Assignment-1: Statistical Machine Learning, Winter-2023

Arnav Singh, 2021019

January, 2024

---

1. Consider two Cauchy distributions in one dimension:

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

Assume $P(\omega_1) = P(\omega_2)$. Find the total probability of error. Note you need to first obtain the decision boundary using $p(\omega_1|x) = p(\omega_2|x)$. Then determine the regions where error occurs and use

$$p(\text{error}) = \int_x p(\text{error}|x)p(x)dx$$

Plot the conditional likelihoods, $p(x|\omega_i)p(\omega_i)$, and mark the regions where error will occur. This shall be a rough hand-drawn sketch. As $p(x)$ is the same when equating posteriors, we can simply use $p(x|\omega_i)p(\omega_i)$ [1].

Assumption:- $P(\omega_1) = P(\omega_2)$

Now, to find the decision boundary:-

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

<u>By Bayes' Theorem</u>

$$P(\omega_i|x) \times \frac{P(x)}{P(\omega_i)} = P(x|\omega_i)$$

$$P(\omega_i|x) = \frac{P(x|\omega_i) \times P(\omega_i)}{P(x)}$$

So,

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

$\because$ Assumed that $P(x) \neq 0$, $P(\omega_1) \neq 0$

Given: $P(\omega_1) = P(\omega_2)$

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

$\therefore$ Assumpt[illegible] the de[illegible] are non ze[illegible] Values

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

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

$$\frac{x-a_2}{b} = \frac{x-a_1}{b} \quad \text{or} \quad \frac{x-a_2}{b} = \frac{a_1-x}{b}$$

$a_2 = a_1$

This is impossible, since it's given that $a1 < a2$

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

this serves as the decision boundary for the 2 classes

$$P(\text{error}) = \int_x P(\text{error}|x)\, p(x)\, dx$$

$P(\text{error}|x) = P(\omega_1|x)$ if ω_2 is chosen

$P(\omega_2|x)$ if ω_1 is chosen

$$P(\text{error}) = \int_{-\infty}^{\frac{a1+a2}{2}} \min\left(P(\omega_1|x), P(\omega_2|x)\right) p(x)\,dx + \int_{\frac{a_1+a_2}{2}}^{\infty} \min\left(P(\omega_1|x), P(\omega_2|x)\right)p(x)\,dx$$

$$= \int_{-\infty}^{\frac{a_1+a_2}{2}} P(\omega_2|x)\,p(x)\,dx + \int_{\frac{a_1+a_2}{2}}^{\infty} P(\omega_1|x)\,p(x)\,dx$$

$$= \int_{-\infty}^{\frac{a_1+a_2}{2}} P(x|\omega_2)\,P(\omega_2)\,dx + \int_{\frac{a_1+a_2}{2}}^{\infty} P(x|\omega_1)\,P(\omega_1)\,dx$$

