Name: ....

1. State the difference between supervised learning and reinforcement learning? (5 points)

2. Why object recognition is considered a hard problem? (5 points)

3. State and briefly explain the two types of the classification problems. Use sketches to illustrate your answer. (3 points)

4. Compare histogram density estimation to kernel density estimation? No formulas are required. (10 points)

5. State the difference between feature selection and feature extraction. (10 points)

6. Discuss the main idea of the AdaBoost classifier? (10 points)

7. Write down the classifier weight equation of the AdaBoost? How to extend it to obtain multi-class classification? Explain your answers. (5 points)

8. Consider the following problem:

   Class 1 patterns: $(1 \quad 1)^T, (1.3 \quad 0.7)^T, (1.5 \quad -0.1)^T$

   Class 2 patterns: $(0 \quad -1)^T, (0.1 \quad 0.3)^T, (1.2 \quad 0.1)^T$

   a) Assume that we would like to use the nearest neighbor classifier. What would be the classification of the following pattern $(1.3 \quad 0)^T$? What would be the classification is we use K-nearest neighbor classifier where K = 3. (5 points)

   b) Compare between the results obtained in (a). (2 points)

   c) What would be the classification of the same point $(1.3 \quad 0)^T$ if we use a Bayes classifier along with a kernel density estimation. (Assume $P(C_1) = P(C_2) = 0.5$, assume an independent Gaussian kernel function and take h = 0.5). (10 points) sec2 هبة أحمد علي

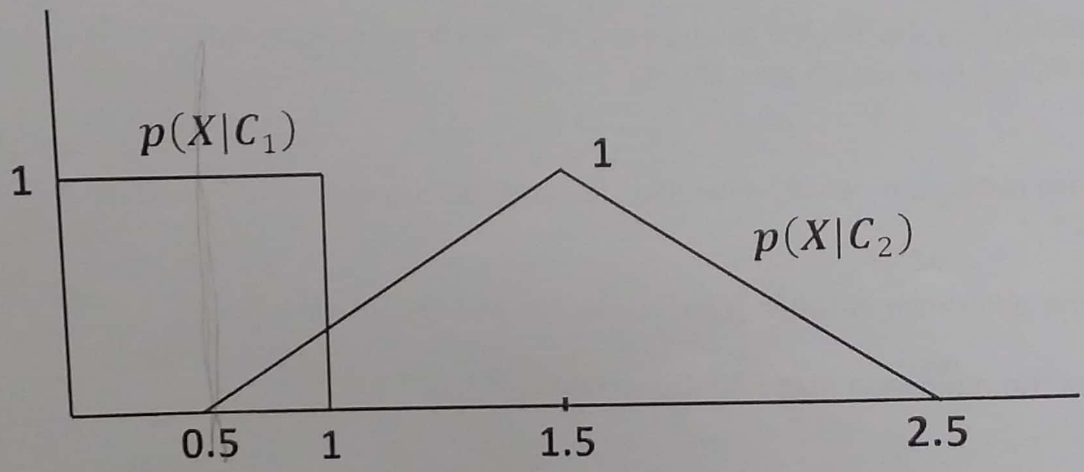
Cairo University
Faculty of Engineering
Computer Engineering Dept.

Pattern Classification
Fall 2018 - Midterm Exam

9. Consider the one-dimensional two-class classification problem, where we assume that $P(C_1) = 0.3$ and $P(C_2) = 0.7$. The class conditional densities are shown in the figure below:

a) Plot the decision regions and the decision boundaries for the Bayes classifier. (10 points)

b) Find the classification error for the Bayes classifier. (10 points)



10. Consider a two-dimensional two-class classification problem, where the class-conditional densities are given by:

$$p(\underline{X}|C_1) = 0.5\frac{1}{2\pi}e^{\frac{-((x_1+1)^2+(x_2+1)^2)}{2}} + 0.5\frac{1}{2\pi}e^{\frac{-((x_1-1)^2+(x_2-1)^2)}{2}}$$

$$p(\underline{X}|C_2) = 0.5\frac{1}{2\pi}e^{\frac{-((x_1-1)^2+x_2^2)}{2}} + 0.5\frac{1}{2\pi}e^{\frac{-(x_1^2+(x_2-1)^2)}{2}}$$

Assume that $P(C_1) = P(C_2) = 0.5$.

a) Sketch the approximate decision boundary (10 points)
b) The classification regions for the Bayes classifier (10 points)

1) Supervised learning : the goal is To learn the function that maps inputs to outputs.

