QUESTION A1

(a) $0 \leq P \leq 1$

(b) $P = P_1 \cdot P_2$ $(P_1 \wedge P_2)$

(c) $P = P_1 + P_2$ $(P_1 \vee P_2)$ [3]

(d)

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$

B = HYPOTHESIS

A = DATA

P(B) = PRIOR PROBABILITY

P(A|B) = A PRIORI PROBABILITY OF DATA GIVEN HYPOTHESIS

P(A) = PROBABILITY OF DATA

$= \sum_i P(A|B_i)\,P(B_i)$ FOR i HYPOTHESES.

P(B|A) = POSTERIOR PROBABILITY [4]

(e) $P = \lim_{N\to\infty}\left(\frac{n}{N}\right)$ [1]

(f) $\int_{-\infty}^{\infty} P(x)\,dx = 1$ (PROBABILITY DENSITY FUNCTION)

$\max\left[\mathcal{L}(x)\right] = 1$

$\mathcal{L}(x)$ IS $P(x)$ NORMALISED TO SATISFY THE ABOVE CONDITION [2]

TOTAL A1 = 10 MARKS

QUESTION A2

(a) ARITHMETIC MEAN: $\bar{x} = \frac{1}{n} \sum_i x_i$ (sum over data i)

STANDARD DEVIATION (UNBIASED VARIANCE) $\sigma = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$

PEARSONS SKEW $\gamma = \frac{1}{n\sigma^3} \sum (x_i - \bar{x})^3$

COVARIANCE $\sigma_{xy} = \frac{1}{n-1} \sum_i (x_i - \bar{x})(y_i - \bar{y})$

PEARSONS CORRELATION $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

[5]

(b) $\langle x \rangle = \int_{-\infty}^{\infty} x\, P(x)\, dx$ FOR SOME PDF $P(x)$

$$\langle V \rangle = \langle x^2 \rangle - \langle x \rangle^2$$
$$= \int_{-\infty}^{\infty} x^2\, P(x)\, dx - \left[\int_{-\infty}^{\infty} x\, P(x)\, dx \right]^2$$

[2]

(c) i) BINOMIAL $\bar{x} = np$
$\sigma^2 = npq$

[1]

ii) POISSON $\bar{x} = \lambda$
$\sigma^2 = \lambda$; $\therefore \sigma = \lambda^{1/2}$

[1]

d) MEDIAN: WHEN ORDERING DATA IN TERMS OF INCREASING VALUE THE MEDIAN OF THE DATA IS THE CENTRAL VALUE (ODD NUMBER OF DATA), OR MID POINT BETWEEN THE TWO MOST CENTRAL VALUES (EVEN # OF DATA)

MODE: WHEN DISPLAYING DATA GRAPHICALLY IN A BINNED HISTOGRAM, THE MODE IS THE BIN WITH THE MOST ENTRIES.

[1]

TOTAL A2 = 10 MARKS

QUESTION A3

(a) $$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma_i} \right)^2$$

i = index, sum over n data
x_i = THE i^{th} VALUE OF x
$\bar{x}$ = MEAN OF x
σ_i = UNCERTAINTY ON i^{th} VALUE OF x. [3]

(b) ν = NUMBER OF DEGREES OF FREEDOM
$= n - n_c$

n_c: NUMBER OF CONSTRAINTS.

FOR A SAMPLE OF n DATA $\nu = n - 1$ [2]

(c) FOR $T(x) = f$ WE COMPUTE

(i) $\frac{dT}{dx}$ NUMERICALLY BY EVALUATING $T(x_0)$ AND $T(x_0 + \epsilon)$, WHERE
x_0 = INITIAL VALUE
ϵ = PREDEFINED STEP SIZE

i.e. $$\frac{dT}{dx} \simeq \frac{T(x_0+\epsilon) - T(x_0)}{\epsilon} = \frac{\Delta T}{\Delta x}$$

(ii) GIVEN THIS ESTIMATE OF THE GRADIENT ONE CAN NOW COMPUTE A NEW ESTIMATE FOR x; x_1 GIVEN BY

$$x_1 = x_0 + \delta$$
$$= x_0 + D \cdot \frac{\Delta T}{\Delta x}$$

WHERE IN GENERAL D IS A SEARCH STEP SIZE, NOT NECESSARILY ϵ. NOW IN GENERAL

$$x_{i+1} = x_i + \delta$$

(iii) STOP ITERATION OF STEP (ii) WHEN $\frac{\Delta T}{\Delta x} \simeq 0$ (WITHIN ACCEPTIBLE LIMIT) [5]

