MSML/DATA 603 MIDTERM EXAM

Fall 2021 — Session 0101

Closed book, closed no

No calculator or electr

Please, turn off cell phones & smart phones

No speaking or whispering to classmates during exam

Date: October 26, 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 $\vec{\mu_1} = [0 \quad 3]^T$ and the mean of the likelihood for ω_2 is $\vec{\mu_2} = [6 \quad 9]^T$. A new instance is studied and it has feature vector $\vec{x} = [4 \quad 6]^T$. (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_2

(b) Answer: Becau[illegible]ptions underlined above, this is a "minimum distance classifier".

Wanda's features are closer to $\vec{\mu_2}$.

You could just draw the boundary to get full credit, but if you [illegible] $g_1(\vec{x}) = -\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 3 \end{bmatrix} \right\|^2$

$g_2(\vec{x}) = -\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 6 \\ 9 \end{bmatrix} \right\|^2$

(c) Answer:

$g_1(\vec{x}) = g_2(\vec{x})$ at all $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ where

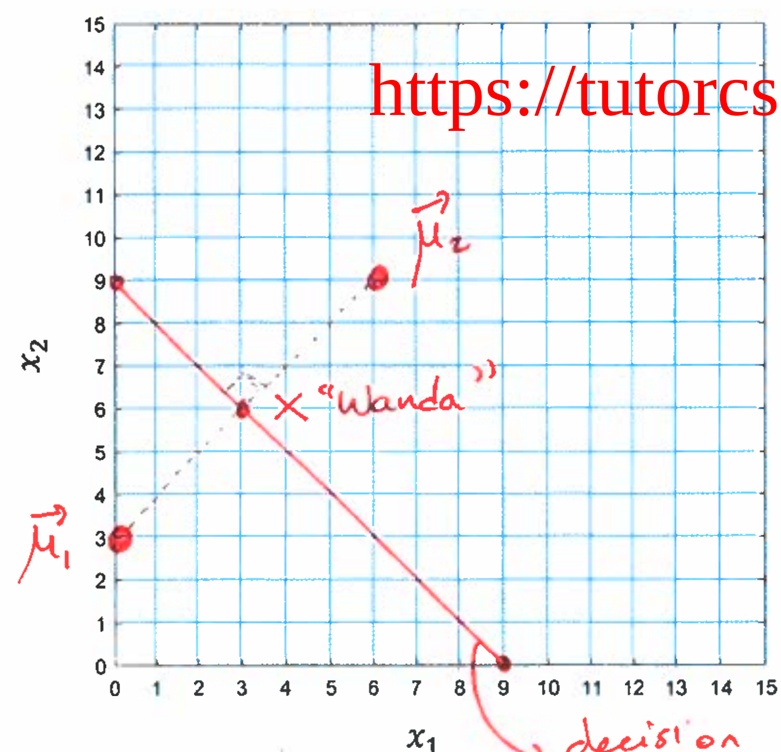
$\{\cancel{x_1^2}\} + \{\cancel{x_2^2} - 6x_2 + 9\} = \{\cancel{x_1^2} - 12x_1 + 36\} + \{\cancel{x_2^2} - 18x_2 + 81\}$

