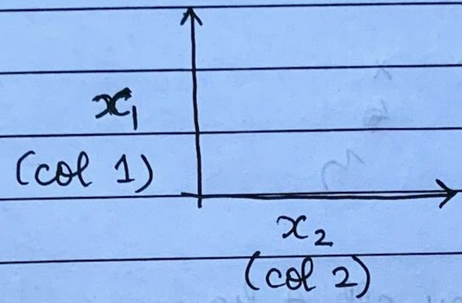
Derivation of Bayes' Error:

Step 1: Visualizing the Gaussians present

Distribution of data:

481 + 510 + 492 + 517 (ranges in each obv)
blob 1, blob 2, blob 3, blob 4



color of dot = class

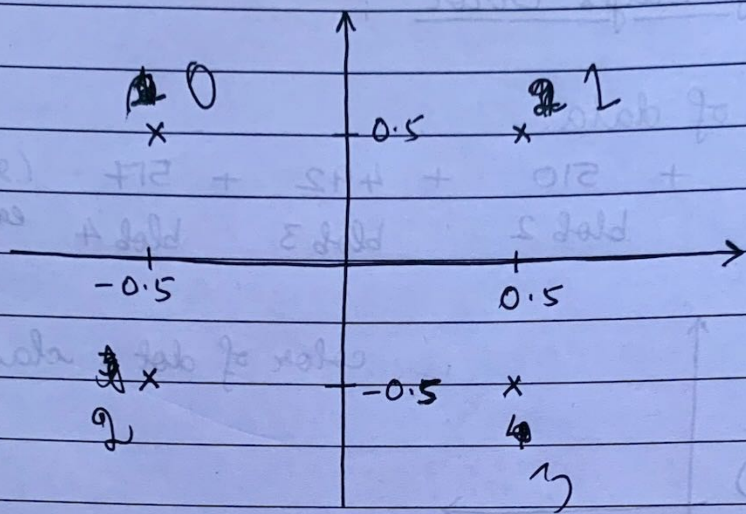
Classes allocated:

0	1	1	0
blob 1	blob 2	blob 3	blob 4

Means: $\Leftrightarrow$ Centers of ~~data~~ Gaussians

$$\left.\begin{array}{c} -0.5, 0.5 \\ 0.5, 0.5 \\ -0.5, -0.5 \\ 0.5, -0.5 \end{array}\right\} \text{means of clusters}$$

Std. dev. of Gaussians: 0.5



Quadrant 1 : Influenced by all 4 Gaussians

Bayes' error = 4 x Error in quadrant 1

Does not apply because all Gaussians appear in all quadrants

4-2D Gaussians, bivariate distribution

Orange → 0
Green → 1

~~Quad 2 & 3 → 0~~
~~Quad 1 & 4 → 1~~ } Primarily

Based on Visualization in Code

Index 0 & 3 → class 0
1 & 2 → class 1

$$p(x|C_1) = N(\mu_1, \Sigma_1) + N(\mu_3, \Sigma_3)$$

Step 3: Analytical definition of Bayes' error

$$p(x|C_2) = N(\mu_2, \Sigma_2) + N(\mu_4, \Sigma_4)$$

Assign: C_1 if $P(C_1|x) > P(C_2|x)$
otherwise, assign C_2

$$p(C_i|x) = \frac{p(x|C_i) \cdot p(C_i)}{p(x)}$$

If divides into regions R_i such that $x \in R_i$, implies x is in C_i

Bayes error = Integrate probability of incorrect decision over decision regions

$$= P(x \in R_1 | C_2) \cdot P(C_2) + P(x \in R_2 | C_1) \cdot P(C_1)$$

$$= \int_{R_1} p(x|C_2) \cdot P(C_2)\, dx + \int_{R_2} p(x|C_1) \cdot p(C_1)\, dx$$

Q: How are regions defined?

Step 4: Finding the PDF of the Gaussian
Ref: https://stats.stackexchange.com/questions/24772/finding-the-bayesian-classifier-for-a-bivariate-gaussian-distribution

papergrid
Date: / /

Quad 1

$$f(x|c_1) = \frac{1}{2\pi\sqrt{2}} \exp\left(-\frac{1}{2}\left(\frac{(x_1-0.5)^2}{0.5} + \frac{(x_2-0.5)^2}{0.5}\right)\right)$$

Quad 3

$$+ \frac{1}{2\pi\sqrt{2}} \exp\left(\frac{-1}{2}\left(\frac{(x_1+0.5)^2}{0.5} + \frac{(x_2+0.5)^2}{0.5}\right)\right)$$

$$\cancel{= \frac{1}{2\pi\sqrt{2}} \exp\left(\frac{-1}{2}\left(\cdots^2 -\right.\right.}$$

$$= \frac{1}{2\pi\sqrt{2}} \exp\left(\frac{-1}{\cancel{2}}\left(\frac{x_1^2 - x_1 + 0.5^2 + x_2^2 - x_2 + 0.5^2}{0.5}\right) \times \cancel{2}\right)$$

$$= \frac{1}{2\pi\sqrt{2}} \exp\left(\cancel{\frac{-1}{2}}\left(-2\left(x_1^2 - x_1 + 0.5^2 + x_2^2 - x_2 + 0.5^2\right)\right)\right)$$

$f(x|c_2)$

Even if I work this out, I will get the PDF & will assume both classes occur with equal probability
How do I get to Bayes' error?
What regions do I consider?

papergrid

Date: / /

Step 5: Assume Gaussians are independent (drawn at random)

$f(x_1 \& x_2) = 0 \ \forall \ x_1 \& x_2$ ~~ie~~, all the Gaussians are uncorrelated.

drawn randomly & independently

$\therefore$

$\therefore p(x|C_i) = p(x_1|C_i) \times p(x_2|C_i)$

0 & 1 vary along x1 & 0 & 3 vary along x2 (Ref Python Code)

Find overlap b/n 0 & 1 along x_1

$\mu = -0.5$ $\quad$ $\mu = 0.5$

$\sigma = 0.5$ $\quad$ $\sigma = 0.5$

Instead of actual values from code the original estimates of mean & std dev were used, assuming many samples -> tends to original values

pt. of intersection : 0, 0.4839 **Online calculator**

Limits & description from code

-1 to $0 \rightarrow \mu = 0.5, \sigma = 0.5$ } G1

0 to $1 \rightarrow \mu = -0.5, \sigma = 0.5$ } G0

Area : 0.197 + 0.197
-1 to 0 $\quad$ 0 to 1

} Online integrator Gaussian : $\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$

Area : 0.394 **Total error of G0 with G1 & G2**

Min. possible error 0.394 ?

combining along x_1 & x_2 : multiply 0.394×0.394
$= 0.155$

& similarly exists for G3 with 1 & 2 $\quad$ x2

Total error = Error of G0 with G1 & G2 + Error of G3 with G1 & G2 = 0.310

(G0 & G3), (G1 & G2) are same class, so no error