ML_HW1

Mohammad Javad Pesarakloo
810100103

March 20, 2024

Question 1

A

$$P(w_1|x) = p(w_2|x)$$
$$\Rightarrow \frac{1}{\pi b}\frac{1}{1+(\frac{x-a_1}{b})^2} = \frac{1}{\pi b}\frac{1}{1+(\frac{x-a_2}{b})^2}$$
$$\Rightarrow (\frac{x-a_1}{b})^2 = (\frac{x-a_2}{b})^2$$

condition one:

$$x - a_1 = x - a_2$$

Which is contradiction and the other condition:

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

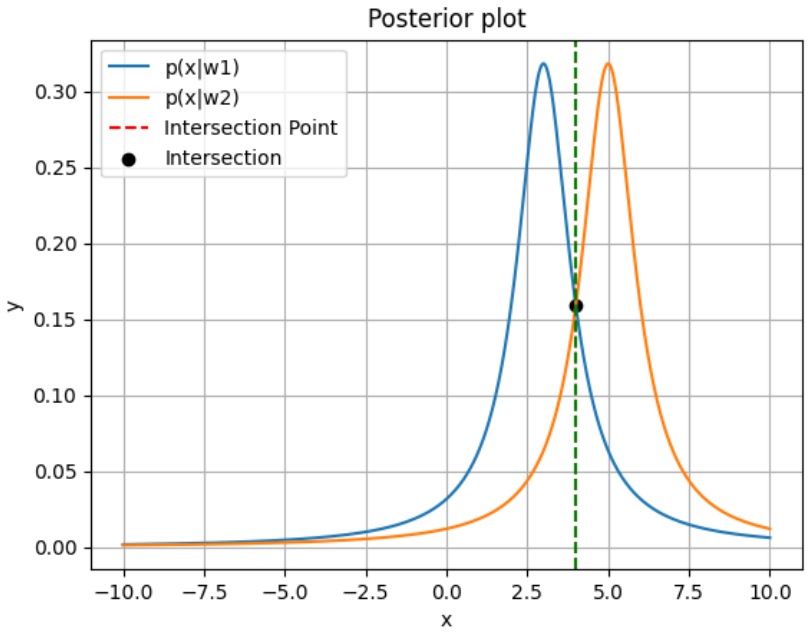

Figure 1: posterior plot

B

Minimum error occurs when we use bayes classifier.As the following figure suggests, probabality of error is surface of the red region:

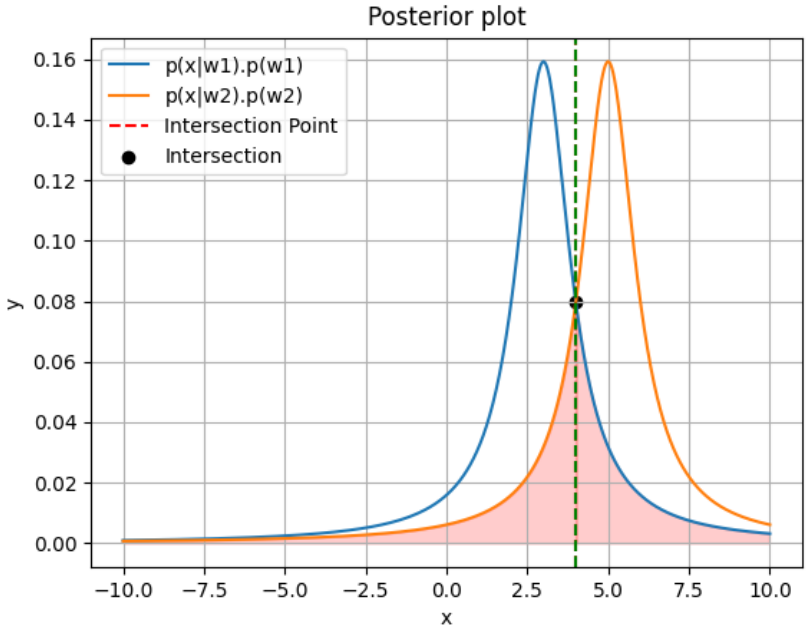

Figure 2: error

To calculate this surface, we have:

$$\frac{1}{2}\left[\int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{1}{\pi b}\frac{1}{1+(\frac{x-a_1}{b})^2}dx + \int_{\frac{a_1+a_2}{2}}^{\infty} \frac{1}{\pi b}\frac{1}{1+(\frac{x-a_2}{b})^2}dx\right]$$

$$= \frac{1}{2\pi b}\left[\tan^{-1}(\frac{a_1-a_2}{2b}) + \frac{b\pi}{2} + \frac{b\pi}{2} - \tan^{-1}(\frac{a_2-a_1}{2b})\right]$$

$$\xrightarrow{a_2>a_1} P(error)$$

$$= \frac{1}{2\pi b}\left[b\pi - 2\tan^{-1}|\frac{a_2-a_1}{2b}|\right]$$

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

C

Maximum value of error occurs when the value of $\tan^{-1}$ function has its minimum value which is 0.This condition occurs under two circumstances:

- $b \to \infty$: under this circumstance, distribution of both classes are zero and it is not wise for a classifier to choose either of them because both of them have zero probabality.
- $a_1 = a_2$: under this circumstance, distribution of both classes are totally identical and classifier is acting totally random and as the relation of the previous section suggests, random classifier has accuracy and error rate of 50%.

D

As calculated in section A and B, the decision boundary occurs at point $\frac{a_1+a_2}{2}$ and probabality of error is :

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

In the following figure, yellow area is decision area of class a_1 and blue area is decision area of class a_2.

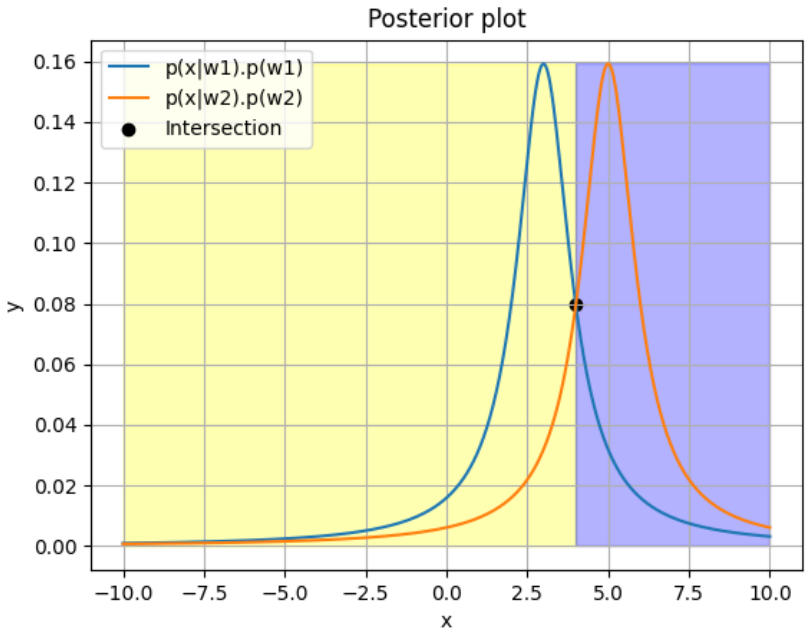

Figure 3: decision area

E

To minimize risk, the classifier has to meet the following condition:

$$R(w_1|X) \overset{2}{>} R(w_2|X)$$

or:

$$R(w_1|X) \overset{1}{<} R(w_2|X)$$

As values on the diagonal of risk matrix are zero, we have the following relations:

$$R(w_1|X) = \lambda_{21} P(w_2|X)$$
$$R(w_2|X) = \lambda_{12} p(w_1|X)$$

As perior probabalities are equal, to find the decision boundary, we have to solve the following equation:

$$\frac{2}{1 + (\frac{x-a_2}{b})^2} = \frac{1}{1 + (\frac{x-a_1}{b})^2}$$
$$Delta = \sqrt{2a_2^2 - 4a_2a_1 + 2a_1^2 - b^2}$$
$$\Rightarrow x = -a_2 + 2a_1 \overset{+}{-} Delta$$

The following figure shows the decision boundary and area of error which can be calculated using integration:

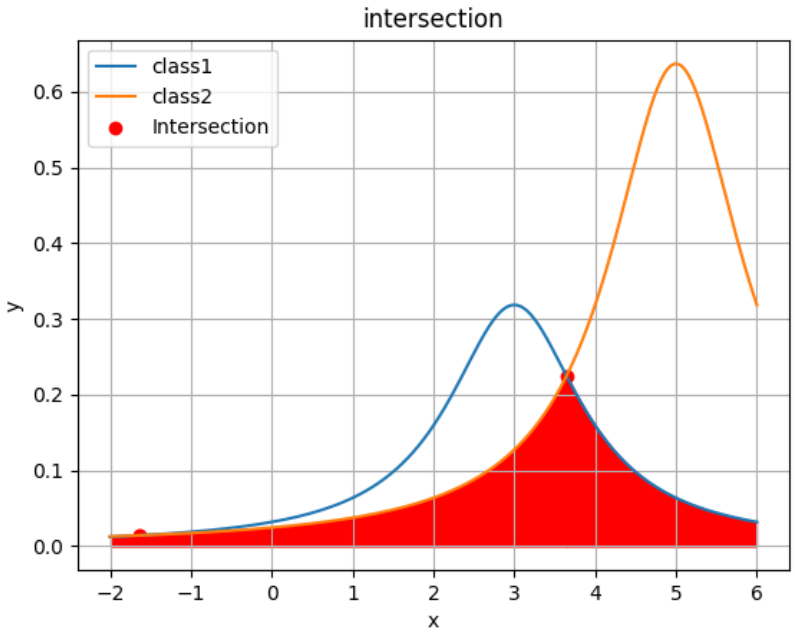

Figure 4: decision area minimum risk

Question 2

A

$$R(w_1|X) \overset{1}{<} R(w_2|X)$$
$$R(w_1|X) = \lambda_{11}p(w_1|X) + \lambda_{12}p(w_2|X)$$
$$= \lambda_{11}\frac{p(X|w_1)p(w_1)}{p(X)} + \lambda_{12}\frac{p(X|w_2)p(w_2)}{p(X)}$$

and similarly:

$$R(w_2|X) = \lambda_{21}\frac{p(X|w_1)p(w_1)}{p(X)} + \lambda_{22}\frac{p(X|w_2)p(w_2)}{p(X)}$$

To find the decision boundary, we intersect these two risks:

$$R(w_1|X) = R(w_2|X)$$
$$\Rightarrow \frac{p(X|w_1)}{p(X|w_1)} \overset{1}{>} \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}\frac{p(w_2)}{p(w_1)}$$

B

$$\frac{p(X|w_1)}{p(X|w_2)} \overset{1}{>} \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}\frac{p(w_2)}{p(w_1)}$$
$$\Rightarrow \frac{p(X|w_2)}{p(X|w_1)} < \frac{\lambda_{21} - \lambda_{11}}{\lambda_{12} - \lambda_{22}}\frac{p(w_1)}{p(w_2)}$$

Question 3

To find the decision boundary, we have to intersect posterior probabalities:

$$p(w_1|x) = p(w_2|x)$$
$$\Rightarrow \frac{p(x|w_1)p(w_1)}{p(x)} = \frac{p(x|w_2)p(w_2)}{p(x)}$$
$$\xrightarrow{p(w_1)=p(w_2)} p(x|w_1) = p(x|w_2)$$
$$\Rightarrow \frac{x}{\sigma_1^2}exp(-\frac{x^2}{2\sigma_1^2}) = \frac{x}{\sigma_2^2}exp(-\frac{x^2}{2\sigma_2^2})$$

The above equation suggests that for $x \leq 0$, two distributions are identical and the classifier acts randomly.for $x > 0$ we have:

$$\frac{\sigma_2^2}{\sigma_1^2} = exp(-\frac{x^2}{2}(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2})) \xrightarrow[x>0]{\ln}$$
$$x = \sqrt{2\ln\frac{\sigma_2^2}{\sigma_1^2}(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}))}$$