We give it the labeled possible outputs To choose from. ex: classify written characters.

Reinforcement learning: here we don't give it the labeled outputs. The input is a state it takes an action and learn from it's score/penalty.

~~Ex→ when we~~ give a child a toy and see what his action would be. if he played with it give him a score. if he broke it give him a penalty.

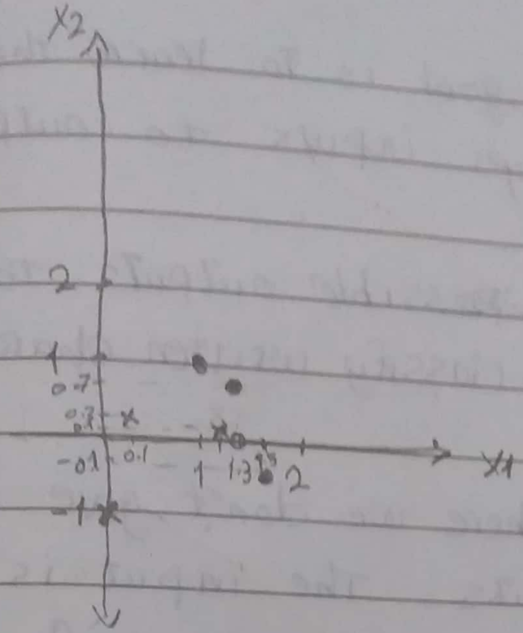
10

5

2) object recognision is considered a hard problem because of the diversity in the same object which can take different shapes (ex: when the angle changed)

5

⑧



$d = (x_{11} - x_{21})^2 + (x_{12} - x_{22})^2$

Points	$(1.3\ 0)^T$ distance	class
$(1\ 1)^T$	1.09	C_1
③ $(1.3\ 0.7)^T$	0.49	C_1
② $(1.5\ -0.1)^T$	0.05	C_1
$(0\ -1)^T$	2.69	C_2
$(0.1\ 0.3)^T$	1.53	C_2
① nearest neighbour $(1.2\ 0.1)^T$	0.02	C_2

5

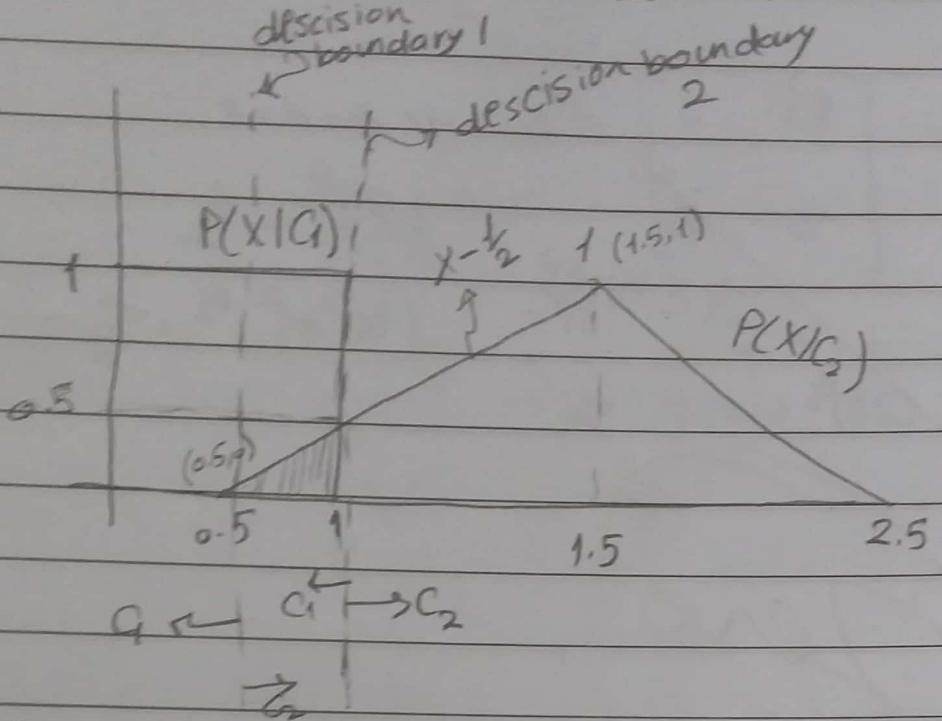
to the nearest neighbour the classification of $(1.3\ 0)^T \Rightarrow C_2$

when $k = 3 \Rightarrow C_1$ as 2 points are from C_1 and 1 from C_2

b) the second result is more accurate as the strange values won't affect it like in the ~~neighrest~~ nearest neighbour.

