, Aris Podotas, A.M: 7115152400040, Machine midterms
this is the serial number

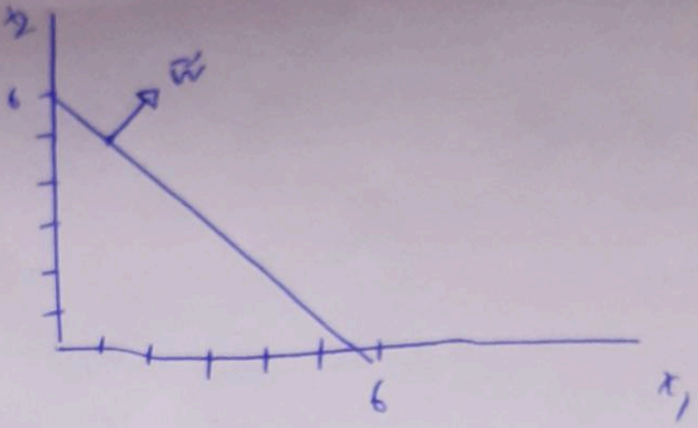
(P1)

(Q1) a) given $w_0 = 6$ the red class has $t = 1$
the green -//- has $t = -1$

to find the w_1, w_2 components we can look to two points on the plot. Namely when $(x_1 = 0, x_2)$, $(x_1, x_2 = 0)$ these two point will help for the general form of the line equation Being: $w_1 x_1 + w_2 x_2 - w_0$. Substituting what we know we get $w_1 x_1 + w_2 x_2 - 6$
then for $(x_1 = 0, x_2) \Rightarrow w_1 0 + w_2 x_2 - 6 = 0$
we see that for $x_1 = 0$ $x_2 = 6$ on the plot
$\Rightarrow \cancel{w_1 0} + 6 w_2 - 6 = 0 \Rightarrow 6 w_2 = 6 \Rightarrow w_2 = 1$
0

and we get $w_1 x_1 + 1 x_1 - 6$
then for $(x_1, x_2 = 0)$ $w_1 x_1 + 0 - 6 = 0 \Rightarrow w_1 x_1 = 6$
we see that for $x_2 = 0$ $x_1 = 6$ on the plot
$\Rightarrow w_1 = \frac{6}{6} = 1$ so finally we have $\underline{1x_1 + 1x_2 - 6 = 0}$
and we can make sure of this by re-applying our points
$1 \cdot 6 + 1 \cdot 0 - 6 = 0$: and this holds true. this will suffice for w_1, w_2. The weight vector has a direction towards the positive semi-space. This space has all points $f(x_1, x_2) > w_1 x_1 + w_2 x_2 - w_0$. we see the point $(5, 5)$
$\Rightarrow 1 \cdot 5 + 1 \cdot 5 - 6 = 10 - 6 = 4 > 0$ so the positive semi space is "towards the green points"

(Q1) a) continued: then drawing the weight vector



Q1) b)