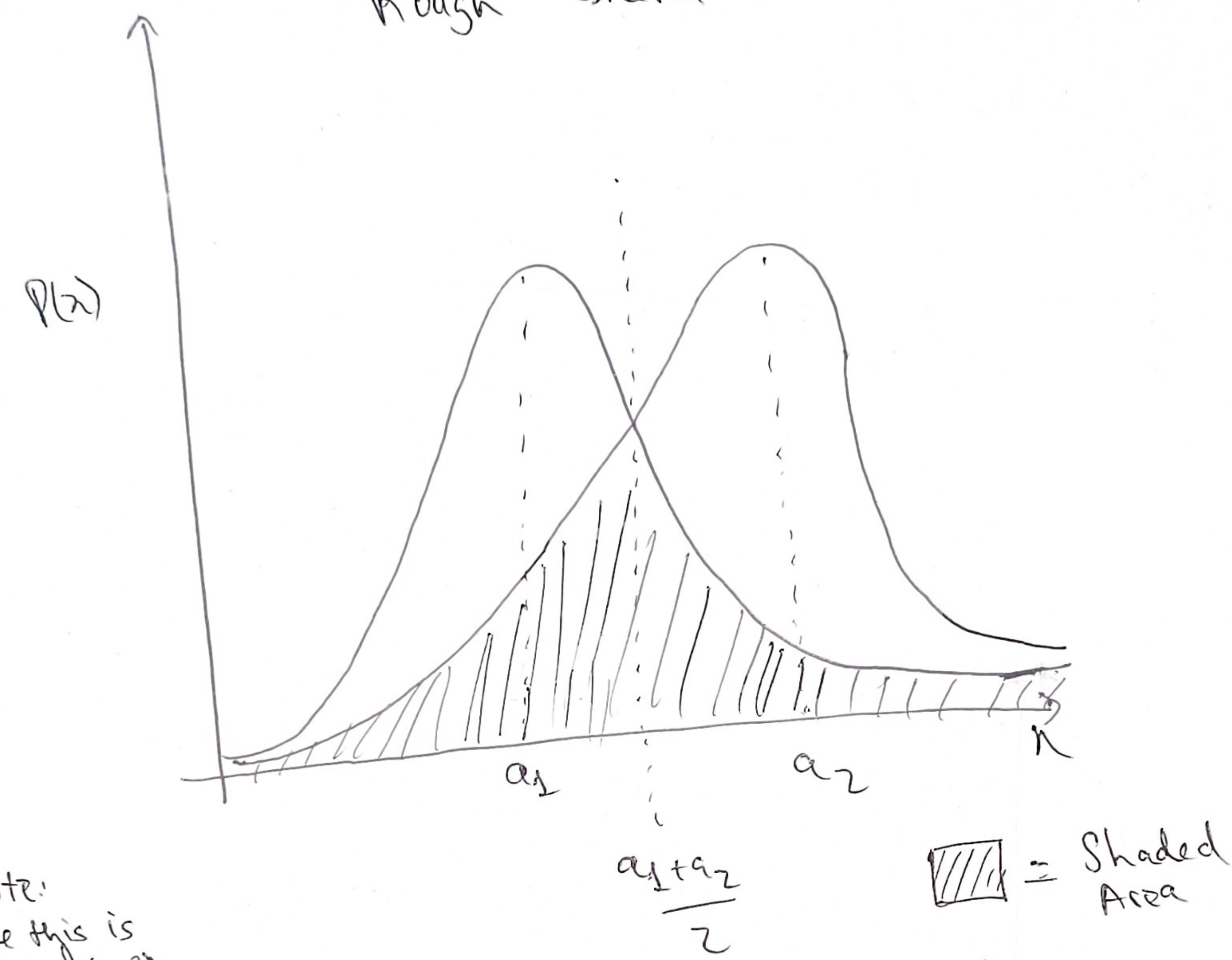
$$P(\text{error}) = P(\omega_1)\int_{-\infty}^{\frac{a_1+a_2}{2}} P(x|\omega_2)\,dx + \int_{\frac{a_1+a_2}{2}}^{\infty} P(x|\omega_1)\,dx$$

Assuming ~~that~~ the prior is independent of x as it always is

So, the region of error would be the area under the "smaller" curve over the domain, this can be represented as:-

ssump
the de
non ze

Rough Sketch



Note:
Since this is a handrawn graph, it might not be an accurate sketch, but the shaded area relative to the curve should be just the saw

The curve with maxima at $x = a_1$ represents $p(x|\omega_1)\,p(\omega_1)$

The curve with maxima at $x = a_2$ represents $p(x|\omega_2)\,p(\omega_2)$

Further simplifying $p(\text{error})$,

$$p(\text{error}) = p(\omega_1) \int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{1}{\pi b\left(1+\left(\frac{x-a_2}{b}\right)^2\right)} + \int_{\frac{a_1+a_2}{2}}^{\infty} \frac{1}{\pi b\left(1+\left(\frac{x-a_1}{b}\right)^2\right)}$$