TOTAL A3 = 10 MARKS

QUESTION A4

(a) i) 63.7 %
ii) 99.7%
iii) 99.99% [2]

(b) i) yes
ii) yes
iii) no [2]

(c) FOR $f = f(x,y)$

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2 \frac{\partial f}{\partial x}\frac{\partial f}{\partial y} \sigma_{xy}$$ [2]

(d) STATISTICAL ERRORS:
UNCERTAINTY RELATED TO THE NUMBER OF DATA SAMPLED IN AN EXPERIMENT

SYSTEMATIC ERRORS:
UNCERTAINTY RELATED TO THE PROCEDURE USED TO PERFORM A MEASUREMENT. [2]

(e) BLIND ANALYSIS:
BLIND ANALYS TECHNIQUE IS A FORMAL PROCEDURE OF PERFORMING A MEASUREMENT OF SOME OBSERVABLE O, KEEPING THE TRUE VALUE OF O OBSCURED FROM THE EXPERIMENTER UNTIL THE LAST POSSIBLE MOMENT.

USEFUL TO MITIGATE EXPERIMENTER BIAS WHEN ANALYSING DATA [2]

TOTAL A4 = 10 MARKS

QUESTION A5

a) ~~normal likelihood function~~

$$L = \frac{e^{-N'}}{N!} \prod_i N_A P_A + N_B P_B \qquad \text{EXTENDED LIKELIHOOD}$$

$N' = N_A + N_B$ TOTAL FITTED SIGNAL YIELD

N = # OF DATA

P_A = PDF FOR COMPONENT A

P_B = PDF FOR COMPONENT B [2]

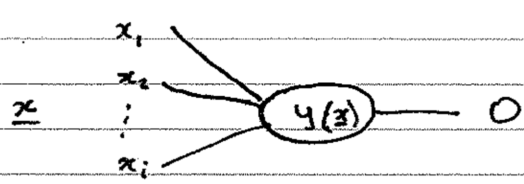
(b)
$$L = G(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$\therefore -\ln L = \underbrace{-\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)}_{\text{MINIMUM VALUE}} + \underbrace{\frac{(x-\mu)^2}{2\sigma^2}}_{\text{CHANGES WITH } x.}$$

$\therefore$ IF $x = \mu \pm 1\sigma$, THEN $-\ln L$ CHANGES BY $1/2$ FROM MINIMUM

SO IF $x = \mu \pm 3\sigma$, THEN $-\ln L$ CHANGES BY $9/2$ FROM THE MINIMUM VALUE. [3]

(c) A PERCEPTRON IS A LOGICAL UNIT HAVING ONE OR MORE INPUTS ACTED ON BY AN ACTIVATION FUNCTION y IN ORDER TO PRODUCE SOME OUTPUT O:



y IS A SURFACE IN THE HYPERSPACE OF $\underline{x}$.

POSSIBLE ACTIVATION FUNCTIONS INCLUDE

RADIAL: $y = e^{-\underline{\omega}\cdot\underline{x}}$

HYPERBOLIC TANGENT: $y = \tanh(\underline{\omega}\cdot\underline{x})$

BINARY: $y = \underline{\omega}\cdot\underline{x} + b \rightarrow O = 1$ for $y > 0$, else $O = 0$

SIGMOID $y = \dfrac{1}{1 + e^{(\underline{\omega}\cdot\underline{x} + \beta)}}$ [5]

TOTAL A5 = 10 MARKS

QUESTION B1

PART a

STEPS INVOLVED:

i) CLEARLY DEFINE THE NULL HYPOTHESIS H_0

ii) THUS DEFINE $H_1 = \overline{H_0}$.

iii) CHOOSE THE CONFIDENCE LEVEL TO ~~TEST~~ TEST H_0 AGAINST, i.e. WHAT LEVEL OF FALSE NEGATIVE AND FALSE POSITIVES IS ACCEPTIBLE FOR THE PROBLEM.

iv) MAKE THE COMPARISON H_0 TO DATA.

v) IF THE DATA ARE FOUND TO BE IN AGREEMENT WITH THE HYPOTHESIS THEN THIS CAN BE STATED WITH THE CORRESPONDING C.L. (SEE iii). OTHERWISE THE LEVEL OF DISAGREEMENT WITH THE DATA CAN BE REPORTED

[5]

(b) USING BAYES THEOREM TO DETERMINE PR. OF RAIN:

~~Know~~ 90% OF THE TIME WHEN IT RAINS, RAIN WAS FORECAST

5% OF THE TIME WHEN NO RAIN IS FORECAST, IT RAINS.