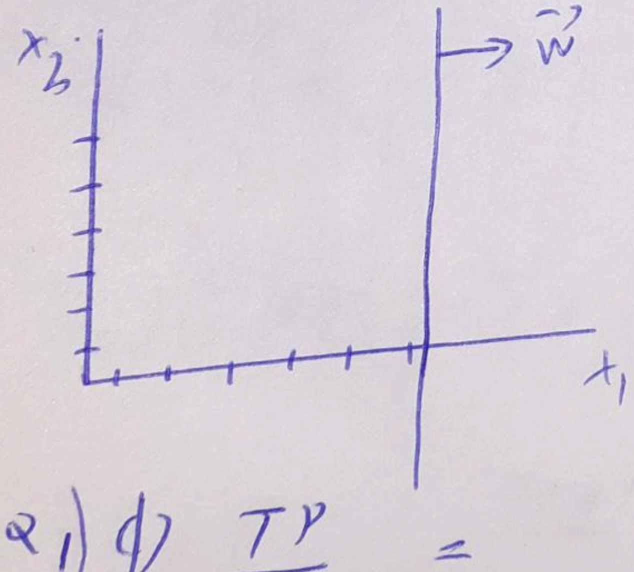
$f = 1x_1 + 1x_2 - 6$

w_1	w_2	w_0	tx_1	tx_2	$t x$	update Y/N	Δw		
1	1	-6	~~+6,08~~ +3,9	~~0~~ -6,3	0,05	Y	+3,9	-6,3	0,05
4,9	-5,3	-5,95	5	-5	0,05	N	0	0	0
4,9	-5,3	-5,95	5	-4	0,05	N	0	0	0
4,9	-5,3	-5,95	3	-4	0,05	N	0	0	0
4,9	-5,3	-5,95	5	-3	0,05	N	0	0	0
4,9	-5,3	-5,95	6	-3	0,05	N	0	0	0
4,9	-5,3	-5,95	6	-1	0,05	N	0	0	0
4,9	-5,3	-5,95	2	-3	0,05	N	0	0	0
4,9	-5,3	-5,95	2	-2	0,05	N	0	0	0
4,9	-5,3	-5,95	3	-2	0,05	N	0	0	0

Q2 b) perceptron for $1x_1 + 1x_2 - 6$ from top to Bottom left to right with rounding of coordinance

w_1	w_2	w_0 -6,9	tx_1	tx_2	$t\varepsilon$	ub Y/N	Δw
1	1	-6,5	-4	-6	-0,05	Y	-4 -6 -0,05
-3	-5	-5,55	-5	-5	-0,05	N	0 0 0
-3	-5	-5,95	-3	-4	-0,05	N	0 0 0
-3	-5	-5,95	-5	-4	-0,05	N	0 0 0
-3	-5	-5,95	2	3	-0,05	Y	2 3 -0,05
-1	-2	-6,10	-3	-5	-0,05	N	0 0 0
-1	-2	-6,1	-6	-3	-0,05	N	0 0 0
-1	-2	-6,1	2	2	-0,05	Y	2 2 -0,05
1	0	-6,15	3	2	-0,05	N	0 0 0
1	0	-6,15	6	1	-0,05	N	0 0 0

=> our new line is $1x_1 + 0x_2 - 6,15 = 0$



Q1) d) $\frac{TP}{TP + FP} =$

Q2/ a) Lets start with the why.

A Bayesian network is a acyclic graph (then G2 is not our graph). => G1 is a Bayesian network

Q2 B)

P(A)

	True	False
A	A_1	A_2

P(B)

	T	F
B	B_1	B_2

P(E | C, B)

B	$E_1 \mid B_1$	$E_1 \mid B_0$
C	$E_1 \mid C_1$	$E_1 \mid C_0$
B	$E_0 \mid B_1$	$E_0 \mid B_0$
C	$E_0 \mid C_1$	$E_0 \mid C_0$

P(C | A, B)

A	$C_1 \mid A_1$	$C_1 \mid A_0$
B	$C_1 \mid B_1$	$C_1 \mid B_0$
A	$C_0 \mid A_1$	$C_0 \mid A_0$
B	$C_0 \mid B_1$	$C_0 \mid B_0$

P(D | A, C)

A	$D_1 \mid A_1$	$D_1 \mid A_0$
C	$D_1 \mid C_1$	$D_1 \mid C_0$
A	$D_0\ A_1$	$D_0\ A_0$
C	$D_0\ C_1$	$D_0\ C_0$

P(F | E)

E	$F_1 E_1$	$F_1 E_0$
E	$F_0 E_1$	$F_0 E_0$

Q2 C) $P(A, B, C, D, E, F)$

$= P(F \mid E) \cdot P(E \mid C) \cdot P(E \mid B)$
$\cdot P(B) \cdot P(C \mid B) \cdot P(B) \cdot P(C \mid A) \cdot P(A)$
$\cdot P(D \mid A) \cdot P(A) \cdot P(D \mid C)$
$\cdot P(C \mid A) \cdot P(A) \cdot P(C \mid B) \cdot P(B)$

Q2 d_b))

$$P(F_1 \mid A_1) = \frac{\sum_{x\in, y\in, i\in, j\in} (P(A_1, B_x, C_y, D_i, E_j, F_1)}{P(A)}$$

$$P(A_1 \mid F_1) = \frac{P(A)}{\sum_{x\in, y\in, i\in, j\in} (P(A_1, B_x, C_y, D_i, E_j, F_1)}$$

Q3) the classes are equiprobable => $P(\omega_1) = P(\omega_2)$

a) considering that we can see our classes have different covariance matrixes Σ_1, Σ_2 we expect our classifier to be non linear.