2

g) $P(C_1)=0.3$ $P(C_2)=0.7$



$\frac{y-0}{x-0.5} = \frac{1-0}{1} =$

$\therefore y = x - \frac{1}{2}$

at $x=1$

$y = \frac{1}{2}$

For decision 1

$$P(\text{error}) = \int_{0.5}^{\infty} P(X|C_1)\,P(C_1)\,dx = (0.5 * 1) * 0.3 = 0.15$$

For decision 2

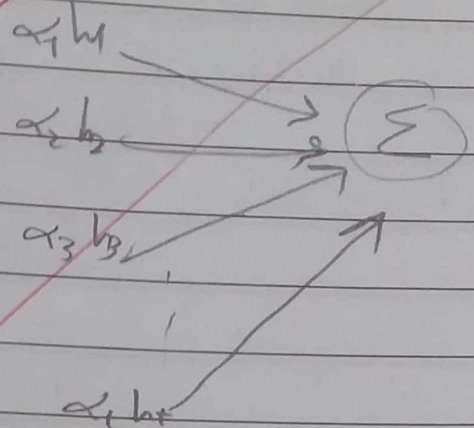
$$P(\text{error}) = \int_{-\infty}^{1} P(X|C_2)\,P(C_2)\,dx = 0.5 * 0.5 * 0.5 * 0.7 = 0.0875$$

$\therefore$ choose decision boundary 2

16

6) the main idea of adaBoost is Combine weak classifiers to get a strong classifier

it firstly choose classifier h1 which fit the data. then get the error and the wait of this classifier



and increase the weight of the points which wrongly classified by h1 then choose h2 who fit the data with the new waits and repeat to get a strong classifier from the t ones.

10

7) $\alpha_t = \log\left(\frac{1-err_t}{err_t}\right)$ $\quad err_t < 0.5$ binary

for multi $err_t = \frac{k-1}{k}$ then α_t would be $-ve$ so

to obtain multi class classification

$$\alpha_t = \log\left(\frac{1-err_t}{err_t}\right) + \log(k-1)$$

α_t will be $+ve$ only if $(1-err_t) > \frac{1}{k}$

$1 - \frac{k-1}{k}$

$\frac{k-k+1}{k}$

So for weak classifiers with $err_t >$ random err_t α_t will be $+ve$.

10

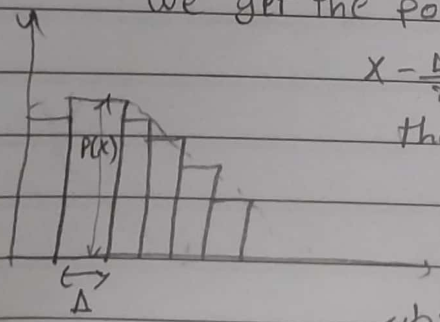
Feature selection is to select some features from which we already have and ignore other features to git rid of the problem of curse of diminsion.

9

Feature extraction is to try to exctract the features to use from the problem like no of pixels black in each column in classifing problem between ~~A~~ and I.

(4) histogram: for each Bin in the histogram we get the points that fall in range

$x - \frac{\Delta}{2}$, $x + \frac{\Delta}{2}$ we draw the histogram → say m points



$$P(x) = \frac{m}{M\,(\text{Bin size})}$$

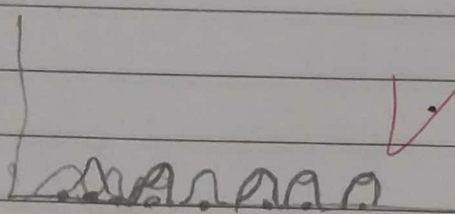
وهنا بيكون متقطع
not smooth

kernel denisty estimator

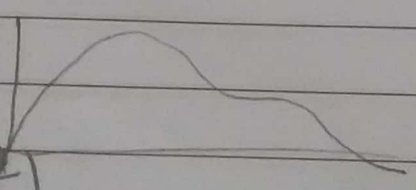
بناخذ kernel زي Gaussian لكل point الـ mean هو النقطة

