HW1 Solution, CSE 569 Fall 2018

Name: Kunal Vinay Kumar Suthar
ASUrite ID: 1215112535

- ## Question 1:

Q1 → SOLUTION :

- Given :

(1.) It is a two-class one dimensional problem.

(2.) PDFs $p(x|\omega_1)$ and $p(x|\omega_2)$ are Gaussians $N(0,\sigma^2)$ and $N(1,\sigma^2)$ respectively.

(3.) Assumption : $\lambda_{11} = \lambda_{22} = 0$

- To show : The threshold x_0 minimizing the average risk is equal to $x_0 = \frac{1}{2} - \sigma^2 \ln \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}$

We know that average loss (or risk) for taking action α_i is :

$$R = \int R(\alpha_i|x)\, p(x)\, dx$$

Let R_1 denote the feature space where classifier decides ω_1 & likewise for R_2 & ω_2. Then Expanding R for the two classes we get :

$$R = \int_{R_1} (\lambda_{11} P(\omega_1) p(x|\omega_1) + \lambda_{12} P(\omega_2) p(x|\omega_2)] \, dx + \int_{R_2} (\lambda_{21} P(\omega_1) p(x|\omega_1) + \lambda_{22} P(\omega_2) p(x|\omega_2)) \, dx$$

By Differentiating the above with respect to x and equating it to 0, we get the threshold x_0 for minimizing the average risk

$$\Rightarrow \left.\frac{dR}{dx}\right|_{x=x_0} = 0 \Rightarrow \lambda_{11} P(\omega_1) p(x_0|\omega_1) + \lambda_{12} P(\omega_2) p(x_0|\omega_2) - \lambda_{21} P(\omega_1) p(x_0|\omega_1) + \lambda_{22} P(\omega_2) p(x_0|\omega_2) = 0$$

Putting $\lambda_{11} = \lambda_{22} = 0$, we get :

$$\cancel{\lambda_{12} P(\omega_2) p(x_0|\omega_2) - \lambda_{21} p(\omega_1) p(x_0|\omega_1) = 0}$$

$$\Rightarrow \lambda_{12} P(\omega_2) P(x_0|\omega_2) - \lambda_{21} P(\omega_1) P(x_0|\omega_1) = 0$$

→ Rearranging terms, we get :

$$\Rightarrow \frac{p(x_0|\omega_1)}{P(x_0|\omega_2)} = \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}$$

$$\Rightarrow \frac{\cancel{\frac{1}{\sqrt{2\pi}\sigma}} e^{-\frac{1}{2}\frac{(x_0-\mu_1)^2}{\sigma^2}}}{\cancel{\frac{1}{\sqrt{2\pi}\sigma}} e^{-1/2\frac{(x_0-\mu_2)^2}{\sigma^2}}} = \frac{\lambda_{12}}{\lambda_{21}} \frac{P(\omega_2)}{P(\omega_1)}$$

$\Rightarrow$ Putting $\mu_1 = 0$ and $\mu_2 = 1$, we get

$$\Rightarrow \frac{\left(e^{-\frac{1}{2}\frac{x_0^2}{\sigma^2}}\right)}{e^{-1/2\frac{(x_0-1)^2}{\sigma^2}}} = \frac{\lambda_{12}}{\lambda_{21}} \frac{P(\omega_2)}{P(\omega_1)}$$

$$\Rightarrow e^{-\frac{1}{2}\left(\frac{x_0^2-(x_0-1)^2}{\sigma^2}\right)} = \frac{\lambda_{12}}{\lambda_{21}} \frac{P(\omega_2)}{P(\omega_1)}$$

$$\Rightarrow e^{-\frac{1}{2}\left(\frac{x_0^2-x_0^2+2x_0-1}{\sigma^2}\right)} = \frac{\lambda_{12}}{\lambda_{21}} \frac{P(\omega_2)}{P(\omega_1)}$$

$\Rightarrow$ taking natural logarithm on both sides, we get

$$\Rightarrow \frac{+1}{2} \frac{(2x_0-1)}{\sigma^2} = -\ln \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}$$

$$\Rightarrow 2x_0 - 1 = -2\sigma^2 \ln \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}$$

$$\Rightarrow x_0 = \frac{1}{2} - \sigma^2 \ln\left(\frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}\right)$$

Hence proved

Question 2:

1) **Part 1 Solution:**

Page

PROBLEM 2: SOLUTION

GIVEN:

(i) There are c classes. Therefore $\sum_{i=1}^{c} P(\omega_i|x) = 1$.

(ii) $P(\omega_{max}|x) \geq P(\omega_i|x)$

(1) TO PROVE: $P(\omega_{max}|x) \geq 1/c$

PROOF: We know that: $P(\omega_{max}|x) \geq P(\omega_i|x)$ (from (ii) in given)

Now applying summation operator from i=1 to c, on the above equation, we get:

$\Rightarrow \sum_{i=1}^{c} P(\omega_{max}|x) \geq \sum_{i=1}^{c} P(\omega_i|x)$

$\Rightarrow \sum_{i=1}^{c} P(\omega_{max}|x) \geq 1$ (from (i) in given)

$\Rightarrow c\,P(\omega_{max}|x) \geq 1$

$\Rightarrow P(\omega_{max}|x) \geq 1/c$

Hence proved.

Question 2:

2) Part 2 Solution:

(2.) TO PROVE: For the minimum-error rate decision rule, the average probability of error is given by:

$$P(error) = 1 - \int P(\omega_{max}|x)\,p(x)\,dx$$

Total average probability of error over all x's is given by: (in space R)

$$\Rightarrow \quad P(error) = \int_R P(error|x)\,p(x)\,dx \quad \text{—— equation } (1)$$

Now, $P(error|x)$ = Probability sum of the posteriors for which the observation did not belong to the correct class =

$$\Rightarrow P(error|x) = 1 - P(\omega_{max}|x) = (1 - (\text{probability for belonging to correct class}))$$

Substituting value of $P(error|x)$ in equation (1) we get:

$$P(error) = \int_R (1 - P(\omega_{max}|x))\,p(x)\,dx$$

$$= \int_R p(x)\,dx - \int P(\omega_{max}|x)\,p(x)\,dx$$

$$\Rightarrow \quad 1 - \int P(\omega_{max}|x)\,p(x)\,dx$$

Question 2:

3) Part 3 Solution

(2.3) SOLUTION:

TO PROVE: $P(error) \leq (c-1)/c$

from part-1:

$$P(\omega_{max}|x) \geq 1/c$$

multiplying the above inequality with -1 we get:

$$\Rightarrow \quad -P(\omega_{max}|x) \leq -1/c$$

~~integrating the~~ multiplying the above with $p(x)$, we get

$$\Rightarrow \quad -p(\omega_{max}|x)\,p(x) \leq -\frac{p(x)}{c}$$

integrating the above with respect to x on space R, we get:

$$\Rightarrow \quad -\int_R p(\omega_{max}|x)\,p(x)\,dx \leq -\frac{1}{c}\int p(x)\,dx$$

we know that $\int p(x)\,dx = 1$

$\Rightarrow$ Adding 1 to the inequality, we get:

$$1 - \int_R p(\omega_{max}|x)\,p(x)\,dx \leq 1 - 1/c$$

$$\Rightarrow \quad 1 - \int_R p(\omega_{max}|x)\,p(x)\,dx \leq \frac{c-1}{c}$$

$$\Rightarrow \quad \boxed{P(error) \leq \frac{c-1}{c}}$$ Hence proved

Question 2:

4) Part 4 Solution

(2.4) A situation for which $P(error) = (c-1)/c$, would be when

$P(\omega_1|x) = P(\omega_2|x) = P(\omega_3|x) = \ldots\ldots = P(\omega_c|x) = P(\omega_{max}|x) = 1/c$

i.e. ~~given x~~, all the posteriors have equal probability.

In that case, according to part(2):

$$P(error) = 1 - \int_R P(\omega_{max}|x)p(x)dx$$

$$\Rightarrow P(error) = 1 - \frac{1}{c}\int_R p(x)dx$$

$$\Rightarrow P(error) = 1 - \frac{1}{c} = \frac{(c-1)}{c}.$$

Question 3:

1) Part 1 Solution

(Q3) GIVEN:

(i) two-class two dimensional classification task.