Q3 B) yes. Because the naive Bayes classifier assumes that each dimension for each feature is independent of the others. Tis can be seen on the plot because both the isolevel curves for our 2 Gaussians have no rotation and that means that the Σ_1, Σ_2 are diagonal ($\Sigma_1 = \sigma_{\sigma_1} I$, $\Sigma_2 = \sigma_{\sigma_2} I$)

Q3 C) from B) we see that

$\Sigma_1 = \begin{bmatrix} \sigma_{12} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} \sigma_{12} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$ (note σ_{12} for Σ_1 and Σ_2 are different they were written like this for readability)

An eigen vector is defined as a vector that satisfies

$A\vec{x} = \lambda \vec{x}$ (were λ is a scalar) we also know that the

=> $\Sigma_1 x_1 = \lambda_1 x_1$ => direction of one of our vectors points to the ~~middle~~ minimum distance from the mean to the isolevel curves and the other to the most

remember that

$\|x\| = \sqrt{x}\sqrt{x}$

continued then o-v Σ_1 eigen vectors are $z_1 = \begin{bmatrix} 0,3 \\ 0 \end{bmatrix}$ $z_2 = \begin{bmatrix} 0 \\ 0,7 \end{bmatrix}$ and Σ_2 =>

$z_1 \begin{bmatrix} 0,7 \\ 0 \end{bmatrix}$ $z_2 \begin{bmatrix} 0 \\ 0,3 \end{bmatrix}$

q3 d) $\Sigma_1 = 1 \cdot \begin{bmatrix} 0,3 \\ 0 \end{bmatrix} [0,3 \quad 0]^T + 1 \begin{bmatrix} 0 \\ 0,7 \end{bmatrix} [0 \quad 0,7] = \begin{bmatrix} 0,09 & 0 \\ 0 & 0,49 \end{bmatrix}$

$\Sigma_2 = \begin{bmatrix} 0,49 & 0 \\ 0 & 0,9 \end{bmatrix}$

Q 3 e) Lets notice that $\mu_1 = (0,0)$ $\mu_2 = (4,4)$

$-x^T \begin{bmatrix} 0,09 & 0 \\ 0 & 0,49 \end{bmatrix}^{-1} + x^T \begin{bmatrix} 0,09 & 0 \\ 0 & 0,49 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \cancel{\ldots}$

$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0,09 & 0 \\ 0 & 0,49 \end{bmatrix}^{-1} x - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0,09 & 0 \\ 0 & 0,49 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 2\ln \cancel{P(w_1)} - 2\ln \det$

$= -x^T \begin{bmatrix} 0,49 & 0 \\ 0 & 0,09 \end{bmatrix} x + x^T \begin{bmatrix} 0,49 & 0 \\ 0 & 0,09 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 4 \end{bmatrix} + [4 \quad 4]^T \begin{bmatrix} 0,49 & 0 \\ 0 & 0,09 \end{bmatrix}^{-1} x$ $\begin{bmatrix} 0,09 & 0 \\ 0 & 0,49 \end{bmatrix}$

$- [4 \quad 4]^T \begin{bmatrix} 0,49 & 0 \\ 0 & 0,09 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 4 \end{bmatrix} + 2\ln \cancel{P(w_2)}) - 2\ln \det \begin{bmatrix} 0,49 & 0 \\ 0 & 0,09 \end{bmatrix}$

q3 f) we must calculate the mahalanobis Distance

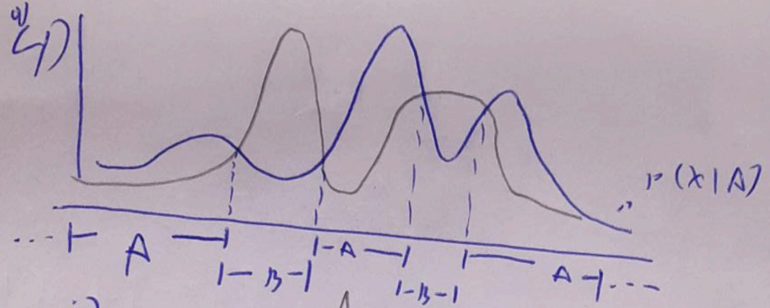
$(x-\mu_1)^T \Sigma_1^{-1} (x-\mu_1) = [3 \quad 1] \begin{bmatrix} 11,1 & 0 \\ 0 & 2,04 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = [33,3 \quad 2,04] \begin{bmatrix} 3 \\ 1 \end{bmatrix}$
$= 99,9 + 2,04$