$12x_2 = -12x_1 + 108$

$\boxed{x_2 = -x_1 + 9}$

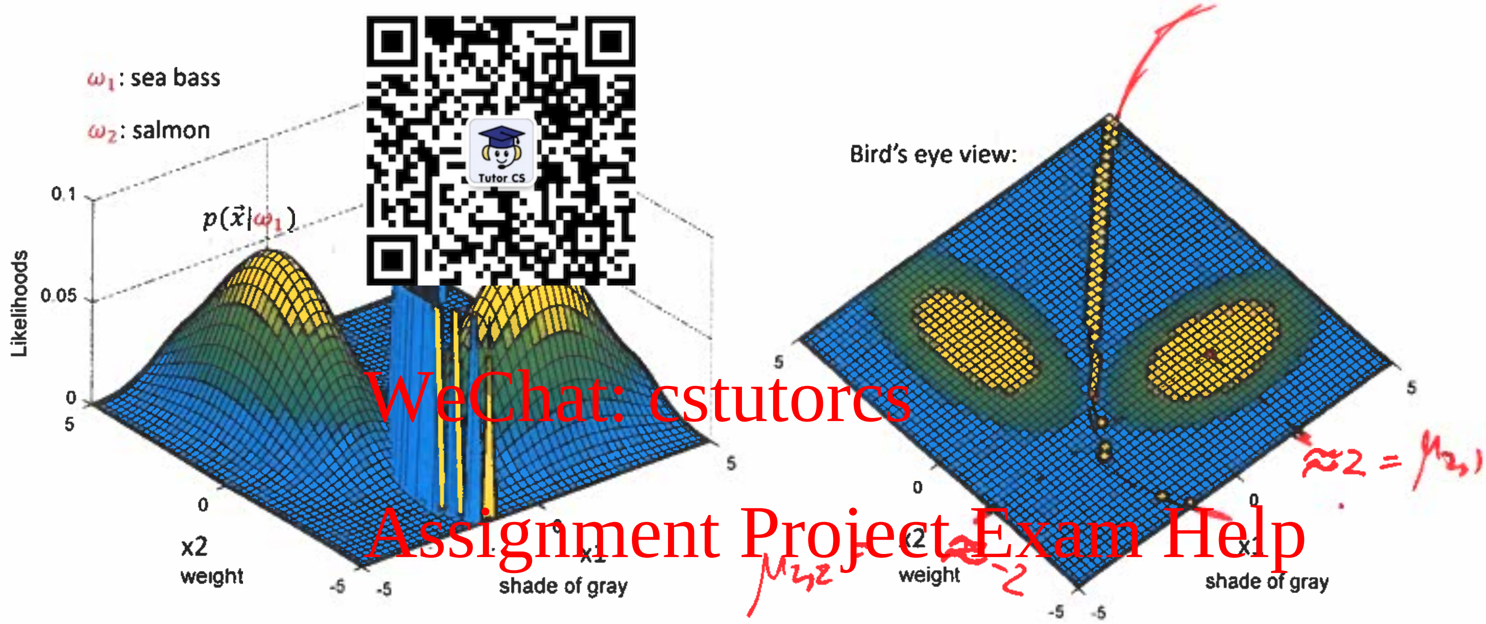
↳ equation of line

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



decision boundary is a line that is equal distance from each μ

(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_2 (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 have larger variance(s), ω_1 or ω_2?



$\vec{\mu}_2 = \begin{bmatrix} \mu_{2,1} \\ \mu_{2,2} \end{bmatrix}$

Answer (a): $\vec{\mu}_2 \approx \begin{bmatrix} 2 \\ -2 \end{bmatrix}$

Answer (b): Features are independent, so covariance matrices are diagonal. Features do not have the same variance (not scaled). For w_1, x_2's variance > x_1's variance. For w_2, x_1's variance > x_2's variance.

Answer (c): w_2 seems to have smaller variances, since boundary wraps closurer to $\vec{\mu_2}$.

(3) [2 points] Please, recall our discussion in lecture and (a) explain the difference between a Regression Problem and a Classification Problem, and, (b) explain the difference between Instance Based Learning and Model Based Learning.

Answer (a): Regression: predict a real value from a [illegible]uous interval.

classif[illegible]assify an item into exactly [illegible]e of several (countable) categories.

Answer (b): Instance based: basis for decision is empirical [illegible] and not formulas

Model based: create a model formula to predict label

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

(a) Given the information ab[illegible] 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	grade in organic chemistry
1	Admitted	4	3
2	Admitted	3	2
3	Not Admitted	2	1
4	Not Admitted	1	3

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

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

$$\vec{\mu}_1 = \frac{1}{2}\left\{\begin{bmatrix}4\\3\end{bmatrix} + \begin{bmatrix}3\\2\end{bmatrix}\right\} = \begin{bmatrix}3.5\\2.5\end{bmatrix} = \begin{bmatrix}7/2\\5/2\end{bmatrix}$$

$$\Sigma_1 = \frac{1}{2}\left\{\left(\begin{bmatrix}4\\3\end{bmatrix} - \begin{bmatrix}3.5\\2.5\end{bmatrix}\right)\cdot\left(\begin{bmatrix}4\\3\end{bmatrix} - \begin{bmatrix}3.5\\2.5\end{bmatrix}\right)^T + \left(\begin{bmatrix}3\\2\end{bmatrix} - \begin{bmatrix}3.5\\2.5\end{bmatrix}\right)\cdot\left(\begin{bmatrix}3\\2\end{bmatrix} - \begin{bmatrix}3.5\\2.5\end{bmatrix}\right)^T\right\}$$

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

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

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

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

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

$$\Sigma_2 = \frac{1}{2}\left\{\begin{bmatrix} 0.25 & -0.5 \\ -0.5 & 1 \end{bmatrix} + \begin{bmatrix} 0.25 & -0.5 \\ -0.5 & 1 \end{bmatrix}\right\} = \begin{bmatrix} 0.25 & -0.5 \\ -0.5 & 1 \end{bmatrix}.$$

(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):

$\vec{w_1} = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$ and $\vec{w_2} = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix}$ and $\vec{w_3} =$ … instance has the following feature vector: $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

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

Answer: w_3 "ransom ware"

w_1: $[0 \ 2 \ 2]\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 0+4+2 = 6$

w_2: $[4 \ 0 \ 2]\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 4+0+2 = 6$

w_3: $[1 \ 5 \ 0]\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 1+10+0 = 11$ ✓ biggest discriminant function

Good luck!!