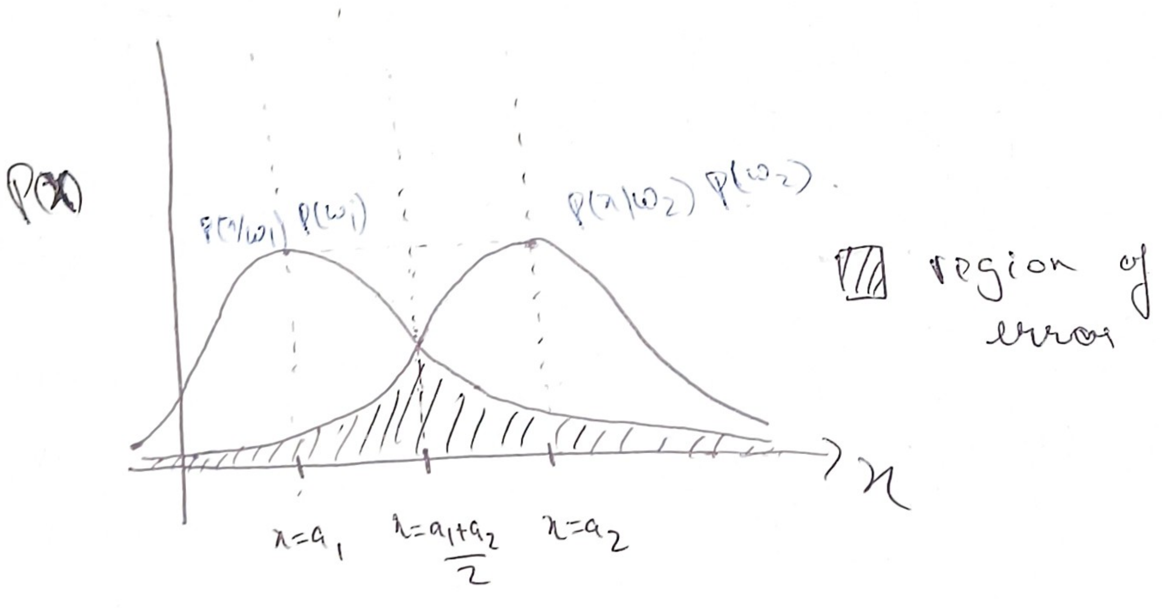
$$p(\text{error}) = p(\omega_1)\cdot 2 \cdot \int_{\frac{a_1+a_2}{2}}^{\infty} \frac{1}{\pi b\left(1+\left(\frac{x-a_1}{b}\right)^2\right)}$$

$$p(\text{error}) = p(\omega_1)\cdot 2 \cdot \left[\frac{\tan^{-1}\left(\frac{x-a_1}{b}\right)}{\pi}\right]_{\frac{a_1+a_2}{2}}^{\infty}$$

$$\Rightarrow p(\omega_1) \cdot 2 \cdot \frac{1}{\pi} \left(\tan^{-1}\infty - \tan^{-1}\left(\frac{a_2-a_1}{2b}\right)\right)$$

$$\Rightarrow p(\omega_1) \cdot 2 \cdot \frac{1}{\pi} \left(\frac{\pi}{2} - \tan^{-1}\left(\frac{a_2-a_1}{2b}\right)\right)$$

$$p(\text{error}) \Rightarrow \boxed{p(\omega_1) - \frac{2p(\omega_1)\tan^{-1}\left(\frac{a_2-a_1}{2b}\right)}{\pi}}$$



another sketch of the curves on the previous page

Note
As mentioned earlier the graphs represent
$p(x|\omega_i)p(\omega_i)$

2. Compute the unbiased covariance matrix: [0.5]

$$X = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

[redacted]
$d = 3$ variables
$N = 3$ observations

Here, $X \in R^{d \times N}$ form.

Assumptions

Each row represents values of a single var for different observation. Also, even if this is not the case if measured columnwise the variables' orders might have changed, but the magnitudes remain the same. Also, I haven't inferred anything about the nature of data from the matrix upon writing $d \times N$ in the way that it's ~~presented~~.

