MSML/DATA 603 MIDTERM EXAM

Fall 2021 – Section 01D2

Closed book, closed n

No calculator or electr

Please, turn off cell phones & smart phones

No speaking or whispering to classmates during exam

Date: October 29, 2021

HONOR PLEDGE:

I pledge on my honor that I have not given or received any unauthorized assistance on this examination. I pledge that I have not intentionally used or attempted to use unauthorized materials or information to assist me in this examination, and I pledge on my honor that I have not looked at or read anything from any classmate's exam papers or scrap-material sheets.

SAMPLE

Student's name and UID (required)

(1) [3 points] Consider a two-label Bayesian Classification case where: all features are statistically independent, their likelihoods are Gaussian, and the features have been scaled to have the same variance. Furthermore, the priors are all the same. The mean of the likelihood for ω_1 is $\vec{\mu_1} = [5 \quad 10]^T$ and the mean of the likelihood for ω_2 is $\vec{\mu_2} = [11 \quad 4]^T$. A new instance is studied and it has features $\vec{x} = [7 \quad 7]^T$. (a) What classification would our Bayesian classifier apply to this new instance? (b) Please, justify your response in question (a). (c) Please, sketch the decision boundary in the feature space.

(a) Answer: ω_1

(b) Answer: Because of the terms underlined above, this is a "minimum distance classifier": Wanda's vector $\vec{x}$ is closer to $\vec{\mu_1}$

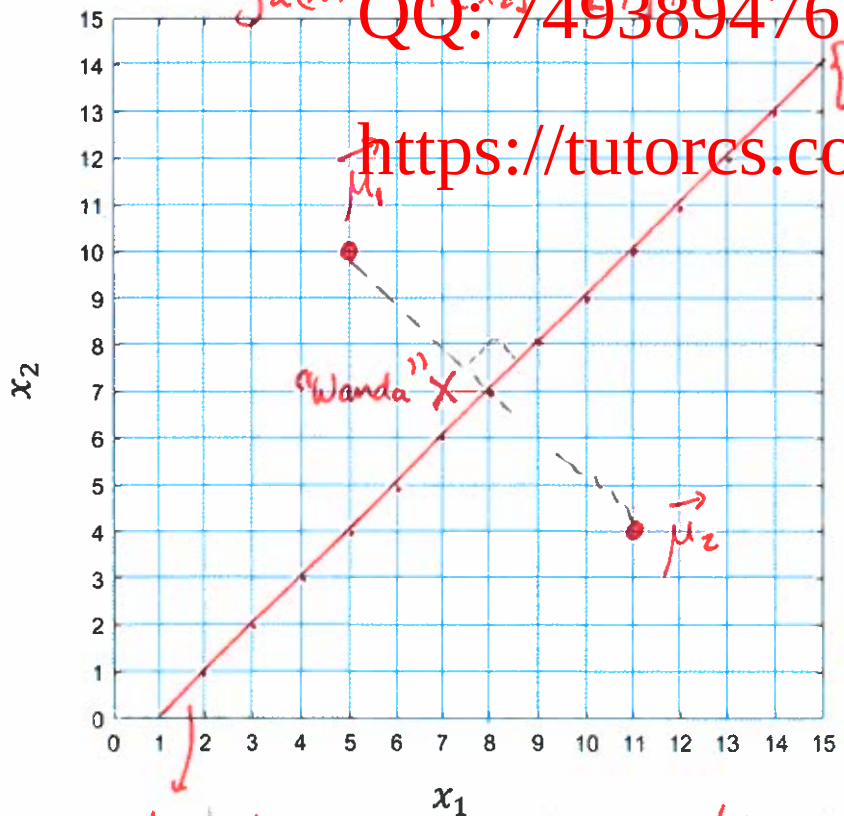
You could just draw the boundary to get full credit, but if you want to solve for it: $g_1(\vec{x}) = g_2(\vec{x})$ at all $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$g_1(\vec{x}) = -\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5 \\ 10 \end{bmatrix} \right\|^2$

$g_2(\vec{x}) = -\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 11 \\ 4 \end{bmatrix} \right\|^2$

where:

(c) Answer:



