{error}) = \sum_{k=0}^{\infty} P(\text{error}, k) = \sum_{k=0}^{\infty} P(\text{error}|k)p(k)$$
$$= \sum_{k=0}^{k_B-1} P(\text{error}|k)p(k) + \sum_{k=k_B}^{\infty} P(\text{error}|k)p(k)$$

Let $k_B = \left\lceil\frac{\lambda_2-\lambda_1}{\ln\lambda_2-\ln\lambda_1}\right\rceil$, which is the ceiling of $\frac{\lambda_2-\lambda_1}{\ln\lambda_2-\ln\lambda_1}$.

$$= \sum_{k=0}^{k_B-1} p(k|\lambda_1)P(\lambda_1) + \sum_{k=k_B}^{\infty} p(k|\lambda_2)P(\lambda_2) = 0.5\left[\sum_{k=0}^{k_B-1} p(k|\lambda_1) + \sum_{k=k_B}^{\infty} p(k|\lambda_2)\right]$$
$$= \boxed{0.5\left[\sum_{k=0}^{k_B-1} e^{-\lambda_1}\frac{\lambda_1^k}{k!} + \sum_{k=k_B}^{\infty} e^{-\lambda_2}\frac{\lambda_2^k}{k!}\right]}$$

Here, $0.5\left[\sum_{k=0}^{k_B-1} e^{-\lambda_1}\frac{\lambda_1^k}{k!} + \sum_{k=k_B}^{\infty} e^{-\lambda_2}\frac{\lambda_2^k}{k!}\right]$ Is the Bayes error rate.

3. Let $p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \sigma^2 I)$ for a two-category k-dimensional problem with $P(\omega_1) = P(\omega_2) = \frac{1}{2}$.
   a. Find P_e, the minimum probability of error.

<u>Ans:</u>

We can choose to classify an observation to ω_1 if
$$P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x}).$$
Then let the following be shown:
$$\frac{p(\mathbf{x}|\omega_1)P(\omega_1)}{p(\mathbf{x})} \geq \frac{p(\mathbf{x}|\omega_2)P(\omega_2)}{p(\mathbf{x})}$$
$$p(\mathbf{x}|\omega_1) \geq p(\mathbf{x}|\omega_2)$$

$$\frac{1}{(2\pi)^{\frac{k}{2}}|\sigma^2 I|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^\top(\sigma^2 I)^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\right]$$
$$\geq \frac{1}{(2\pi)^{\frac{k}{2}}|\sigma^2 I|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^\top(\sigma^2 I)^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\right]$$
$$\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^\top(\sigma^2 I)^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\right] \geq \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^\top(\sigma^2 I)^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\right]$$
$$(\mathbf{x}-\boldsymbol{\mu}_1)^\top\left(\frac{1}{\sigma^2}I\right)(\mathbf{x}-\boldsymbol{\mu}_1) \geq (\mathbf{x}-\boldsymbol{\mu}_2)^\top\left(\frac{1}{\sigma^2}I\right)(\mathbf{x}-\boldsymbol{\mu}_2)$$
$$(\mathbf{x}-\boldsymbol{\mu}_1)^\top(\mathbf{x}-\boldsymbol{\mu}_1) \geq (\mathbf{x}-\boldsymbol{\mu}_2)^\top(\mathbf{x}-\boldsymbol{\mu}_2)$$
$$\mathbf{x}^\top\mathbf{x}-\mathbf{x}^\top\boldsymbol{\mu}_1-\boldsymbol{\mu}_1^\top\mathbf{x}+\boldsymbol{\mu}_1^\top\boldsymbol{\mu}_1 \geq \mathbf{x}^\top\mathbf{x}-\mathbf{x}^\top\boldsymbol{\mu}_2-\boldsymbol{\mu}_2^\top\mathbf{x}+\boldsymbol{\mu}_2^\top\boldsymbol{\mu}_2$$
$$-2\mathbf{x}^\top\boldsymbol{\mu}_1+\boldsymbol{\mu}_1^\top\boldsymbol{\mu}_1 \geq -2\mathbf{x}^\top\boldsymbol{\mu}_2+\boldsymbol{\mu}_2^\top\boldsymbol{\mu}_2$$
$$2\mathbf{x}^\top\boldsymbol{\mu}_2-2\mathbf{x}^\top\boldsymbol{\mu}_1 \geq \boldsymbol{\mu}_2^\top\boldsymbol{\mu}_2-\boldsymbol{\mu}_1^\top\boldsymbol{\mu}_1$$
$$2\mathbf{x}^\top(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1) \geq \boldsymbol{\mu}_2^\top\boldsymbol{\mu}_2-\boldsymbol{\mu}_1^\top\boldsymbol{\mu}_1$$