IT RAINS 50 DAYS OF THE YEAR.

PRIORS: $P(\text{RAIN}) = 50/365 = 0.137$

$P(\text{NO RAIN}) = 315/365 = 0.863$

$$P(\text{RAIN} \mid \text{RAIN FORECAST}) = \frac{P(\text{RAIN FORECAST} \mid \text{RAIN})\, P(\text{RAIN})}{P(\text{DATA})} = P_R$$

$P(\text{RAIN FORECAST} \mid \text{RAIN}) = 0.9$

$P(\text{RAIN NOT FORECAST} \mid \text{RAIN}) = 0.05$

$$P(\text{DATA}) = P(\text{RAIN FORECAST} \mid \text{RAIN})\, P(\text{RAIN}) + P(\text{RAIN NOT FORECAST} \mid \text{RAIN}) . P(\text{NO RAIN})$$

$$\therefore P_R = \frac{0.9 \times 0.137}{0.9 \times 0.137 + 0.05 \times 0.863}$$

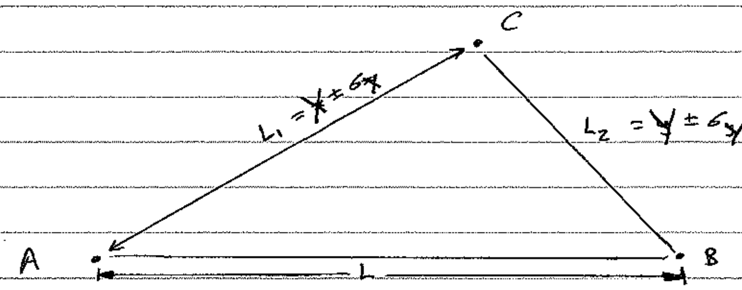
$= 0.741$

[5]

CONTD.

QUESTION B1 CONTD

c) UNSEEN



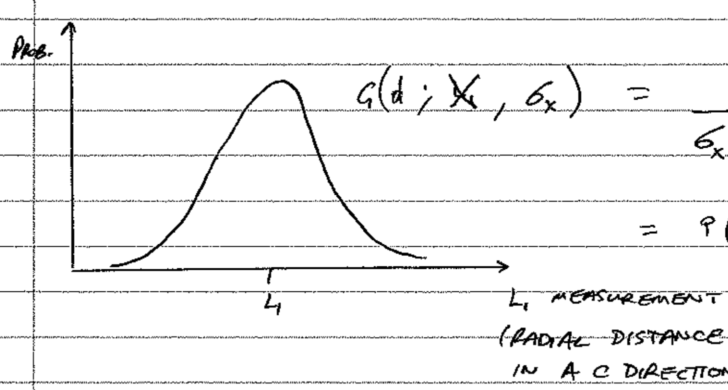
- CHOOSE COORDINATES $\hat{x}$ IS THE DIRECTION AB, & $\hat{y}$ is VERTICAL (OUT OF $\hat{x}$ TOWARD C).

- EACH MEASUREMENT L_1 AND L_2 COMPRISES A CENTRAL VALUE AND SOME UNCERTAINTY

$$L_1 = X \pm \sigma_x$$
$$L_2 = Y \pm \sigma_y$$

- ASSUME UNCERTAINTIES ARE GAUSSIAN, AND L IS SUFFICIENTLY WELL MEASURED THAT ONE CAN IGNORE ANY UNCERTAINTY ARISING FROM IT.