$\{x_1^2 - 10x_1 + 25\} + \{x_2^2 - 20x_2 + 100\} =$

$\{x_1^2 - 22x_1 + 121\} + \{x_2^2 - 8x_2 + 16\}$

$(-20+8)x_2 = (-22+10)x_1 + 121 + 16 - 25 - 100$

$-12x_2 = -12x_1 + 12$

$\Rightarrow \boxed{x_2 = x_1 - 1}$ → equation of line

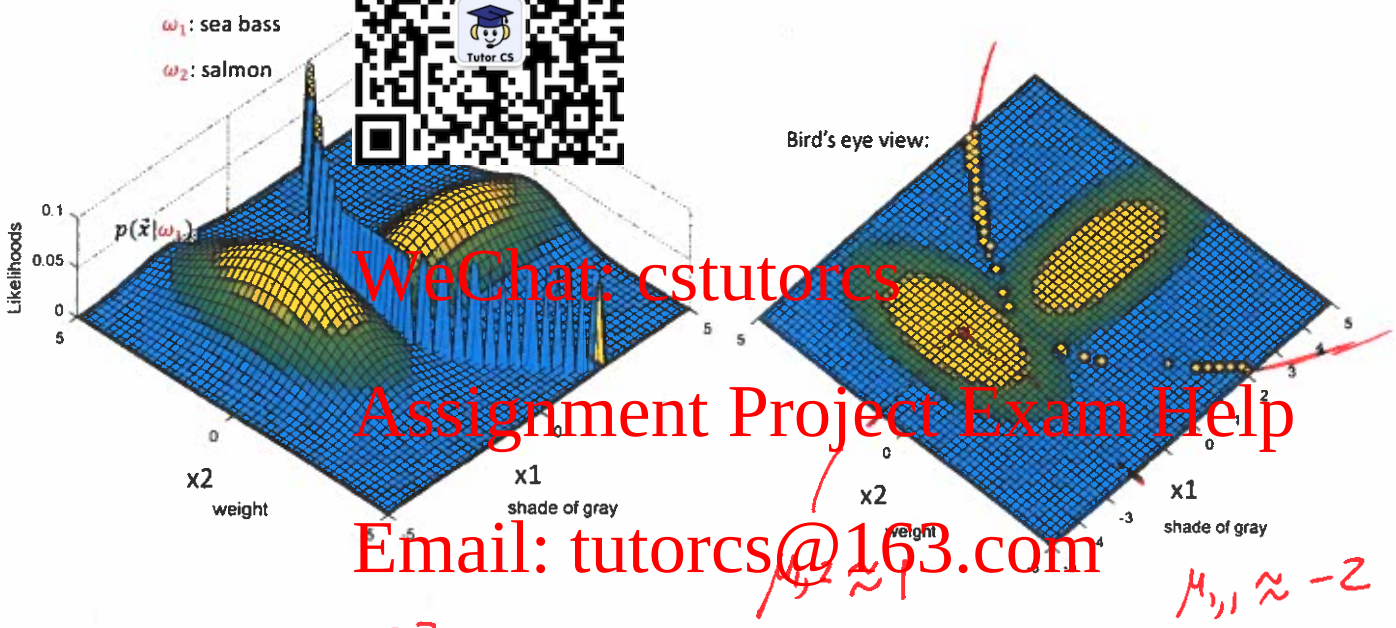
(Draw it or derive it, either way you get full credit)

$+16 \quad -25$

$37 \quad 12$

decision boundary is a line that is equal distance from each $\vec{\mu}$

(2) [3 points] Consider the Bayesian Decision problem we saw in class: viewing the likelihoods for both labels, ω_1 and ω_2, below, please, (a) provide the expectation of the likelihood for ω_1 (reading the graphs is not easy, so you don't need to be exact) (b) what can you say about the covariance matrices for both likelihoods? (c) Estimate which one seems to [illegible](s), ω_1 or ω_2?



$\mu_{2,1} \approx -1$

$\mu_{1,1} \approx -2$

Answer (a): $\vec{\mu}_1 \approx \begin{bmatrix} -2 \\ \end{bmatrix}$

Answer (b): Features are independant, so covariance matrices are diagonal. Features do not have the same variances (not scaled). For ω_1 variance of x_2 > variance of x_1. For ω_2 variance of x_1 > variance of x_2.

Answer (c): ω_2 seems to have smaller variances, as the boundary wraps closer to $\vec{\mu}_2$

(3) [2 points] Please, recall our discussion in lecture and (a) explain the difference between a Regression Problem and an Anomaly Detection Problem, and, (b) explain the difference between Batch Learning and Online Learning.

Answer (a): Regression: predict a real value from a continuous interval.

Anomaly detect outlier behavior

Answer (b): Batch learning: system does not learn incrementally with instances. All the training data is used offline before launching system.

Online learning: system is trained incrementally with one-by-one training instances. System designed to change quickly and autonomously as new instances arrive.

(4) [3 points] Consider the 2-label classification case we saw in lecture in reference to admitting a student to "X" Medical School. We have training instances that each include two attributes: (i) undergraduate GPA and (ii) grade in organic chemistry. Assume the following:

- all instances (applicar re independent from each other,
- all instances within a "admitted" or ω_2 ="not admitted) are identically distributed,
- the likelihood functio Gaussian with pdf $\mathcal{N}(\vec{\mu_1}, \Sigma_1)$ and $\mathcal{N}(\vec{\mu_2}, \Sigma_2)$, respectively.

(a) Given the information above, what simple, fast method did we use to estimate $\vec{\mu_1}, \Sigma_1, \vec{\mu_2}$, and Σ_2?