If $\mathbf{x}=\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$, then the following holds:

$$2\left[\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\right]^\top(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)=\boldsymbol{\mu}_2^\top\boldsymbol{\mu}_2-\boldsymbol{\mu}_1^\top\boldsymbol{\mu}_1.$$

So, we classify an example to ω_1 if $\mathbf{x}\geq\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$.

Let $R_1=\left\{\mathbf{x}\colon \mathbf{x}\geq\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\right\}$ and $R_2=\left\{\mathbf{x}\colon \mathbf{x}<\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\right\}$.

$$P_e=P(\text{error})=\int_{R_1}P(\text{error}|\mathbf{x})p(\mathbf{x})d\mathbf{x}+\int_{R_2}P(\text{error}|\mathbf{x})p(\mathbf{x})d\mathbf{x}$$
$$=\int_{R_1}P(\mathbf{x}|\omega_2)P(\omega_2)d\mathbf{x}+\int_{R_2}P(\mathbf{x}|\omega_1)P(\omega_1)d\mathbf{x}$$
$$=\frac{1}{2}\left[\int_{R_1}P(\omega_2|\mathbf{x})d\mathbf{x}+\int_{R_2}P(\omega_1|\mathbf{x})d\mathbf{x}\right]$$
$$\boxed{=\frac{1}{2}\left[\int_{R_1}\frac{1}{(2\pi\sigma^2)^{k/2}}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_2)^\top(\mathbf{x}-\boldsymbol{\mu}_2)\right]d\mathbf{x}+\int_{R_2}\frac{1}{(2\pi\sigma^2)^{k/2}}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_1)^\top(\mathbf{x}-\boldsymbol{\mu}_1)\right]d\mathbf{x}\right]}$$

b. Let $\mu_1=\mathbf{0}$ and $\mu_2=(m_1,\cdots,m_k)^\top\neq\mathbf{0}$. Show that $P_e\to 0$ as the dimension k approaches infinity. Assume that $\sum_{k=1}^{\infty}m_k^2\to\infty$.

Ans:

From part a, we have that

$$P_e=\frac{1}{2}\left[\int_{R_1}\frac{1}{(2\pi\sigma^2)^{k/2}}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_2)^\top(\mathbf{x}-\boldsymbol{\mu}_2)\right]d\mathbf{x}\right.$$
$$\left.+\int_{R_2}\frac{1}{(2\pi\sigma^2)^{k/2}}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_1)^\top(\mathbf{x}-\boldsymbol{\mu}_1)\right]d\mathbf{x}\right].$$

Given that $\mu_1=\mathbf{0}$ and $\mu_2=(m_1,\cdots,m_k)^\top\neq\mathbf{0}$, then P_e can be rewritten as follows:

$$P_e=\frac{1}{2}\left[\int_{R_1}\frac{1}{(2\pi\sigma^2)^{k/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{k}(x_i-m_i)^2\right]d\mathbf{x}+\int_{R_2}\frac{1}{(2\pi\sigma^2)^{k/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{k}x_i^2\right]d\mathbf{x}\right]$$

We have also seen before that the decision boundary is at $\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$, which in part b evaluates to,

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \frac{1}{2}\boldsymbol{\mu}_2 = \begin{bmatrix} \frac{m_1}{2} \\ \frac{m_2}{2} \\ \vdots \\ \frac{m_k}{2} \end{bmatrix}.$$

Therefore, if an example falls in

$$R_1 = \left\{\mathbf{x}: \mathbf{x} \geq \frac{1}{2}\boldsymbol{\mu}_2\right\},$$

then it is classified at ω_1, otherwise it is classified as ω_2 and falls in

$$R_2 = \left\{\mathbf{x}: \mathbf{x} < \frac{1}{2}\boldsymbol{\mu}_2\right\}.$$

To show that $P_e \to 0$ as the dimension k approaches infinity, let us refer back to P_e. It has been divided into two parts, blue and red. First, we look at the blue section:

$$\int_{R_1} \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{k}(x_i - m_i)^2\right] d\mathbf{x}$$

$$= \prod_{i=1}^{k} \int_{\frac{1}{2}m_i}^{\infty} \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}(x_i - m_i)^2\right]}_{N(m_i,\sigma^2)} dx_i$$

$$= \prod_{i=1}^{k}\left[1 - \Phi\left(\frac{\left(\frac{1}{2}m_i\right) - m_i}{\sigma}\right)\right] = \prod_{i=1}^{k}\left[1 - \Phi\left(-\frac{m_i}{2\sigma}\right)\right]$$

Next, let us look at the red section:

$$\int_{R_2} \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{k} x_i^2\right] d\mathbf{x}$$

$$= \prod_{i=1}^{k} \int_{-\infty}^{\frac{1}{2}m_i} \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}x_i^2\right]}_{N(0,\sigma^2)} dx_i$$

$$= \prod_{i=1}^{k} \Phi\left(\frac{m_i}{2\sigma}\right)$$

So, it follows that

$$P_e = \frac{1}{2}\left\{\prod_{i=1}^{k}\left[1 - \Phi\left(-\frac{m_i}{2\sigma}\right)\right] + \prod_{i=1}^{k}\Phi\left(\frac{m_i}{2\sigma}\right)\right\}$$

$$= \frac{1}{2}\left\{\prod_{i=1}^{k}\Phi\left(\frac{m_i}{2\sigma}\right) + \prod_{i=1}^{k}\Phi\left(\frac{m_i}{2\sigma}\right)\right\} = \prod_{i=1}^{k}\Phi\left(\frac{m_i}{2\sigma}\right)$$

where $0 < \Phi\left(\frac{m_i}{2\sigma}\right) < 1$. Therefore, $P_e \to 0$ as the dimension k approaches infinity. ■

4. Under the assumption that $\lambda_{21} > \lambda_{11}$ and $\lambda_{12} > \lambda_{22}$, show that the general minimum risk discriminant function for a classifier with independent binary features is given by $g(\mathbf{x}) = \mathbf{w}^{\top}\mathbf{x} + w_0$. What are $\mathbf{w}$ and w_0?

Ans: (Reference: [5] pp.52-53)

Let $\mathbf{x} = (x_1, \cdots, x_d)^{\top}$, where x_i are either 0 or 1 with probabilities

$$p_i = \Pr[x_i = 1|\omega_1] \text{ and } q_i = \Pr[x_i = 1|\omega_2].$$

The class-conditional probabilities can be written as follows:

$$P(\mathbf{x}|\omega_1) = \prod_{i=1}^{d} p_i^{x_i}(1-p_i)^{1-x_i} \text{ and } P(\mathbf{x}|\omega_2) = \prod_{i=1}^{d} q_i^{x_i}(1-q_i)^{1-x_i}.$$

The likelihood ratio is as follows,

$$\frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)} = \frac{\prod_{i=1}^{d} p_i^{x_i}(1-p_i)^{1-x_i}}{\prod_{i=1}^{d} q_i^{x_i}(1-q_i)^{1-x_i}} = \prod_{i=1}^{d}\left(\frac{p_i}{q_i}\right)^{x_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-x_i}.$$

We are able to write out $P(\mathbf{x}|\omega_1)$ and $P(\mathbf{x}|\omega_2)$ in that format because of the assumption of conditional independence. This problem is a two-category case and so the classifier is known as a dichotomizer. Therefore, the separate discriminant functions can be combined into a single discriminant function. The one used here will be

$$g(\mathbf{x}) = \ln\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} + \ln\frac{P(\omega_1)}{P(\omega_2)}.$$

Using this discriminant function leads to the following,

$$g(\mathbf{x}) = \ln\prod_{i=1}^{d}\left(\frac{p_i}{q_i}\right)^{x_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-x_i} + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

$$= \sum_{i=1}^{d}\left[\ln\left(\frac{p_i}{q_i}\right)^{x_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-x_i}\right] + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

$$= \sum_{i=1}^{d}\left[x_i\ln\left(\frac{p_i}{q_i}\right) + (1-x_i)\ln\left(\frac{1-p_i}{1-q_i}\right)\right] + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

$$= \sum_{i=1}^{d}\left[x_i\ln\left(\frac{p_i}{q_i}\right) + \ln\left(\frac{1-p_i}{1-q_i}\right) - x_i\ln\left(\frac{1-p_i}{1-q_i}\right)\right] + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

$$= \sum_{i=1}^{d}\left\{x_i\left[\ln\left(\frac{p_i}{q_i}\right) - \ln\left(\frac{1-p_i}{1-q_i}\right)\right] + \ln\left(\frac{1-p_i}{1-q_i}\right)\right\} + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

$$= \sum_{i=1}^{d}\left\{x_i\ln\frac{\left(\frac{p_i}{q_i}\right)}{\left(\frac{1-p_i}{1-q_i}\right)} + \ln\left(\frac{1-p_i}{1-q_i}\right)\right\} + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

$$= \sum_{i=1}^{d} \left\{ x_i \ln \frac{p_i(1-q_i)}{q_i(1-p_i)} + \ln\left(\frac{1-p_i}{1-q_i}\right) \right\} + \ln \frac{P(\omega_1)}{P(\omega_2)}$$

Since this discriminant function is linear in terms of x_i, it can be written in the form of

$$g(\mathbf{x}) = \sum_{i=1}^{d} w_i x_i + w_0 = \mathbf{w}^\top \mathbf{x} + w_0,$$

where

$$\mathbf{w} = (w_1, \cdots, w_d)^\top,$$

$$w_i = \ln \frac{p_i(1-q_i)}{q_i(1-p_i)} \quad i = 1, \cdots, d,$$

and

$$w_0 = \sum_{i=1}^{d} \ln \frac{1-p_i}{1-q_i} + \ln \frac{P(\omega_1)}{P(\omega_2)}.$$

References

[1] https://john.cs.olemiss.edu/~ychen/courses/ENGR691F06/hw1/hw1sol.pdf

[2] https://www.statlect.com/probability-distributions/Poisson-distribution

[3] http://llc.stat.purdue.edu/2014/41600/notes/prob1804.pdf

[4] https://www.youtube.com/watch?v=lkjhwyrW8Io

[5] Pattern Classification 2nd Ed. Duda, Hart, and Stork.

Code Appendix

```
library(latex2exp)
### 1
### a
# sketch P(omega_1 | x) = [p(x|omega_1) * P(omega_1)] / p(x)
cauchy <- function(x, a, b) {
  (1 / (pi * b)) * (1 / (1 + ((x - a) / b)^2))
}

a1 <- 3; a2 <- 2; b <- 5; p_omega1 <- 0.5; p_omega2 <- 0.5
p_x <- function(x, p_omega1, p_omega2, a1, a2, b) {
  p_x_given_omega1 <- cauchy(x = x, a = a1, b = b)
  p_x_given_omega2 <- cauchy(x = x, a = a2, b = b)

  prob_x <- (p_x_given_omega1 * p_omega1) +
    (p_x_given_omega2 * p_omega2)

  return(prob_x)
}

prob_x <- p_x(x = xs,
  p_omega1 = p_omega1, p_omega2 = p_omega2,
  a1 = a1, a2 = a2, b = b)

omega_given_x <- function(x, p_omega1, p_omega2,
  a1 = a1, a2 = a2, b = b) {
  # Calculate p(x)
  prob_x <- p_x(x = x,
```

```r
    p_omega1 = p_omega1, p_omega2 = p_omega2,
    a1 = a1, a2 = a2, b = b)

  # Calculate p(x|omega_i)
  p_x_given_omega <- cauchy(x = x, a = a1, b = b)

  cond_prob <- (p_x_given_omega * p_omega1) / prob_x

  return(cond_prob)
}

xs <- seq(-3e2, 3e2, length.out = 1e3)
ys <- omega_given_x(x = xs,
  p_omega1 = p_omega1, p_omega2 = p_omega2,
  a1 = a1, a2 = a2, b = b)

plot(xs, ys, type = 'l',
     main = TeX('$P(\\omega_1 | x) \\; vs.\\; x$'),
     xlab = TeX('$x$'), ylab = TeX('$P(\\omega_1 | x)$'))
### b
prob_error <- function(x) {
  x <- abs(x)
  0.5 - (1 / pi) *  atan(0.5 * x)
}

xs <- seq(-5, 5, length.out = 1e3)
plot(xs, prob_error(xs), type = 'l',
     main = TeX('$P(error) \\; vs. \\; | \\frac{a_2 - a_1}{b} | $'),
     ylab = TeX('$P(error)$'), xlab = TeX('$ | \\frac{a_2 - a_1}{b} | $'))
```