- CONSTRAINT FROM L_1



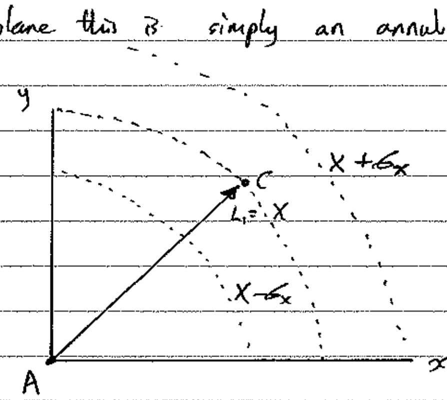
$$G(d\,;\,X,\,\sigma_x) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-(d-X)^2/2\sigma_x^2}$$

$$= P(L_1 \mid B)$$

B = hypothesis that the point (x, y) corresponds to the coordinate of C

Note $d = \sqrt{x^2 + y^2}$

In the x, y plane this is simply an annulus:

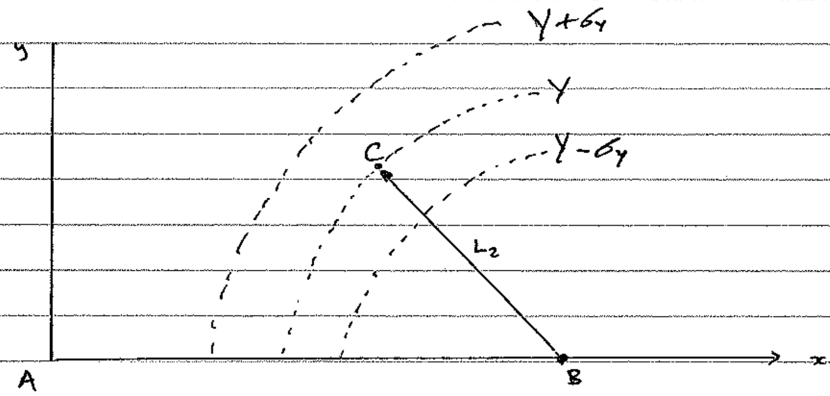


QUESTION B1 CONTD

- CONSTRAINT FROM L_2 IS CONSTRUCTED IN A SIMILAR WAY:

$$G(d'; Y, \sigma_Y) = \frac{1}{\sigma_Y\sqrt{2\pi}} e^{(d'-Y)^2/2\sigma_Y^2} = P(L_2|B)$$

HERE WE NOTE ANNULUS IS CENTRED ABOUT $x = L$ BY CONSTRUCTION, HENCE



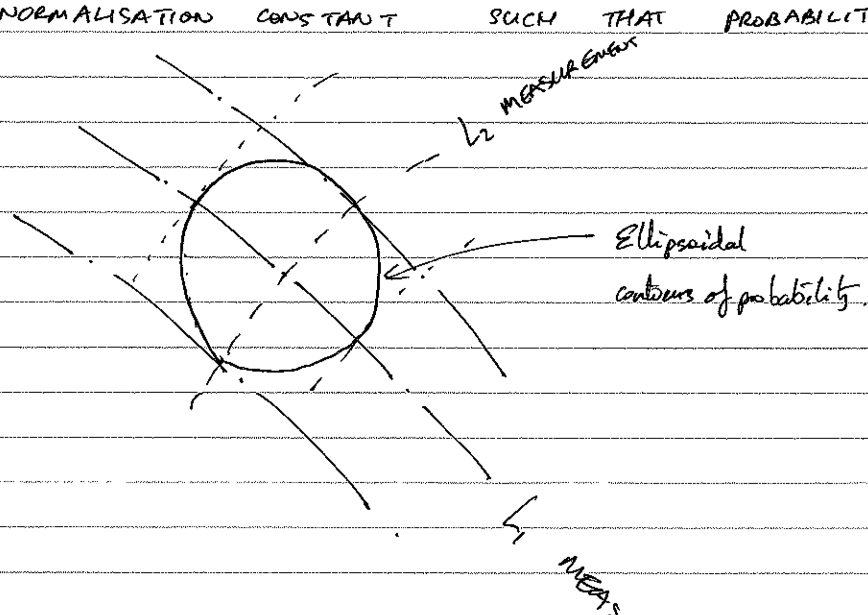
$$d' = \sqrt{(x-L)^2 + y^2}$$

BAYES THEOREM TELLS US

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$

- HERE $P(B)$ IS THE PRIOR BASED ON THE HYPOTHESIS THAT THE APEX C IS AT SOME COORDINATE (x, y). LETS ASSUME $|x| \leq \max(L_1, L_2) + 5\sigma_{L_1/L_2}$ ARE EQUI PROBABLE SOLUTIONS, AND ALL OTHER VALUES HAVE ZERO PROBABILITY.
- $P(A)$ IS A NORMALISATION CONSTANT SUCH THAT PROBABILITY SATISFIES $0 < P(B|A) \leq 1$.

COMBINATION



[15]

THERE ARE TWO SOLUTIONS (SYMMETRY IN Y ABOUT x AXIS).

TOTAL MARKS B1 = 25

QUESTION B2

(UNSEEN)

a) $x_1 = 1.0 \pm 1.5$

$x_2 = 2.0 \pm 0.7$

ESTIMATE μ & σ USING A χ^2 SCAN METHOD:

$$\chi^2 = \left(\frac{x - x_1}{\sigma_1}\right)^2 + \left(\frac{x - x_2}{\sigma_2}\right)^2$$

x	χ^2
0.0	8.61
0.25	6.5
0.5	4.7
0.75	3.2
1.0	2.0
1.25	1.2
1.5	0.6
1.75	0.38
2.0	0.44
2.25	0.82
2.5	1.51
2.75	2.51

ESTIMATE FROM SCAN THAT

$\bar{x} \sim 1.9$

$\sigma \sim 0.6$

BETWEEN THESE POINTS.

WEIGHTED AVERAGE

$$\bar{x} \pm \sigma_x = \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2} \pm \left(\sum_i \frac{1}{\sigma_i^2}\right)^{-1/2}$$

$$= 1.8 \pm 0.6$$

SCAN ESTIMATE & WEIGHTED AVERAGE ARE IN REASONABLE AGREEMENT.

[10]

QUESTION B2 CONTD.

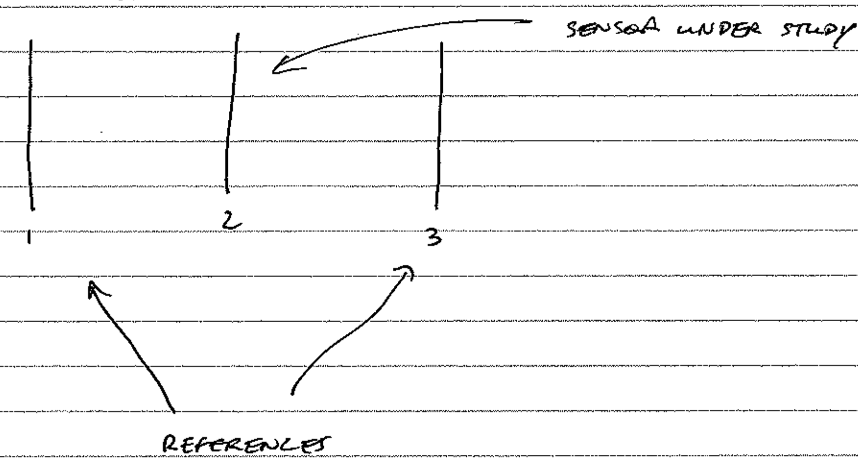
b) DETECTION EFFICIENCY $\epsilon = \frac{N_{DET}}{N_{TOT}}$ (UNSEEN)

WHERE THIS IS A BINOMIAL PROCESS HENCE

$$\sigma_\epsilon = \sqrt{\frac{\epsilon(1-\epsilon)}{N_{TOT}}}$$

(AS VARIANCE IS npq FOR A BINOMIAL DISTRIBUTION).

IN ORDER TO COMPUTE THE EFFICIENCY OF A DETECTOR ONE MUST PERFORM AN EXPERIMENT WITH SUFFICIENT REDUNDANCY TO ENSURE A PARTICLE PASSED THROUGH THE DETECTOR eg. 3 PLANAR COUNTERS



ASSUMPTION: IF PARTICLE IS DETECTED IN 1 & 3 IT MUST PASS THROUGH 2. ERGO COUNT EVENTS THAT SATISFY DETECTION CRITERIA FOR 1 & 3 & SEE HOW MANY OF THESE ARE ALSO SEEN IN 2. (n).

$$\therefore \frac{n}{N_{TOT}} = \epsilon. \quad \& \quad \sigma_\epsilon \text{ ARE CALCULABLE.}$$

$\boxed{5}$

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - a x_i^2 - b}{\sigma_i} \right]^2$$

$$= \frac{1}{\sigma^2} \sum_i \left[y_i - a x_i^2 - b \right]^2$$

$$= \frac{1}{\sigma^2} \sum_i \left(y_i^2 + a^2 x_i^4 + b^2 - 2a y_i x_i^2 - 2b y_i + 2ab x_i^2 \right)$$

$$= \frac{N}{\sigma^2} \left[\overline{y^2} + a^2 \overline{x^4} + b^2 - 2a \overline{yx^2} - 2b \overline{y} + 2ab \overline{x^2} \right]$$

Optimal soln satisfies

$$\frac{\partial \chi^2}{\partial a} = 0 \quad ; \quad \frac{\partial \chi^2}{\partial b} = 0$$

$$\therefore \quad \frac{\partial \chi^2}{\partial a} = \frac{N}{\sigma^2} \left[2a \overline{x^4} - 2\overline{yx^2} + 2b \overline{x^2} \right]$$

$$\frac{\partial \chi^2}{\partial b} = \frac{N}{\sigma^2} \left[2b - 2\overline{y} + 2a \overline{x^2} \right]$$

$$\therefore \quad a = \frac{\overline{yx^2} - \overline{y}\,\overline{x^2}}{\overline{x^4} - \left(\overline{x^2} \right)^2}$$

$$b = \overline{y} - a \overline{x^2}.$$

[10]

B2 TOTAL

25 MARKS

QUESTION B3

(a) $$F = \sum_i \alpha_i x_i + \beta$$

THE α_i ARE WEIGHTS

x_i ARE DISCRIMINATING VARIABLES

β IS AN ARBITARY OFFSET (OFTEN SET THIS TO ZERO)

$$\underline{\alpha} = \omega^{-1} (\underline{\mu}_S - \underline{\mu}_B)$$

WHERE $\underline{\mu}_S$ IS A VECTOR OF MEANS FOR A SIGNAL SAMPLE OF DATA $(\mu_S = \mu_S(\underline{x}))$.

$\underline{\mu}_B$ IS THE MEAN OF BACKGROUNDS (ALSO A VECTOR)

$\omega =$ SUM OF COVARIANCE MATRICES FOR SIGNAL & BACKGROUND i.e. $\omega = V_S + V_B$

$\therefore \omega^{-1} =$ INVERSE OF ω.

[5]

b) $\mu_A = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}$ $\quad \mu_B = \begin{pmatrix} 0.0 \\ 0.0 \end{pmatrix}$ (UNSEEN)

$V_A = \begin{pmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{pmatrix}$ $\quad V_B = \begin{pmatrix} 2.0 & 0.2 \\ 0.2 & 2.0 \end{pmatrix}$

$$\therefore \mu_A - \mu_B = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}$$

$$\omega = \begin{pmatrix} 3.0 & 0.3 \\ 0.3 & 3.0 \end{pmatrix}$$

$$\det \omega = 9 - 0.09 = 8.91$$

$$\omega^T = \begin{pmatrix} 2.0 & 0.3 \\ 0.3 & 2.0 \end{pmatrix}$$

$$C = \cancel{\omega^T} \begin{pmatrix} 2.0 & -0.3 \\ -0.3 & 2.0 \end{pmatrix}$$

$$\omega^{-1} = \frac{1}{\det \omega} \begin{pmatrix} C \end{pmatrix} = \frac{1}{8.91} \begin{pmatrix} 3.0 & -0.3 \\ -0.3 & 3.0 \end{pmatrix}$$

$$\cancel{= \begin{pmatrix} 0.512 & -0.077 \\ -0.077 & 0.512 \end{pmatrix}}$$

QUESTION B3

$$W^{-1} = \begin{pmatrix} 0.3367 & -0.0337 \\ -0.0337 & 0.3367 \end{pmatrix}$$

$$\therefore \alpha = W^{-1}(\mu_A - \mu_B)$$
$$= \begin{pmatrix} 0.3367 & -0.0337 \\ -0.0337 & 0.3367 \end{pmatrix} \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}$$

$$= \begin{pmatrix} 0.320 \\ 0.135 \end{pmatrix}$$

[10]

c) • TRAINING SAMPLE IS REQUIRED WITH $\underline{x}$ VARIABLES FOR EACH EVENT e_i TO BE USED IN THE DETERMINATION OF WEIGHTS $\underline{w}$.

