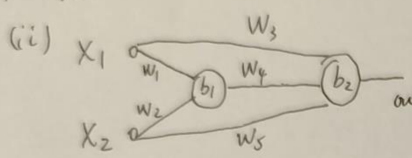
(a)

(i) f_2, f_4, f_5

In the perception, different information has different weights, and the prediction is determined by the computed result ~~[illegible]~~ versus the threshold.

For f_1, no output is 1 which can't learn

(ii)



$W_1 = W_2 = +1 = W_3 = W_5$

$W_4 = -2$

$b_1 = 1.5$

$b_2 = 0.5$

output	X_1	X_2	$W_1X_1 + W_2X_2$ (compare with b_1)	$W_4 \cdot out_{b_1}$	$W_3X_1 + W_4out_{b_1} + W_5X_2$ (compare with b_2)
0	0	0	0	0	0
1	0	1	1	0	1
1	1	0	1	0	1
0	1	1	2	-2	0

b) (i) $\{(5,3),(4,4),(6,3),(5,4),(2,3)\}$

$C_0 = \{(5,2),(4,5)\}$ $\quad O_{0,0} = (5,2)$, $O_{0,1} = (4,5)$

$d_{1,0} = \sqrt{(5-5)^2+(3-2)^2} = 1$

$d_{1,1} = \sqrt{(5-4)^2+(3-5)^2} = \sqrt{5}$

$d_{1,0} < d_{1,1}$

$\therefore (5,3)$ for $O_{0,0}$

$d_{2,0} = \sqrt{(5-4)^2+(2-4)^2} = \sqrt{5}$

$d_{2,1} = \sqrt{(4-4)^2+(4-5)^2} = 1$

$d_{2,0} > d_{2,1}$

$\therefore (4,4)$ for $O_{0,1}$

$d_{3,0} = \sqrt{(6-5)^2+(3-2)^2} = \sqrt{2}$

$d_{3,1} = \sqrt{(6-4)^2+(3-5)^2} = 2\sqrt{2}$

$d_{3,0} < d_{3,1}$

$\therefore (6,3)$ for $O_{0,0}$

$d_{4,0} = \sqrt{(5-5)^2+(2-4)^2} = 2$ $\quad d_{4,0} > d_{4,1}$

$d_{4,1} = \sqrt{(4-5)^2+(5-4)^2} = \sqrt{2}$ $\quad \therefore (5,4)$ for $O_{0,1}$

$d_{5,0} = \sqrt{(2-5)^2+(3-2)^2} = \sqrt{10}$

$d_{5,1} = \sqrt{(2-4)^2+(3-5)^2} = 2\sqrt{2}$

$d_{5,0} > d_{5,1}$

$\therefore (2,3)$ for $O_{0,1}$

$\therefore O_0 = \{(5,3),(6,3)\}$

$O_1 = \{(4,4),(5,4),(2,3)\}$

$\therefore M_1 = ((5+6)/2, (3+3)/2) = (5.5, 3)$

$M_2 = ((4+5+2)/3, (4+4+3)/3) = (\frac{11}{3}, \frac{11}{3})$

$\therefore C_1 = \{(5.5,3), (\frac{11}{3}, \frac{11}{3})\}$

$d_{1,0} = \sqrt{(5.5-5)^2+(3-3)^2} = \frac{\sqrt{2}}{2}$

$d_{1,1} = \sqrt{(\frac{11}{3}-5)^2+(\frac{11}{3}-3)^2} = \frac{2\sqrt{5}}{3}$

$d_{1,0} < d_{1,1}$ $\therefore (5,3)$ for $O_{1,0}$

...... $(4,4)$ for $O_{1,1}$

$(6,3)$ for $O_{1,0}$

$(5,4)$ for $O_{1,0}$

$(2,3)$ for $O_{1,1}$

$\therefore O_0 = \{(5,3),(6,3),(5,4)\}$

$O_1 = \{(4,4),(2,3)\}$

$\therefore M_1 = ((5+6+5)/3, (3+3+4)/3)$

$= (\frac{16}{3}, \frac{10}{3})$

$M_2 = ((4+2)/2, (4+3)/2)$

$= (3, 3.5)$

$\therefore C_2 = \{(\frac{16}{3}, \frac{10}{3}), (3, 3.5)\}$

(ii) Firstly, the large number of features requires significant computational resources to calculate the Euclidean distance, and secondly, the large K values chosen make it difficult to converge the data.

(c) (i) $n = \frac{1000 \times 0.95}{0.98} \doteq 970$

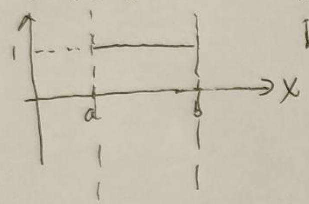
(ii)

(I) True when VC = 4, no set of 4 samples can be shattered

(II) False they may be on the same hyperplane

(III) False

(iii) VC dimension = 3



It looks like decision tree with 2 node so dimension = 2+1 = 3

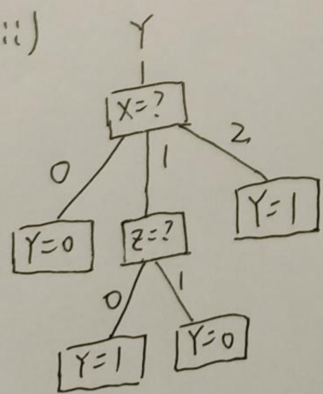
(d) (i) entropy of Y: $-\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} = 1$

(ii) entropy' $= \frac{3}{10}\left(-\frac{3}{3}\log_2\frac{3}{3}\right) - \left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right)\times\frac{4}{10} - \left(\frac{3}{3}\log_2\frac{3}{3}\right)\times\frac{3}{10}$

$= 0.4$

$\therefore$ information gain $= 1 - 0.4 = 0.6$

(iii)



~~Feathe~~

Feature W was not selected as a feature of the decision tree, indicating that it is not significant enough on its own in relation to the target Y.