Question 4

Suppose $p(x|w_1) = \mathcal{N}(\mu_1, \sigma^2)$ and $p(x|w_2) = \mathcal{N}(\mu_2, \sigma^2)$.At first we find decision boundary by intersecting posterior probabalities:

$$p(x|w_1)p(w_1) = p(x|w_2)p(w_2)$$

$$\frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma^2})p(w_1) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2}\frac{(x-\mu_2)^2}{\sigma^2})p(w_2)$$

$$\frac{p(w_1)}{p(w_2)} = exp(-\frac{1}{2\sigma^2}((x-\mu_1)^2 - (x-\mu_2)^2) \xrightarrow{\ln}$$

$$x = \frac{-2\sigma^2 \ln\frac{p(w_1)}{p(w_2)} + \mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)}$$

without loosing any generalization, we can assume that $\mu_2 > \mu_1$.For x to not lie between μ_1 and μ_2, we have the following relation:

$$|\frac{-2\sigma^2 \ln\frac{p(w_1)}{p(w_2)} + \mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} - \frac{\mu_1 + \mu_2}{2}| < \frac{\mu_2 - \mu_1}{2}$$

Question 5

A

For class of red stars, we have:

$$mean = \begin{pmatrix} 1.33 \\ 1.61 \end{pmatrix} cov = \begin{pmatrix} 0.55 & 0.185 \\ 0.185 & 0.987 \end{pmatrix}$$

And for class of black dots, we have:

$$mean = \begin{pmatrix} -0.15 \\ -0.15 \end{pmatrix} cov = \begin{pmatrix} 1.5525 & 0.0025 \\ 0.0025 & 0.5025 \end{pmatrix}$$

B

If we use the following relation to find decision boundary:

$$\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu))$$

we get the following decision boundary:

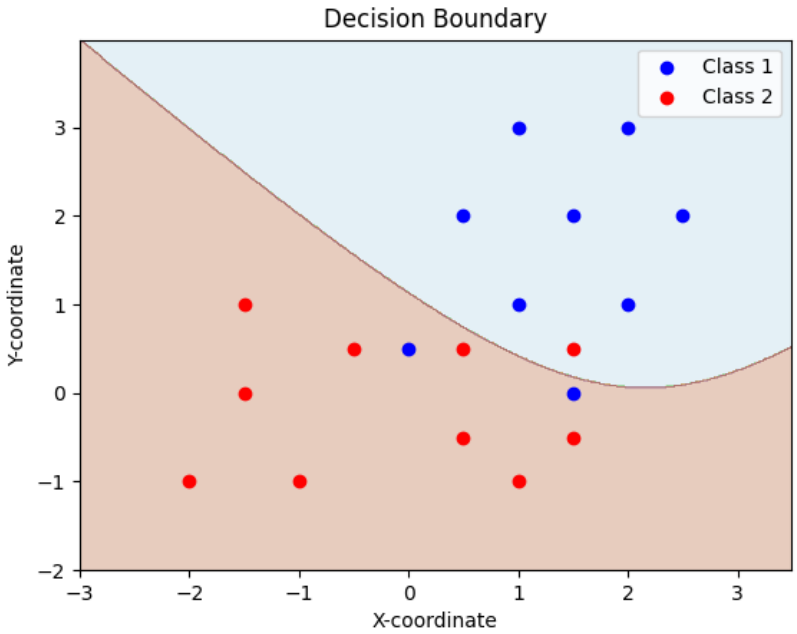

Figure 5: decision boundary

from which we can obtain that 15.7% of data is classified wrong.

B

To minimize risk, we use minimum risk classifier and as the risk matrix suggests, we would like to classify class 1 less wrong than class 2.So the decsision boundary shifts down and changes as follows:

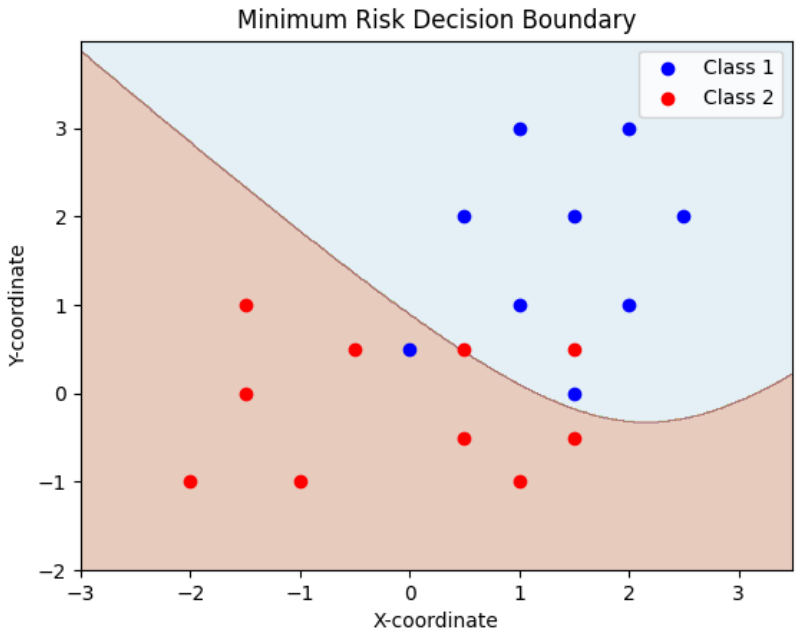

Figure 6: decision boundary in minimum risk

C

If we adjust perior probabalites, we expect to increase the decision area of class 1 which the following figure yields:

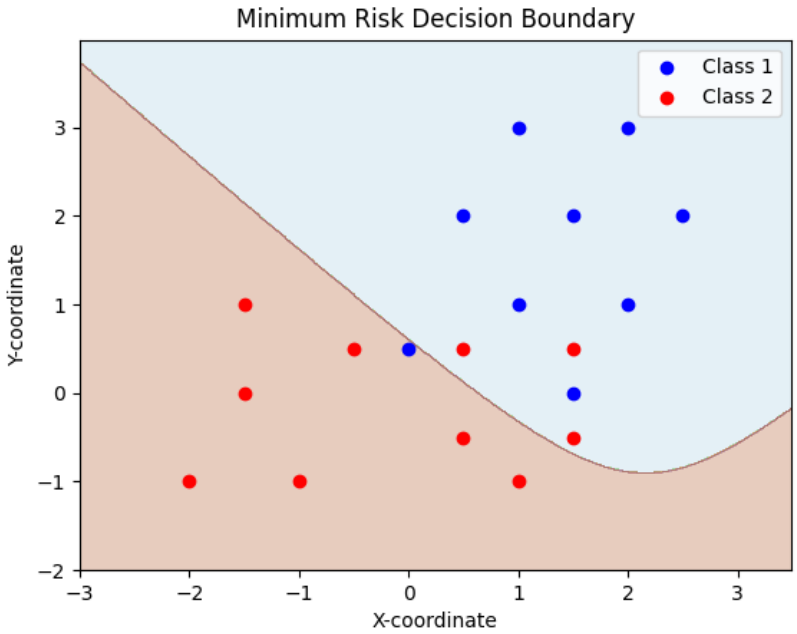

Figure 7: decision boundary in minimum risk, adjusted periors

Question 6

$$p(X) = \frac{\lambda^X e^{-\lambda}}{X!}$$

A

$$p(x_1, x_2, ..., x_n|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \xrightarrow{\log}$$

$$logLikelihood = \sum_{i=1}^{n} \log \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$= \sum_{i=1}^{n} x_i \log \lambda - \lambda - \log x! \xrightarrow{\frac{d}{d\lambda}=0}$$

$$-n + \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \lambda = \frac{\sum_{i=1}^{n} x_i}{n}$$

B

$$p(\lambda|D) \propto p(D|\lambda)p(\lambda) = p(x_1, ..., x_n|\lambda)p(\lambda)$$

$$c\lambda^{\alpha-1} e^{-\beta\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$= \underbrace{c \prod_{i=1}^{n} \frac{1}{x_i}}_{c'} \lambda^{\alpha-1+\sum_{i=1}^{n} x_i} e^{-\lambda(n+\beta)}$$

$$= Gamma(\lambda|\alpha + \sum_{i=1}^{n} x_i, n + \beta)$$

C

Yes it is.Because both the perior and posterior are Gamma

D

$$\lambda_{MAP} = \frac{\alpha_{new} - 1}{\beta_{new}} = \frac{\alpha - 1 + \sum_{i=1}^{n} x_i}{n + \beta}$$

E

Yes:

$$\lim_{n\to\infty} \frac{\alpha - 1 + \sum_{i=1}^{n} x_i}{n + \beta} = \frac{\sum_{i=1}^{n} x_i}{n} = \lambda_{MLE}$$

F

When number of samples is very large, our sample is a good representation of data.So MLE is an easier and better approach.But when number of samples is low, we have to be more cautious and it is better to announce a distribution instead of a single number for λ