• THE TRUE TARGET TYPE t MUST BE KNOWN FOR EACH EVENT.

• MIS CLASSIFICATION ERROR CAN BE ASSIGNED FOR AN EVENT:

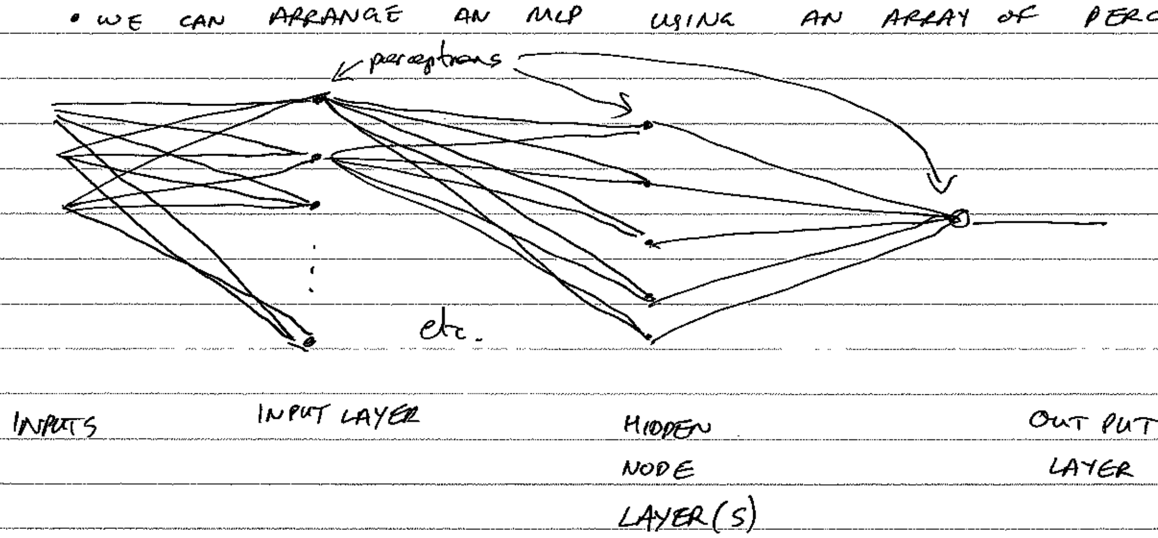
$$\epsilon_i = \frac{1}{2}(y_i - t)^2$$

WHERE y_i IS THE ACTIVATION FUNCTION OUTPUT. THE TOTAL SAMPLE ERROR IS

$$E = \sum_i \epsilon_i$$
$$= \frac{1}{2}\sum_i (y_i - t)^2$$

FOR A PERCEPTRON.

• WE CAN ARRANGE AN MLP USING AN ARRAY OF PERCEPTRONS



- GUESS AN INITIAL SET OF WEIGHTS & USE ERROR TO ESTIMATE A BETTER SET OF WEIGHTS

$$\underline{\omega}_{j+1} = \underline{\omega}_j + \Delta\underline{\omega}$$

$$\Delta\underline{\omega} \doteq -\alpha \frac{dE}{d\underline{\omega}}$$

where α is a small positive learning rate. Set $\underline{\omega}_{j+1}$ WILL HAVE A SMALLER ERROR THAN SET $\underline{\omega}$ AS:

$$\Delta E = \Delta\underline{\omega} \frac{dE}{d\underline{\omega}}$$

$$= -\alpha \left(\frac{dE}{d\underline{\omega}}\right)^2$$

- COMPUTE THE TOTAL ERROR FOR THE NETWORK AS THE SUM OF ERRORS FOR EACH PERCEPTRON. (E_{TOT})

- ITERATE UNTIL CONVERGED

- USE VALIDATION SAMPLE TO ENSURE E_{TOT} IS COMPATIBLE BETWEEN TRAINING & VALIDATION SAMPLES, AS A CREITERIA TO STOP CONVERGENCE

[10]

TOTAL MARKS FOR B 3 = 25