$(x-\mu_2)^T \Sigma_2^{-1} (x-\mu_2)$ $[-1 \quad -3] \begin{bmatrix} 2,04 & 0 \\ 0 & 11,1 \end{bmatrix} \begin{bmatrix} -1 \\ -3 \end{bmatrix} = [-2,04 \quad -33,3] \cdot \begin{bmatrix} -1 \\ -3 \end{bmatrix}$

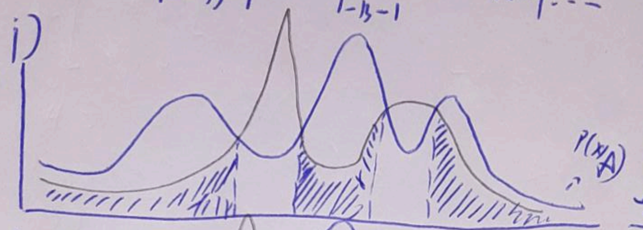
$= 2,04 + 99,9$

or in other words our point is on the Descision curve. Lets assign randomly to w_1

Q4 a)

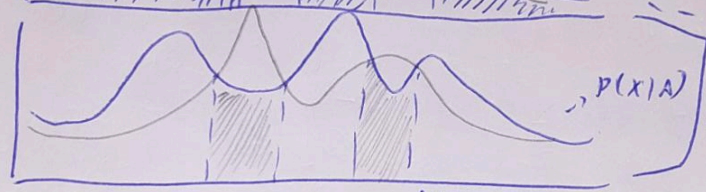


Q4 b) i)



under P(x|A) But Belonging to P(x|B)

ii)



under P(x|B) But Belonging to P(x|A)

Q4 c) Note the area of a ~~square~~ parallelogram $= (1 \cdot 0{,}02)\,([dx \cdot dx])$

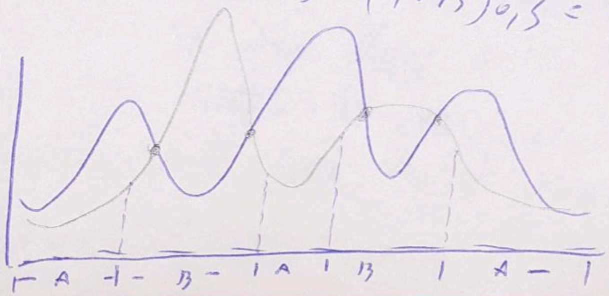
$\cancel{P_e = 0{,}5 \int p(x|A)\,dx + 0{,}5 \int p(x|A)}$

$$P_e = 0{,}5\left(\int_{-\infty}^{-4} p(x|A)\,dx + \int_{-0{,}5}^{2{,}5} p(x|A) + \int_{5{,}01}^{+\infty} p(x|A)\right)$$

$$+\, 0{,}5\left(\int_{-4}^{-0{,}5} p(x|B)\,dx + \int_{2{,}5}^{5{,}01} p(x|B)\,dx\right) \simeq 0{,}5\,(0{,}04 + 0{,}12 + 0{,}1) + 0{,}5\,(0{,}06 + 0{,}105)$$

$$= 0{,}5 \cdot (0{,}17 + 0{,}165) = (0{,}335)\,0{,}5 = 0{,}1675$$

Q4 d)



Notice that the points of our classification shifted such that we increase the ranges Belonging to P(x|B) for 5)

Q5) A) $\hat{\theta}_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{4}\,1+2+3+6 = \frac{12}{4} = 3{,}5 = 3{,}5$

Q5) B) $MSE_{\theta ML} = 1$, $Bias = 1$, $variance = 12{,}25$

Q5) C) $\hat{\theta}_{MAP} = \dfrac{4 \cdot 12{,}25 \cdot 3{,}5 \cdot 2 \cdot \left(\frac{3}{2}\right)^2}{4 \cdot 12{,}25 + \left(\frac{3}{2}\right)^2} = 2{,}83$

$\Delta\theta = \hat{\theta}_{ML} - \hat{\theta}_{MAP} = 0{,}22$

Q6) when I began working with machines to complete my Biological studies i realised two things. first, the computer is better at definite tasks and second that the process of implementing this definite knowledge is one I enjoy quite alot. That was my personal preference, I consider that the ethical problems stemming from machine learning primarily stem from the momentary inequality it produces for us. With this I mean that only the states before and after the reaching-conclusion of the field do the problems cease to exist but ~~we~~ while only a portion have aless, knowledge, use for it thes ethical issues hold merit. I exist on the positive side of the ingeuality st