بعدين نجمعهم كلهم يبين الـ probabilty denisty

وبتكون smooth



10



Summation are gaussian

$Q_h = \frac{1}{h} g\left(\frac{x}{h}\right)$

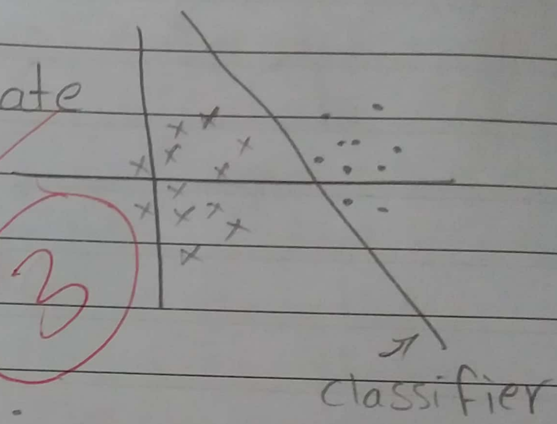
$\int g(x)\,dx = 1$ ⟵ Gaussian الـ kernel مش شرط يكون

... to maximize the prob. of correct classification.

1.5

3) linear classification:

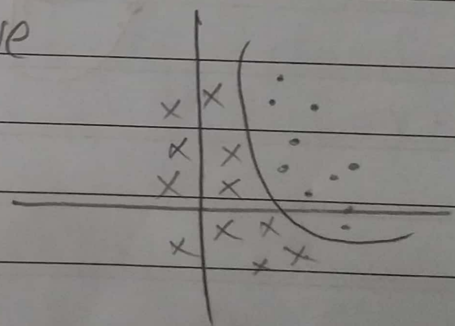
A hyperplane is appropriate (in general) for the training examples. classifing