(ii) class-conditional pdfs:

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}(x-\mu_1)^T(x-\mu_1)\right)$$

$$p(x|\omega_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2\sigma_2^2}(x-\mu_2)^T(x-\mu_2)\right)$$

$\mu_1 = [1,1]^T$, $\mu_2 = [1.5, 1.5]^T$, $\sigma_1^2 = \sigma_2^2 = 0.2 = \sigma^2$

(iii) $P(\omega_1) = P(\omega_2)$

3.1 SOLUTION: Design a Bayesian classifier that minimizes the error probability.

We know that the minimum error-rate classification can be achieved by the discriminant function:

$$g_i(x) = P(\omega_i|x)$$

$$\Rightarrow \quad g_i(x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}$$

As $p(x)$ will be the same for all i's, we treat it as a constant and ignore it.

$\Rightarrow \quad g_i(x) = p(x|\omega_i)\,p(\omega_i)$

taking natural logarithm, we get: $g_i(x) = \cancel{p(x|\omega_i)\,p(\omega_i)}\ \ln p(x|\omega_i) + \ln P(\omega_i)$

$g_1(x) = \ln P(x|\omega_1) + \ln P(\omega_1)$, $g_2(x) = \ln(P(x|\omega_2)) + \ln P(\omega_2)$

Now, for a 2-class problem: $g(x) = g_1(x) - g_2(x)$

$\Rightarrow g(x) = \ln P(x|\omega_1) - \ln P(x|\omega_2)$ ....{as (iii) is given $\ln P(\omega_1)$, $\ln(P(\omega_2))$ cancel each other)

$$\Rightarrow g(x) = \ln \frac{P(x|\omega_1)}{P(x|\omega_2)}$$

$$\Rightarrow \ln \exp\left(-\frac{1}{2\sigma^2}\left[(x-\mu_1)^T(x-\mu_1) - (x-\mu_2)^T(x-\mu_2)\right]\right) \quad .... \text{(ii) } \sigma^2 = \sigma_1^2 = \sigma_2^2 = 0.2$$

$\Rightarrow$ let x be $[x_1, x_2]^T$

$$\Rightarrow \ln \exp\left(\frac{-1}{0.4}\left[[x_1-1 \quad x_2-1.0]\begin{bmatrix} x_1-1 \\ x_2-1.0 \end{bmatrix} - [x_1-1.5 \quad x_2-1.5]\begin{bmatrix} x_1-1.5 \\ x_2-1.5 \end{bmatrix}\right]\right)$$

Question 3:

1) **Part 1 Solution Continued:**

$$\Rightarrow \ln\left(\exp\left(\frac{-1}{0.4}\left((x_1-1)^2+(x_2-1)^2-(x_1-1.5)^2-(x_2-1.5)^2\right)\right)\right)$$

$$\Rightarrow \frac{5}{2}\left[(x_1-1.5)^2+(x_2-1.5)^2-(x_1-1)^2-(x_2-1)^2\right]$$

$$\Rightarrow \frac{5}{2}\left[\cancel{x_1^2}+2.25-3x_1+\cancel{x_2^2}+2.25-3x_2-\cancel{x_1^2}-1+2x_1-\cancel{x_2^2}-1+2x_2\right]$$

$$\Rightarrow \frac{5}{2}\left[4.5-2-(x_1+x_2)\right]$$

$$\Rightarrow 2.5\left[2.5-(x_1+x_2)\right]=g(x)$$

Now ~~select to~~ classify as class ω_1, if $g(x)>0$

$$\Rightarrow 2.5\left[2.5-(x_1+x_2)\right]>0$$

$$\Rightarrow \boxed{x_1+x_2<2.5}$$

else classify as ω_2, if $g(x)<0$

$$2.5\left(2.5-(x_1+x_2)\right)<0$$

$$\Rightarrow \boxed{x_1+x_2>2.5}$$

Question 3:

2) Part 2 Solution:

3.2 We know that average risk $= R = \int R(\alpha_i|x)P(x)\,dx$

$\cancel{\Rightarrow R = \int_{R_1}}$

Let the decision boundary be at $x = \alpha_0 = [x_1 \;\; x_2]^T$

$$\Rightarrow \quad R = \int_0^x \left(\lambda_{11}P(\omega_1)P(x|\omega_1) + \lambda_{12}P(\omega_2)P(x|\omega_2)\right)dx$$

$$+ \int_x^\infty \left(\lambda_{21}P(\omega_1)P(x|\omega_1) + \lambda_{22}P(\omega_2)P(x|\omega_2)\right)dx$$

and we know that : $\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.5 & 0 \end{bmatrix}$

$$\Rightarrow R = \int_0^x \lambda_{12}P(\omega_2)P(x|\omega_2)\,dx - \int_\infty^x \lambda_{21}P(\omega_1)P(x|\omega_1)\,dx$$

Now minimizing R w.r.t x, at $x = \alpha_0 = [x_1 \;\; x_2]^T$

$$\left.\frac{dR}{dx}\right|_{x=\alpha_0} = 0 \Rightarrow \lambda_{12}P(\omega_2)P(\alpha_0|\omega_2) - \lambda_{21}P(\omega_1)P(\alpha_0|\omega_1) = 0$$

Question 3:

2) Part 2 Solution Continued:

$\Rightarrow$ $P(\omega_1) = P(\omega_2) = 1/2$, for 2-class classification

$\Rightarrow$ $(1)\cancel{\left(\frac{1}{2}\right)} P(\alpha_0|\omega_2) - \left(\frac{1}{2}\right)\cancel{\left(\frac{1}{2}\right)} P(\alpha_0|\omega_1) = 0$

$$\Rightarrow \quad \frac{P(\alpha_0|\omega_2)}{P(\alpha_0|\omega_1)} = \frac{1}{2}$$

taking natural logarithm

$$\Rightarrow \quad \ln \frac{P(\alpha_0|\omega_2)}{P(\alpha_0|\omega_1)} = \ln\left(\frac{1}{2}\right)$$

$$\Rightarrow \quad \ln \frac{\exp\left(-1/0.4 \begin{bmatrix} x_1 - 1.5 \\ x_2 - 1.5 \end{bmatrix}^T \begin{bmatrix} x_1 - 1.5 \\ x_2 - 1.5 \end{bmatrix}\right)}{\exp\left(-1/0.4 \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}^T \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}\right)} = \ln\left(\frac{1}{2}\right)$$

$$\Rightarrow \ln \exp\left(\frac{-1}{0.4}\left[\cancel{x_1^2} + 2.25 - 3x_1 + \cancel{x_2^2} + 2.25 - 3x_2 - \cancel{x_1^2} - 1 + 2x_1 - \cancel{x_2^2} - 1 + 2x_2\right]\right) = \ln(1/2)$$

$$\Rightarrow \frac{-1}{0.4}\left[2.5 - (x_1 + x_2)\right] = \ln(1/2)$$

$$\Rightarrow \quad x_1 + x_2 - 2.5 = 0.4(\ln(1) - \ln(2))$$

$$\Rightarrow \quad x_1 + x_2 - 2.5 = 0.4(0 - 0.693)$$

$$\Rightarrow \quad x_1 + x_2 - 2.5 = -0.2772$$

$$\Rightarrow \quad x_1 + x_2 = 2.2228$$

Hence, if the feature vectors $x = [x_1 \; x_2]^T$ follows the above equation, then we get the minimum average risk

Question 4:

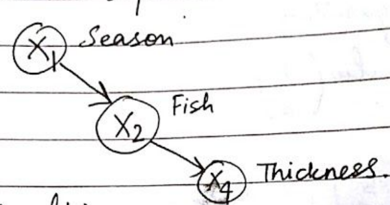
1) Part 1 Solution:

4.(1) SOLUTION:

Given:

(i) Soft evidence on the season ~~of~~ in which the fish was caught: $P(x_1) = (0.5, 0, 0, 0.5)$

(ii) The lightness has not been measured. So, we remove the variable X_3 from the network. New network:



(iii) This fish is thin.

SOL: Now we want to find the probability of the type of fish, given the above factors.

Lets calculate for salmon first:

$$P(X_2 = \text{salmon} \mid X_4 = \text{thin}) = \frac{P(X_2 = \text{salmon}, X_4 = \text{thin})}{P(X_4 = \text{thin})}$$

$$\Rightarrow \alpha \sum_{\forall x_1} P(X_1, X_2 = \text{salmon}, X_4 = \text{thin})$$

$$\Rightarrow \alpha \sum_{\forall x_1} P(x_1)\, P(X_2 = \text{salmon} \mid x_1)\, P(X_4 = \text{thin} \mid X_2 = \text{salmon})$$

$$\Rightarrow \alpha\, P(X_4 = \text{thin} \mid X_2 = \text{salmon}) \sum_{\forall x_1} \left(P(x_1)\, P(x_2 = \text{salmon} \mid x_1) \right)$$

$$\Rightarrow \alpha (0.6) \left[(0.5)(0.9) + (0)(0.3) + (0)(0.4) + (0.5)(0.8) \right]$$

$$\Rightarrow \alpha (0.6) [0.85] = 0.51\alpha$$

Similarly for $P(X_2 = \text{sea-bass} \mid X_4 = \text{thin})$ we get:

$$\Rightarrow \alpha (0.05) \left[(0.5)(0.1) + (0)(0.7) + (0)(0.6) + (0.5)(0.2) \right]$$

$$\Rightarrow \alpha (0.05)(0.15) = 0.0075\alpha$$

Question 4:

1) **Part 1 Solution Continued:**

Now $P(X_2 = salmon \mid X_4 = thin) + P(X_2 = seabass \mid X_4 = thin) = 1$

$\Rightarrow \quad 0.51\alpha + 0.0075\alpha = 1$

$\Rightarrow \quad \alpha = 1.9323$

Thus, after normalizing:

$P(X_2 = salmon \mid X_4 = thin) = 0.985$

$P(X_2 = seabass \mid X_4 = thin) = 0.015$

Hence, the fish is SALMON.

Question 4:

2) Part 2 Solution:

4·(2) Given: This fish is thin and medium lightness.

To find: $P(X_1 \mid X_3 = \text{medium}, X_4 = \text{thin})$

$$\Rightarrow P(X_1 \mid X_3=\text{medium}, X_4=\text{thin}) = \alpha \sum_{\forall x_2} P(X_1) P(X_2 \mid X_1) P(\overset{\text{medium}}{\overset{\shortparallel}{X_3}} \mid X_2) P(\overset{\text{thin}}{\overset{\shortparallel}{X_4}} \mid X_2)$$

(i) For $X_1 = \text{winter}$, we get:

$$P(X_1=\text{winter} \mid X_3=\text{medium}, X_4=\text{thin}) = \alpha\left((0.25)\left[(0.9)(0.33)(0.6) + (0.1)(0.1)(0.05)\right]\right.$$

$$\Rightarrow \alpha\, 0.25\left[0.1782 + 0.0005\right] = 0.044675\,\alpha$$

(ii) For $X_1 = \text{spring}$, we get

$$P(X_1=\text{spring} \mid X_3=\text{med}, X_4=\text{thin}) = \alpha(0.25)\left[(0.3)(0.33)(0.6) + (0.7)(0.1)(0.05)\right]$$

$$\Rightarrow \alpha(0.25)\left[0.0594 + 0.0035\right] = 0.015725\,\alpha$$

(iii) For $X_1 = \text{summer}$:

$$P(X_1=\text{summer} \mid X_3, X_4) = \alpha(0.25)\left[(0.4)(0.33)(0.6) + (0.6)(0.1)(0.05)\right]$$

$$= \alpha(0.25)\left[0.0792 + 0.003\right] = 0.02055\,\alpha$$

(iv) For $X_1 = \text{fall}$:

$$P(X_1=\text{fall} \mid X_3, X_4) = \alpha(0.25)\left[(0.8)(0.33)(0.6) + (0.2)(0.1)(0.05)\right]$$

$$= \alpha(0.25)\left[0.1584 + 0.001\right] = 0.03985\,\alpha$$

Normalizing them, we get: $\alpha = 8.2781457$ & probabities as:

(i) Winter = 0.37 (iv) fall = 0.33

(ii) Spring = 0.13

(iii) Summer = 0.17

Hence, it is most probably/likely WINTER.

Question 5:

Explore how the empirical error does or does not approach the Bhattacharyya bound as follows:

1) Write a procedure to generate sample points in d dimensions with a normal distribution having mean μ and covariance matrix Σ.

Solution 1:

Language used: Python

Libraries used: Numpy, Scipy

Code Snippet for a normal distribution with mean μ and covariance matrix Σ:

```
import matplotlib.pyplot as plt
import math
import numpy as np
import os
from scipy.stats import multivariate_normal
from scipy import random

d=3 # d dimension

# creating random covariance matrix
matrixSize = d
A = random.rand(matrixSize,matrixSize)
covariance = np.dot(A,A.transpose())

mu= random.rand(d) #mean


xlist,ylist=[],[]

# generating 1000 sample points in the value range between 0 and 1
for i in range(0,1000):
	x=np.random.rand(d) # d dimensional x feature vector
	y=multivariate_normal.pdf(x,mean=mu,cov=covariance) # class conditional
probability with x, mean mu and covariance matrix
	xlist.append(x)
	ylist.append(y)

print(xlist)
print(ylist)
```

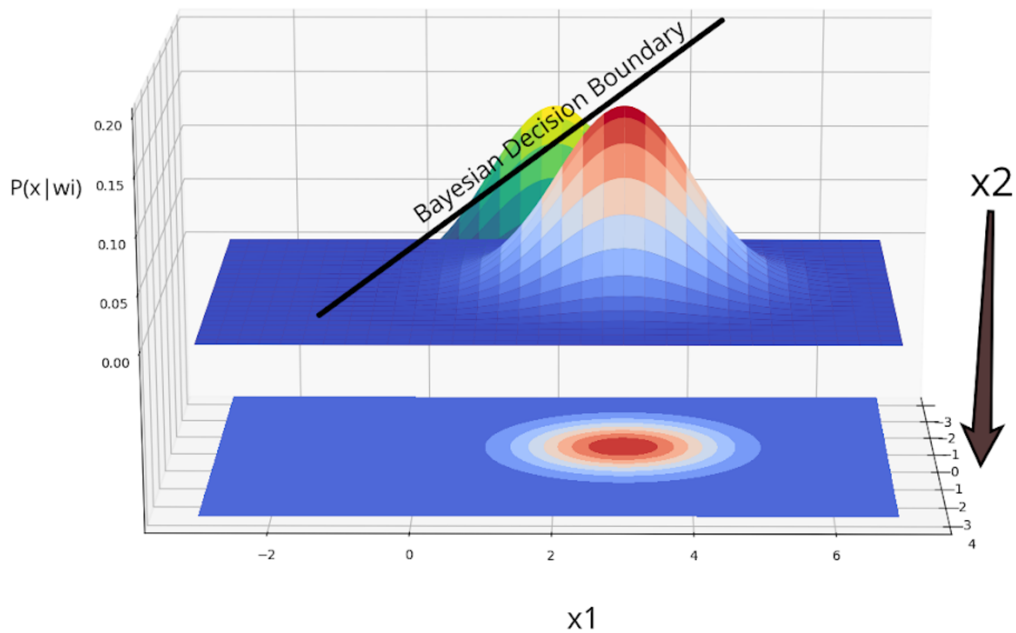
2) Consider the normal distributions p(x | ω1) ~ N ([0, 2]^T, I) and p(x | ω2) ~ N ([0, 3]^T, I) with P(ω1) = P(ω2) = 1/2. By inspection, state the Bayes decision boundary.

Solution 2:

Code snippet for generating the two pdfs and their 3d plots:

```
import matplotlib.pyplot as plt
import math
import numpy as np
import os
from scipy.stats import multivariate_normal
from scipy import random
from scipy.stats import norm
from matplotlib import cm
from mpl_toolkits.mplot3d import Axes3D

mean1= np.array([0, 2])
cov1= np.array([[1, 0],[0, 1]])

mean2= np.array([0, 3])
cov2= np.array([[1, 0],[0, 1]])

x1, x2= np.mgrid[-3:4:.01, -3:7:.01] # feature space for vector x=[x1 x2]
pos = np.dstack((x1, x2))

#generating pdf for the first distribution
rv = multivariate_normal(mean1, cov1)
pdf1=rv.pdf(pos)

#generating pdf for the second distribution
rv2=multivariate_normal(mean2,cov2)
pdf2=rv2.pdf(pos)

#for displaying the pdfs
fig = plt.figure()
ax = fig.gca(projection='3d')
ax.plot_surface(x1, x2, pdf1, rstride=3, cstride=3, linewidth=1,
antialiased=True,cmap=cm.viridis)
cset = ax.contourf(x1, x2, pdf1, zdir='z', offset=-0.15, cmap=cm.viridis)
ax.plot_surface(x1, x2, pdf2, rstride=3, cstride=3, linewidth=1,
antialiased=True,cmap=cm.coolwarm)
cset = ax.contourf(x1, x2, pdf2, zdir='z', offset=-0.15, cmap=cm.coolwarm)

# Adjust the limits, ticks and view angle
ax.set_zlim(-0.15,0.2)
ax.set_zticks(np.linspace(0,0.2,5))
ax.view_init(27, -21)
plt.show()
```

Image of the 3D Plot of the two distributions with the decision boundary: The boundary is actually a plane.



3) Generate n = 100 points (50 for $\omega 1$ and 50 for $\omega 2$) and calculate the empirical error.

```
import matplotlib.pyplot as plt
import math
import numpy as np
import os
from scipy.stats import multivariate_normal
from scipy import random
from scipy.stats import norm
from matplotlib import cm
from mpl_toolkits.mplot3d import Axes3D

mean1= np.array([0, 2])
cov1= np.array([[1, 0],[0, 1]])

mean2= np.array([0, 3])
cov2= np.array([[1, 0],[0, 1]])

# feature space for vector x=[x1 x2] over 50 points
x1=np.linspace(-100,100,50)
x2=np.linspace(-100,100,50)
X1,X2= np.meshgrid(x1,x2)
pos = np.dstack((X1, X2))
```

```
#generating pdf for the first distribution
rv = multivariate_normal(mean1, cov1)
pdf1=rv.pdf(pos)

#generating pdf for the second distribution
rv2=multivariate_normal(mean2,cov2)
pdf2=rv2.pdf(pos)

#calculating sum of probability of misclassified points in pdf1 and pdf2
error= sum(pdf1)+sum(pdf2)

print 0.5*sum(error)
```

This gives an **empirical error** of **0.0325**.

The above solution is based on the following:

5(3) Empirical Error $= \int_{R_2} P(x|\omega_1) P(\omega_1) dx + \int_{R_1} P(x|\omega_2) P(\omega_2) dx$

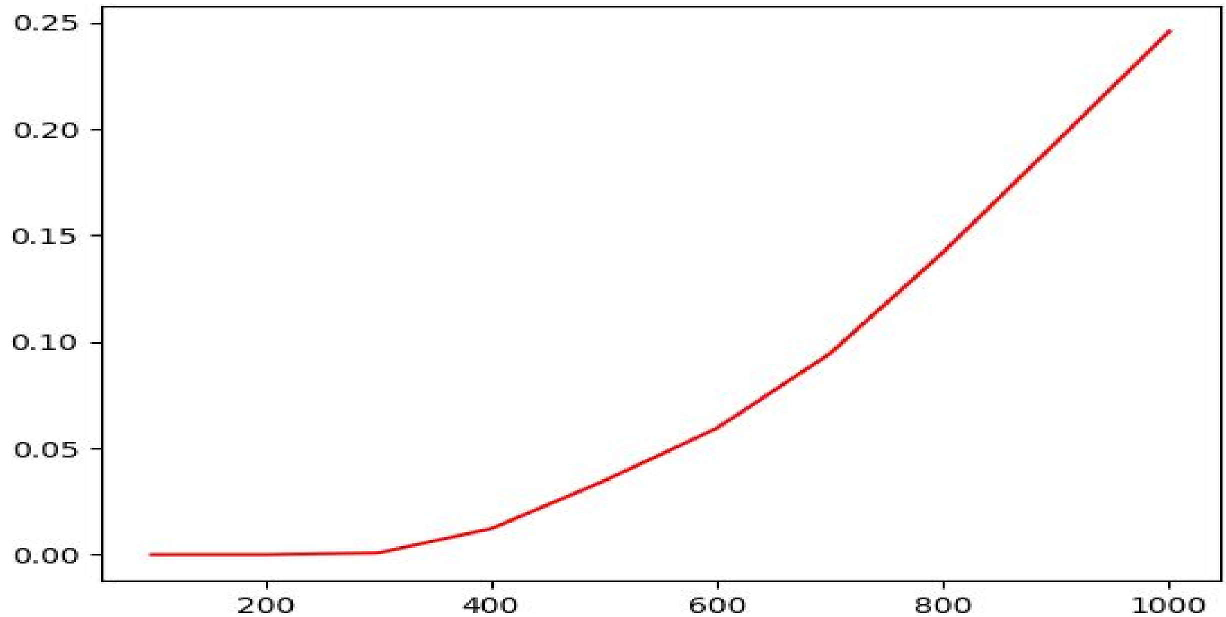
Given $P(\omega_1) = P(\omega_2) = 1/2$

$\Rightarrow \frac{1}{2}\left[\int_{R_2} P(x|\omega_1) dx + \int_{R_1} P(x|\omega_2) dx \right] =$ Empirical Error

4) Repeat for increasing values of n, 100 ≤ n ≤ 1000, in steps of 100 and plot your empirical error.

Solution:

```
import matplotlib.pyplot as plt
import math
import numpy as np
import os
from scipy.stats import multivariate_normal
from scipy import random
from scipy.stats import norm
from matplotlib import cm
from mpl_toolkits.mplot3d import Axes3D
mean1= np.array([0, 2])
cov1= np.array([[1, 0],[0, 1]])

mean2= np.array([0, 3])
cov2= np.array([[1, 0],[0, 1]])

errorlist=[]

# generating feature space for vector x=[x1 x2] over 100,200,.....1000 points
for i in range(1,11):
      x1=np.linspace(-1000,1000,100*i)
      x2=np.linspace(-1000,1000,100*i)
      X1,X2= np.meshgrid(x1,x2)
      pos = np.dstack((X1, X2))


      #generating pdf for the first distribution
      rv = multivariate_normal(mean1, cov1)
      pdf1=rv.pdf(pos)

      #generating pdf for the second distribution
      rv2=multivariate_normal(mean2,cov2)
      pdf2=rv2.pdf(pos)

      #calculating sum of probability of misclassified points in pdf1 and pdf2
      error= sum(pdf1)+sum(pdf2)
      errorlist.append(0.5*sum(error))

xaxis=[100,200,300,400,500,600,700,800,900,1000]
plt.plot(xaxis,errorlist,color='red')
plt.show()
```



Plot of Empirical Error(Y-axis) against the number of sample points(X-axis)

5) Discuss your results. In particular, is it ever possible that the empirical error is greater than the Bhattacharyya or Chernoff bound?

Solution:
The results above show that as the number of sample points increase, the empirical error increases. This might indicate that as the number of sample points increase, there is a possibility that more number of points may get misclassified. Therefore empirical error increases. No, it is never possible that the empirical/true error would be greater than the Bhattacharyya or Chernoff bound. The Chernoff bound analytically gives us an upper bound on the error. Bhattacharyya bound gives a looser bound than the chernoff bound, therefore Bhattacharyya bound would always be slightly higher than the Chernoff bound. Therefore, it is never possible that empirical/true error would be greater than the Bhattacharyya or Chernoff bound.