To find: Unbiased Covariance Matrix

	obs 1	obs. 2	obs. 3
x	1	0	0
y	-1	0	1
z	0	1	1

Mean of $x = \bar{x} = \frac{1+0+0}{3} = \frac{1}{3}$

Mean of $y = \bar{y} = \frac{-1+0+1}{3} = 0$

Mean of $z = \bar{z} = \frac{0+1+1}{3} = \frac{2}{3}$

Now, Covariance of X and X $= \frac{\left(1-\frac{1}{3}\right)^2 + \left(0-\frac{1}{3}\right)^2 + \left(0-\frac{1}{3}\right)^2}{3-1} = \frac{\frac{\cancel{6}\,1}{\cancel{9}\,3}}{\cancel{2}} = \frac{1}{3}$

Covariance of Y and Y $= \frac{(-1)^2 + (0)^2 + 1^2}{3-1} = 1$

Covariance of Z and Z $= \frac{\left(0-\frac{2}{3}\right)^2 + \left(1-\frac{2}{3}\right)^2 + \left(1-\frac{2}{3}\right)^2}{3-1} = \frac{\frac{6}{9}}{2} = \frac{1}{3}$

Covariance of X and Y $= \frac{\left(1-\frac{1}{3}\right)(-1) + \left(0-\frac{1}{3}\right)(0) + \left(0-\frac{1}{3}\right)(1)}{3-1}$

$= \frac{-\frac{2}{3} - \frac{1}{3}}{2} = \frac{-1}{2}$

Covariance of Y and Z $= \frac{\left(0-\frac{2}{3}\right)(-1) + \left(1-\frac{2}{3}\right)(0) + (1)\left(1-\frac{2}{3}\right)}{3-1}$

$= \frac{\frac{2}{3}+\frac{1}{3}}{2} = \frac{1}{2}$

$$\text{Covariance of } X \text{ and } Z = \frac{\left(1-\frac{1}{3}\right)\left(0-\frac{2}{3}\right)+\left(0-\frac{1}{3}\right)\left(1-\frac{2}{3}\right)+\left(0-\frac{1}{3}\right)\ldots}{3-1}$$

$$= \frac{-\frac{4}{9}-\frac{1}{9}-\frac{1}{9}}{2} = \frac{-\frac{6}{9}}{2} = \frac{-\frac{2}{3}}{2} = -\frac{1}{3}$$

Now, Covariance Matrix can be calculated by plugging the values for the variables in the matrix below:-

$$\begin{array}{cccc} & x & y & z \\ x & Cov(x,x) & Cov(x,y) & Cov(x,z) \\ y & Cov(y,x) & Cov(y,y) & Cov(y,z) \\ z & Cov(z,x) & Cov(z,y) & Cov(z,z) \end{array}$$

$$\Rightarrow$$

$$\text{Covariance Matrix} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{2} & 1 & \frac{1}{2} \\ -\frac{1}{3} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

3. **a.** In the multi-category case, the probability of error $p(\text{error})$ is given as $1 - p(\text{correct})$, where $p(\text{correct})$ is the probability of being correct. Consider a case of 3 classes or categories. Draw a rough sketch of $p(x|\omega_i)p(\omega_i)$ for all $i = 1, 2, 3$. Give an expression for $p(\text{error})$. Assume equi-probable priors for simplicity. [1]

b. Mark the regions if the three conditional likelihoods are Gaussians $p(x|\omega_i)$, $N(\mu_i, 1)$. $\mu_1 = -1, \mu_2 = 0, \mu_3 = 1$. Find $p(\text{error})$ in terms of the CDF of the standard distribution. [1]

(a) $p(\text{error}) = 1 - p(\text{correct})$

Classes: $\omega_1, \omega_2, \omega_3$

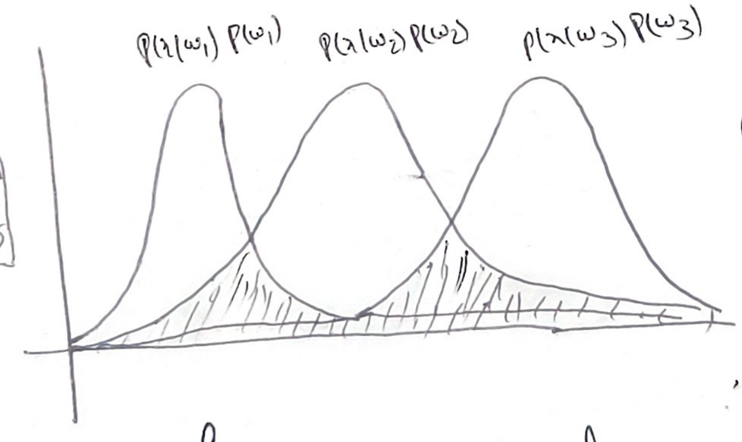
$p(\text{correct}) \Rightarrow$ the prob. that ~~taken~~ the class with the highest posterior probability for a given x is chosen

Equi-probable priors $\Rightarrow P(\omega_1) = P(\omega_2) = P(\omega_3)$

Now, we're supposed to draw rough sketches of $p(x|\omega_i)\,p(\omega_i)\ \forall\, i = 1, 2, 3$. Since the question doesn't explicitly mention the distribution for $p(x|\omega_i)$, I'm going to assume it to be Cauchy, and for them to have equal widths b, and as in the following order $a_1 < a_2 < a_3$

Consequently, this is what it (a combined plot) should look like with the aforementioned assumptions

Considering Cauchy Distributions



= Region of error

Now, $p(\text{correct}) = \int_x p(\text{correct}|x)\, p(x)\, dx$, replacing error by correct in the eqn in Q-1

~~$p(\text{correct}|x) = \int_x \max(P(\omega_1|x), P(\omega_2|x), P(\omega_3|x))\, P(x)\, dx$~~

~~$p(\text{correct}) = \int_x p(x|\text{correct})\, P_1$~~

$$P(\text{correct}) = \int_x \max\left(P(\omega_1|x), P(\omega_2|x), P(\omega_3|x)\right) p(x)\, dx$$

$$P(\text{error}) = 1 - \int_x \max\left(P(\omega_1|x), P(\omega_2|x), P(\omega_3|x)\right) p(x)\, dx$$

By bayes' theorem, $P(\omega|x)\, p(x) = p(x|\omega)\, P(\omega)$

$$P(\text{error}) = 1 - \int_x \max\left(p(x|\omega_1), p(x|\omega_2), p(x|\omega_3)\right) \cdot P(\omega_1)$$

Assuming that $i = 1,2,3$ are the only categories and given that $P(\omega_1) = P(\omega_2) = P(\omega_3)$, and are constant

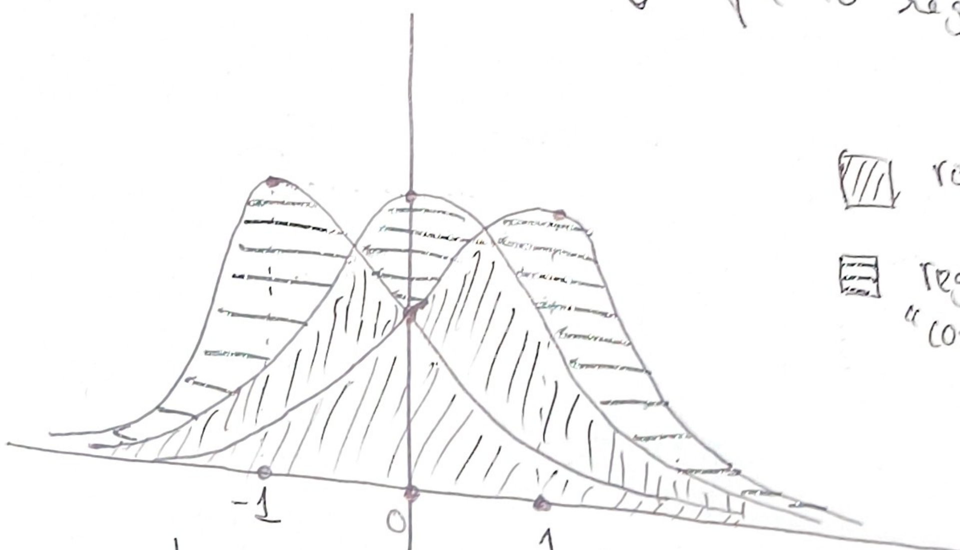
$3 \cdot P(\omega_1) = 1$; $P(\omega_1) = \frac{1}{3}$

Consequently,
$$P(\text{error}) = 1 - \frac{\int_x \max\left(p(x|\omega_1), p(x|\omega_2), p(x|\omega_3)\right)}{3}$$

(b) Now, we're given that the conditional likelihoods are gaussians $p(x|\omega_i)$, $N(\mu_i, 1)$, $\mu_1 = -1$, $\mu_2 = 0$, and $\mu_3 = 1$

ostensibly, the "regions", refer to regions of error.

$F_{X|\omega_i}(x)$ cdf for $P(X|\omega_i)$ at x gaussian distribution



region of error

region of "correctness"

Now,

$$P(\text{correct}) = \int_{-\infty}^{-\frac{1}{2}} p(x|\omega_1) P(\omega_1)\, dx + \int_{-\frac{1}{2}}^{\frac{1}{2}} p(x|\omega_2) P(\omega_2)\, dx + \int_{\frac{1}{2}}^{\infty} p(x|\omega_3) P(\omega_3)\, dx$$

$$P(\text{correct}) = P(\omega_1)\left[F_{X|\omega_1}\left(-\tfrac{1}{2}\right) + F_{X|\omega_2}\left(\tfrac{1}{2}\right) - F_{X|\omega_2}\left(-\tfrac{1}{2}\right) + 1 - F_{X|\omega_3}\left(\tfrac{1}{2}\right)\right]$$

Q2) i)

$p(\text{error}) = 1 - p(\omega_1)\left[F_{X|\omega_1}\left(\frac{-1}{2}\right) + F_{X|\omega_2}\left(\frac{1}{2}\right) - F_{X|\omega_2}\left(\frac{-1}{2}\right) + \underbrace{1 - F_{X|\omega_3}\left(\frac{1}{2}\right)}_{\text{Total area under curve} - \text{area till } x = \frac{1}{2}}\right]$

∵ F represents the cdf

$p(\omega_1) = \frac{1}{3}$ ∵ from last part

Formula for cdf of the standard normal distribution is

:- $\cancel{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-u^2}{2}}\,dx}$ $\int_{-\infty}^{x=X} \text{pdf}$

$$\text{pdf} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$\sigma = 1$, μ depends on the class

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

Class $i=1$, $\text{pdf} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+1)^2}$, $\text{cdf} = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x+1}{\sqrt{2}}\right)\right)$

" $i=2$, $\text{pdf} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $\text{cdf} = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$

" $i=3$, $\text{pdf} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2}$, $\text{cdf} = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x-1}{\sqrt{2}}\right)\right)$

$$p(\text{error}) = 1 - \frac{p(\omega_1)}{2}\left[1 + \text{erf}\left(\frac{1}{2\sqrt{2}}\right) + 1 + \text{erf}\left(\frac{1}{2\sqrt{2}}\right) - \text{erf}\left(\frac{-1}{2\sqrt{2}}\right) - \cancel{1 + 1} - \text{erf}\left(\frac{-1}{2\sqrt{2}}\right)\right]$$

$$= 1 - \frac{p(\omega_1)}{2}\left[2 + 2\left(\text{erf}\left(\frac{1}{2\sqrt{2}}\right) - \text{erf}\left(\frac{-1}{2\sqrt{2}}\right)\right)\right]$$

∵ Since erf is an odd function

$$= 1 - \frac{p(\omega_1)}{2}\left(2 + 4\,\text{erf}\left(\frac{1}{2\sqrt{2}}\right)\right)$$

$$= 1 - p(\omega_1)\left(1 + 2\,\text{erf}\left(\frac{1}{2\sqrt{2}}\right)\right)$$

Acc. to the scipy function 'erf', the value of $\text{erf}\left(\frac{1}{2\sqrt{2}}\right)$

$= \sim 0.382$, so

$$p(\text{error}) = 1 - p(\omega_1)\left[\cancel{}\ 1.764\right]$$

Since, $p(\omega_1) = \frac{1}{3}$ [part (a)

$$p(\text{error}) = 1 - \frac{1}{3}(\cancel{}\ 1.764)$$

$$= \cancel{}\ \boxed{\text{approx } 0.412}$$

also, p(error) in terms of the cdf function is

#
$$\boxed{p(\text{error}) = 1 - \frac{1}{3}\left(cdf_{x|\omega_1}\left(\frac{-1}{2}\right) + cdf_{x|\omega_2}\left(\frac{1}{2}\right) - cdf_{x|\omega_2}\left(\frac{-1}{2}\right) + 1 - cdf_{x|\omega_3}\left(\frac{1}{2}\right)\right)}$$

4. Find the likelihood ratio test for the following Cauchy pdf:

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

Assume $P(\omega_1) = P(\omega_2)$ and use 0-1 loss. [1]

According to Pattern Classification by Duda, Hart and Stork

$$R(\alpha_i | x) = \sum_{J=1}^{c} \lambda(\alpha_i | \omega_J) \, P(\omega_J | x)$$

conditional Risk of decision $\alpha_i = R(\alpha_i|x)$

Loss function for 0-1 loss :-

$$\lambda(\alpha_i | \omega_J) = \begin{cases} 0 & i = J \\ 1 & i \neq J \end{cases} \quad i, J = 1, \ldots, c$$

For simplicity $\lambda_{iJ} = \lambda(\alpha_i | \omega_J)$

$$R(\alpha_1 | x) = \lambda_{11} P(\omega_1 | x) + \lambda_{12} P(\omega_2 | x)$$
$$R(\alpha_2 | x) = \lambda_{21} P(\omega_1 | x) + \lambda_{22} P(\omega_2 | x)$$

if $R(\alpha_2|x) > R(\alpha_1|x)$, it's advantageous to choose $i = 1$

$$\lambda_{11} P(\omega_1 | x) + \lambda_{12} P(\omega_2 | x) < \lambda_{21} P(\omega_1|x) + \lambda_{22} P(\omega_2|x)$$

$$(\lambda_{12} - \lambda_{22})(P(\omega_2|x)) < (\lambda_{21} - \lambda_{11})(P(\omega_1|x))$$

Or, using bayes' theorem

$$\frac{(\lambda_{12} - \lambda_{22}) \cdot P(x|\omega_2) \cdot \cancel{P(\omega_2)}}{\cancel{P(x)}} < \frac{(\lambda_{21} - \lambda_{11}) \cdot P(x|\omega_1) \cdot \cancel{P(\omega_1)}}{\cancel{P(x)}}$$

$\because$ $P(\omega_1) = P(\omega_2)$
and $P(x) = P(x)$

$$P(x|\omega_1)(\lambda_{21}-\lambda_{11}) > P(x|\omega_2)(\lambda_{12}-\lambda_{22})$$

By 0-1 loss, $\lambda_{12}=1, \lambda_{21}=1, \lambda_{11}=0, \lambda_{22}=0,$

On inputting these values :-

$$P(x|\omega_1)(1-0) > P(x|\omega_2)(1-0)$$

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

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

$$\cancel{1}+\left(\frac{x-a_2}{b}\right)^2 > \cancel{1}+\left(\frac{x-a_1}{b}\right)^2$$

$$\frac{\cancel{x^2}-2a_2x+a_2^2}{\cancel{b^2}} > \frac{\cancel{x^2}-2a_1x+a_1^2}{\cancel{b^2}}$$

$$a_2^2-a_1^2 > 2a_2x-2a_1x$$

Given: $a_2 > a_1$

$$(a_2+a_1)\frac{\cancel{a_2^2-a_1^2}}{2\cancel{(a_2-a_1)}} > x$$

Also, $x=\frac{a_1+a_2}{2}$ is the decision boundary

could also be inferred if likelihood ratio $=1$

So, if $x < \left(\frac{a_2+a_1}{2}\right)$, it's 'advantageous' to choose ω_1,
and if $x > \left(\frac{a_2+a_1}{2}\right)$, it's 'advantageous' to choose ω_2

Also, the likelihood ratio is $\frac{P(x|\omega_1)}{P(x|\omega_2)}$, which is equal to

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

if $x > \frac{a_2+a_1}{2}$ i.e., likelihood ratio < 1, it is advantageous to choose ω_2

and if $x < \frac{a_2+a_1}{2}$ i.e. likelihood ratio > 1, it is advantageous to choose ω_1

Note: By 'advantage' or 'advantageous', it's implied that the strategy or choice situation is less likely to incur a loss.

it's in distinct