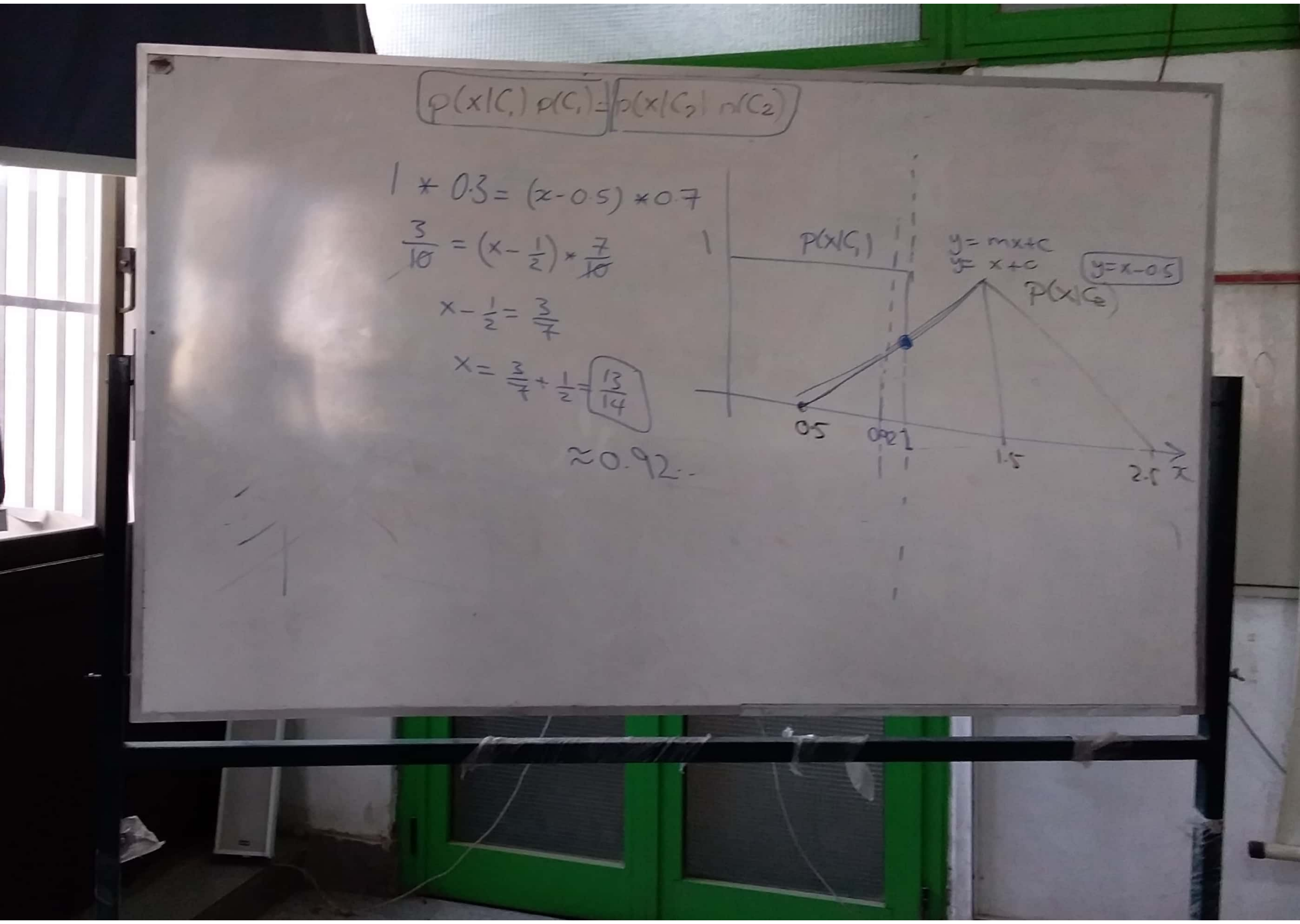


3

non linear classification:

a non-linear plane or curve is required to correctly classify the training data




p(x|C1) p(C1) = p(x|C2) p(C2)
1 * 0.3 = (x-0.5) * 0.7
3/10 = (x - 1/2) * 7/10
x - 1/2 = 3/7
x = 3/7 + 1/2 = 13/14
≈ 0.92..
P(x|C1)
y = mx+c
y = x+c
y=x-0.5
P(x|C2)
1
0.5
0.92
1
1.5
2.5
x