Maximum Likelihood estimation

(b) The training instances are provided in the Table 1, here. Using the method you named in (a), please find the estimated for the four parameters.

Instance number	Label	undergraduate GPA (x_1)	grade in organic chemistry (x_2)
1	Admitted	3	4
2	Admitted	4	2
3	Not Admitted	2	3
4	Not Admitted	1	4

$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

[It's OK to leave answers as fractions, such as for example 7/4]

$$\vec{\mu}_1 = \frac{1}{2}\left\{\begin{bmatrix}3\\4\end{bmatrix} + \begin{bmatrix}4\\2\end{bmatrix}\right\} = \begin{bmatrix}3.5\\3\end{bmatrix} = \begin{bmatrix}7/2\\3\end{bmatrix}$$

$$\Sigma_1 = \frac{1}{2}\left\{\left(\begin{bmatrix}3\\4\end{bmatrix} - \begin{bmatrix}3.5\\3\end{bmatrix}\right)\left(\begin{bmatrix}3\\4\end{bmatrix} - \begin{bmatrix}3.5\\3\end{bmatrix}\right)^T + \left(\begin{bmatrix}4\\2\end{bmatrix} - \begin{bmatrix}3.5\\3\end{bmatrix}\right)\left(\begin{bmatrix}4\\2\end{bmatrix} - \begin{bmatrix}3.5\\3\end{bmatrix}\right)^T\right\}$$

$$= \frac{1}{2}\left\{\begin{bmatrix}-0.5\\1\end{bmatrix}\begin{bmatrix}-0.5 & 1\end{bmatrix} + \begin{bmatrix}+0.5\\-1\end{bmatrix}\begin{bmatrix}+0.5 & -1\end{bmatrix}\right\}$$

$$= \frac{1}{2}\left\{\begin{bmatrix}0.25 & -0.5\\-0.5 & 1\end{bmatrix} + \begin{bmatrix}0.25 & -0.5\\-0.5 & 1\end{bmatrix}\right\} = \begin{bmatrix}\frac{1}{4} & -\frac{1}{2}\\-\frac{1}{2} & 1\end{bmatrix}$$

$$\vec{\mu}_2 = \frac{1}{2}\left\{\begin{bmatrix}2\\3\end{bmatrix} + \begin{bmatrix}1\\4\end{bmatrix}\right\} = \begin{bmatrix}1.5\\3.5\end{bmatrix} = \begin{bmatrix}3/2\\7/2\end{bmatrix}$$

$$\Sigma_2 = \frac{1}{2}\left\{\left(\begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}1.5\\3.5\end{bmatrix}\right)\left(\begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}1.5\\3.5\end{bmatrix}\right)^T + \left(\begin{bmatrix}1\\4\end{bmatrix} - \begin{bmatrix}1.5\\3.5\end{bmatrix}\right)\left(\begin{bmatrix}1\\4\end{bmatrix} - \begin{bmatrix}1.5\\3.5\end{bmatrix}\right)^T\right\}$$

$$\Sigma_2 = \frac{1}{2}\left\{ \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}\begin{bmatrix} 0.5 & -0.5 \end{bmatrix} + \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}\begin{bmatrix} -0.5 & 0.5 \end{bmatrix} \right\}$$

$$= \frac{1}{2}\left\{ \begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 0.25 \end{bmatrix} + \begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 0.25 \end{bmatrix} \right\}$$

$$= \begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 0.25 \end{bmatrix}$$

(5) [1 point] From what we saw in class about Linear Discriminant Functions, if we have a three-label classification case (ω_1 ="good email", ω_2 = "phishing email", and ω_3 ="ransomware email"). Say, we are given the weight vectors for the linear discriminant functions (Note: these weight vectors and feature vectors already include the bias integrated into them):

$\overrightarrow{w_1} = \begin{bmatrix}2\\1\\0\end{bmatrix}$ and $\overrightarrow{w_2} = \begin{bmatrix}1\\3\\1\end{bmatrix}$ and $\overrightarrow{w_3} = \begin{bmatrix}0\\2\\5\end{bmatrix}$. Say a new incoming instance has the following feature vector: $\vec{x} = \begin{bmatrix}2\\3\\1\end{bmatrix}$.

Please, determine what label our l dict for this new instance.

Answer: ω_2 " phi

ω_1 : $[2 \;\; 1 \;\; 0] \begin{bmatrix}2\\3\\1\end{bmatrix} = 4 + 3 + 0 = 7$

ω_2 : $[1 \;\; 3 \;\; 1] \begin{bmatrix}2\\3\\1\end{bmatrix} = 2 + 9 + 1 = 12$ ✓ biggest discriminant function

ω_3 : $[0 \;\; 2 \;\; 5] \begin{bmatrix}2\\3\\1\end{bmatrix} = 0 + 6 + 5 = 11$

Good luck!!