$$\boxed{p(x|C_1)\,p(C_1)} = \boxed{p(x|C_2)\,p(C_2)}$$

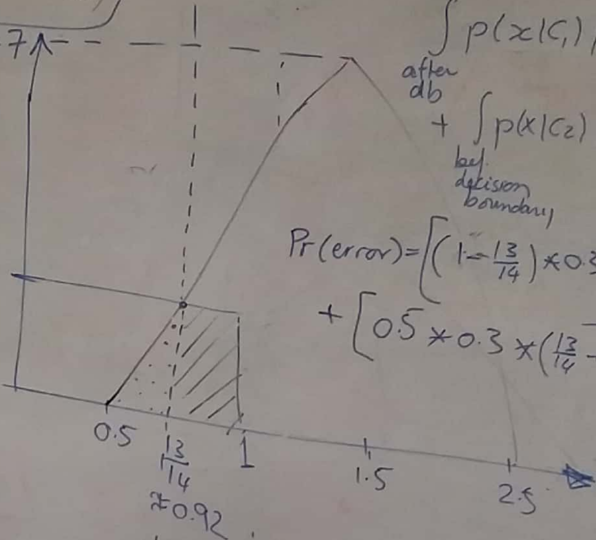
$$* 0.3 = (x - 0.5) * 0.7$$

$$\frac{3}{10} = (\;\;\;-\frac{1}{2}) * \frac{7}{\cancel{10}}$$

$$x - \;\;= \frac{3}{7}$$

$$x = \frac{3}{7} + \frac{1}{2} = \boxed{\frac{13}{14}}$$

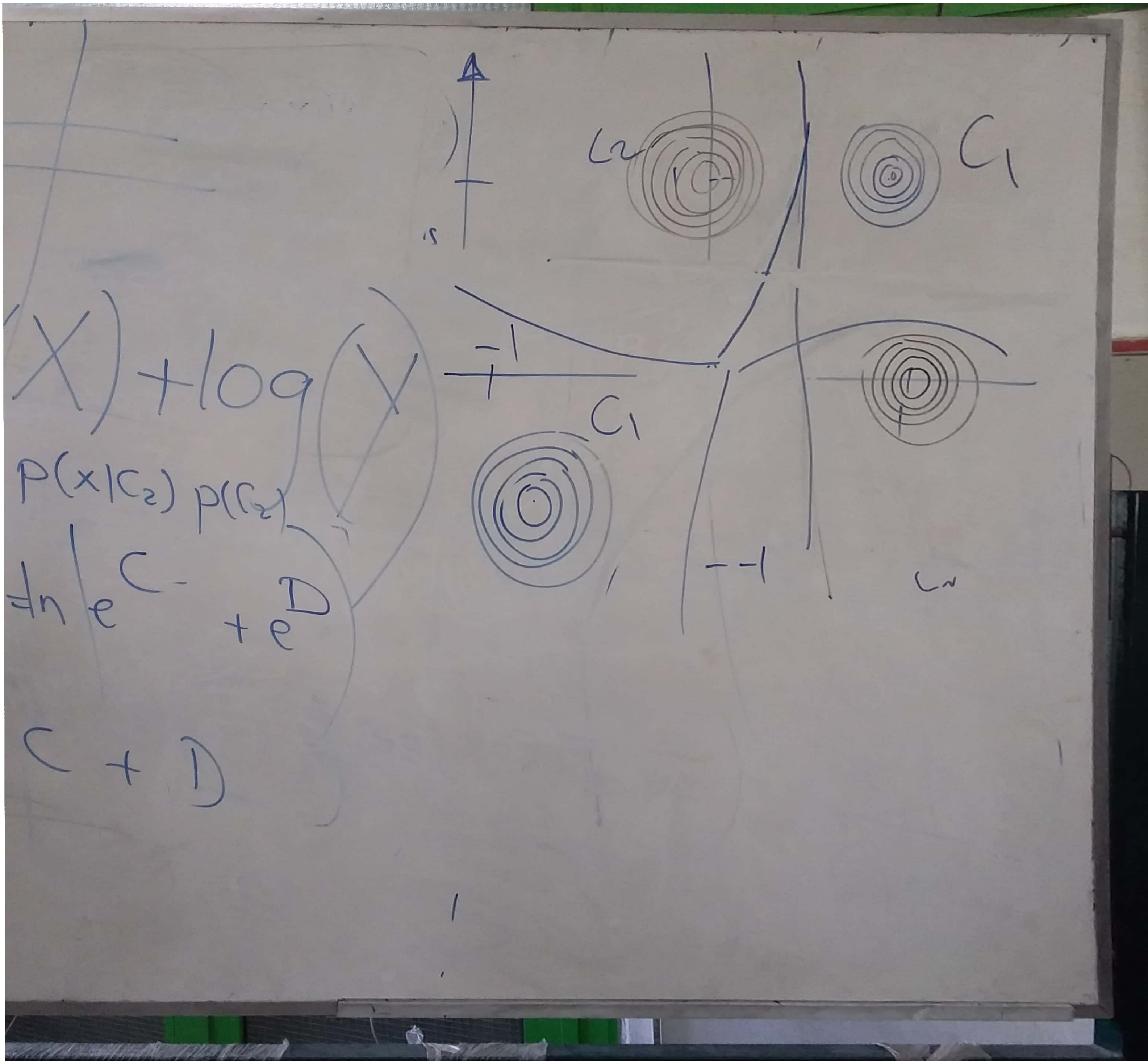
$$\approx 0.92\ldots$$



$\int p(x|C_1)\,p(C_1)$ after db

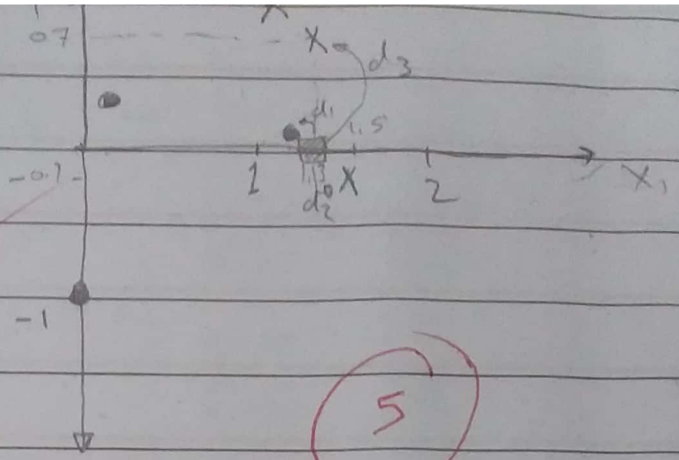
$+ \int p(x|C_2)\,p(C_2)$ bef. decision boundary

$$Pr(error) = \left[\left(1 - \frac{13}{14}\right) * 0.3\right] + \left[0.5 * 0.3 * \left(\frac{13}{14} - 0.5\right)\right]$$


X) + log (Y
p(x|C2) p(C2)
ln e^C + e^D
C + D
C2
C1
C1
-1
-1
-1

$d_2 = \sqrt{(1.5-1.3)^2 + (-0.1-0)^2}$

$= \frac{\sqrt{5}}{10}$



$d_1 < d_2$ $\boxed{\text{Class 2}}$

$d_3 = \frac{7}{10}$

@ $K=3$ (C_2, C_1, C_1) $\therefore$ $\boxed{\text{Class 1}}$

5

2

b) The K-NN seems to be more robust towards outliers (Points far from their original class and close to another class, which in case of NN, its confusing the classifi

c) max $P(C_i|X) \rightarrow$ max $P(X|C_i)P(C_i)$

$$P(\underline{X}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\underline{X}-\underline{\mu})^T \Sigma^{-1}(\underline{X}-\underline{\mu})\right)$$

indep $\rightarrow$
$$= \frac{1}{2\pi\, \sigma_1 \sigma_2} \exp\left[-\frac{1}{2}\sum_{i=1}^{2} \frac{(x_i-\mu_i)^2}{\sigma_i^2}\right]$$

$h=\sigma_i \rightarrow$
$$= \frac{1}{2\pi h^2} \exp\left[-\frac{1}{2}\sum_{i=1}^{2} \frac{(x_i-\mu_i)^2}{h^2}\right] = \frac{2}{\pi}\exp\left\{-2\sum_{i=1}^{2}(x_i-\mu_i)^2\right\}$$

We calculate the total prob. for each class by adding the Kernels of its points

8c cont.

9.5

	Point (mean)	Kernel
Class 1	1, 1	$\frac{2}{\pi}\exp\left[-\frac{1}{2}\left[\frac{(x_1-1)^2}{0.25}+\frac{(x_2-1)^2}{0.25}\right]\right] = \frac{2}{\pi}e\left[-2\left[(x_1-1)^2+(x_2-1)^2\right]\right]$
	1.3, 0.7	$\frac{2}{\pi}\exp\left[-2\left[(x_1-1.3)^2+(x_2-0.7)^2\right]\right]$
	1.5, -0.1	$\frac{2}{\pi}\exp\left[-2\left[(x_1-1.5)^2+(x_2+0.1)^2\right]\right]$
Class 2	0, -1	$\frac{2}{\pi}\exp\left[-2\left[(x_1)^2+(x_2+1)^2\right]\right]$
	0.1, 0.3	$\frac{2}{\pi}\exp\left[-2\left[(x_1-0.1)^2+(x_2-0.3)^2\right]\right]$
	1.2, 0.1	$\frac{2}{\pi}\exp\left[-2\left[(x_1-1.2)^2+(x_2-0.1)^2\right]\right]$

To classify $(1.3, 0) \rightarrow$ Sum the probabilities

$$P_{c_1} = \frac{1}{3}\left[0.072 + 0.239 + 0.576\right] = 0.887$$

$$P_{c_2} = \frac{1}{3}\left[0.0029 + 0.0298 + 0.6117\right] = 0.644$$

$\therefore P_{c_1} > P_{c_2}$ $\boxed{\text{Class 1}}$

The numbers make sense, the closer the (1.3, 0) to the point the higher the prob. that (1.3, 0) follows the point effect

↳ problem 5:

Feature selection = if we have N feature we will select from them L features where $L << N$

Feature extraction is transforming the N features to another domain & select from it L features, the problem is that those features don't have physical meaning. "the L features"

10

↳ problem 6: