For XOR data set

x_1	x_2	Y	
0	0	0	-1
0	1	1	1
1	0	1	1
1	1	0	-1

Here we are using the kernal function

for x_2 $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ $x' = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$

$K(x, x') = (x^T x' + 1)^2$

$(u_1 v_1 + u_2 v_2 + 1)^2$

$= 1 + (u_1 v_1 + 2 u_2 v_2)^2 + 2(u_1 v_1 + u_2 v_2)$

$= 1 + u_1^2 v_1^2 + u_2^2 v_2^2 + 2 u_1 v_1 u_2 v_2 + 2u_1 v_1 + 2u_2 v_2$

So the equivalent transformation

$x \Rightarrow [1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1 x_2]^T$

for $(x_1, x_2) = (0, 0)$

$\Phi(0,0) = [1, 0, 0, 0, 0, 0]^T$

$\Phi(0,1) = [1, 0, \sqrt{2}, 0, 1, 0]^T$

$\Phi(1,0) = [1, \sqrt{2}, 0, 1, 0, 0]^T$

$\Phi(1,1) = [1, \sqrt{2}, \sqrt{2}, 1, 1, \sqrt{2}]^T$

$L(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j K_{ij}$

S.t $\alpha_i y_i = 0 \Rightarrow \alpha_1 + \alpha_4 = \alpha_2 + \alpha_3$

$\alpha_i \geq 0$

inner product = K
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 4 \\ 1 & 1 & 4 & 4 \\ 1 & 4 & 4 & 9 \end{bmatrix}$$

One lagrange equation =

$$\alpha_1+\alpha_2+\alpha_3+\alpha_4 - \frac{1}{2}\left(\alpha_1^2 + 4\alpha_2^2 + 4\alpha_3^2 + 9\alpha_4^2 + \right.$$
$$-\alpha_1\alpha_2 - \alpha_1\alpha_3 + \alpha_1\alpha_4$$
$$-\alpha_1\alpha_2 - \alpha_1\alpha_3 + \alpha_1\alpha_4$$
$$\left. +\alpha_2\alpha_3 + \alpha_2\alpha_3 - 4\alpha_2\alpha_4 - 4\alpha_4\alpha_2 - 4\alpha_3\alpha_4 - 4\alpha_4\alpha_3 \right)$$

$$= \alpha_1+\alpha_2+\alpha_3+\alpha_4 + \frac{\alpha_1^2}{2} + 2\alpha_2^2 + 2\alpha_3^2 + 9\alpha_4^2 - \alpha_1\alpha_2$$
$$-\alpha_1\alpha_3 + \alpha_1\alpha_4 + \alpha_2\alpha_3 - 4\alpha_2\alpha_4 - 4\alpha_3\alpha_4$$

$$\frac{\partial L}{\partial \alpha_1} = 0 \Rightarrow \alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1 \quad (I)$$

$$\frac{\partial L}{\partial \alpha_2} = 0 \Rightarrow 4\alpha_2 - \alpha_1 + \alpha_3 - 4\alpha_4 = 1 \quad (II)$$

$$\frac{\partial L}{\partial \alpha_3} = 0 \Rightarrow 4\alpha_3 - \alpha_1 + \alpha_2 - 4\alpha_4 = 1 \quad (III)$$

$$\frac{\partial L}{\partial \alpha_4} = 0 \Rightarrow 9\alpha_4 + \alpha_1 - 4\alpha_2 - 4\alpha_3 = 1 \quad (IV)$$

from the following equations

$$\alpha_1 = \frac{13}{3} \quad \alpha_2 = \frac{8}{3} \quad \alpha_3 = \frac{8}{3} \quad \alpha_4 = 2$$

as we know

$$w = \sum_{i=1}^{4} \alpha_i y_i \Phi(x_i)$$

$$-\frac{13}{3}\begin{bmatrix}1\\0\\0\\0\\0\\0\end{bmatrix}+\frac{8}{3}\begin{bmatrix}1\\0\\\sqrt{2}\\0\\1\\0\end{bmatrix}+\frac{8}{3}\begin{bmatrix}1\\\sqrt{2}\\0\\1\\0\\0\end{bmatrix}-2\begin{bmatrix}1\\\sqrt{2}\\\sqrt{2}\\1\\1\\\sqrt{2}\end{bmatrix}$$

$$z=\begin{bmatrix}-1\\\frac{2\sqrt{2}}{3}\\\frac{2\sqrt{2}}{3}\\-2\sqrt{2}\\\frac{2}{3}\\\frac{2}{3}\end{bmatrix}$$

$$g(x_1,x_2)=w^T\phi(x)$$

$$z=\begin{bmatrix}-1 & \frac{2\sqrt{2}}{3} & \frac{2\sqrt{2}}{3} & \frac{2}{3} & \frac{2}{3} & -2\sqrt{2}\end{bmatrix}\begin{bmatrix}1\\\sqrt{2}x_1\\\sqrt{2}x_2\\x_1^2\\x_2^2\\\sqrt{2}x_1x_2\end{bmatrix}$$

- decision boundary

$$z=-1+\frac{4}{3}x_1+\frac{4}{3}x_2+\frac{2}{3}x_1^2+\frac{2}{3}x_2^2-4x_1x_2\geq 0$$

Q. Bayes classifier means classification of datapoints into 2 or more classes based on their probability of occurence of their attributes in class1 or class2.

Q3 Let N be equal to the total number of data points and N_1 be the data points for Class 1 and N_2 for Class 2

we label t as $\frac{N}{N_1}$ for C_1 and $-\frac{N}{N_2}$ for C_2

Then the sum of square error function

$$E = \frac{1}{2} \sum_{n=1}^{N} (\omega^T x_n + \omega_0 - t_n)^2$$

$$\frac{\partial E}{\partial \omega} = 0 \Rightarrow \sum_{n=1}^{N} (\omega^T x_n + \omega_0 - t_n) x_n = 0 \quad (1)$$

$$\frac{\partial E}{\partial \omega_0} = 0 \Rightarrow \sum_{n=1}^{N} (\omega^T x_n + \omega_0 - t_n) = 0 \quad (11)$$

from 11 we can compute

$$\omega_0 = -\omega^T \bar{m}$$

here $\bar{m} = \frac{1}{N} \sum_{n=1}^{N} x_n$

ie $\bar{m} = \frac{1}{N} (N_1 \bar{m}_1 + N_2 \bar{m}_2)$

Here $\bar{m}_1$ and $\bar{m}_2$ are corresponding mean of dataset belong to class C_1 and C_2

and we can see

$$\sum_{n=1}^{N} t_n = N_1 \frac{N}{N_1} - N_2 \frac{N}{N_2} = 0$$

equation (1) can also be written as

$$\left(S_W + \frac{N_1 N_2}{N} S_B\right) w = N(\bar{m}_1 - \bar{m}_2)$$

Here $S_W = \sum_{n \in C_1} (x_n - \bar{m}_1)(x_n - \bar{m}_1)^T + \sum_{n \in C_2} (x_n - \bar{m}_2)(x_n - m_2)^T$

$$S_B = (\bar{m}_2 - \bar{m}_1)(\bar{m}_2 - \bar{m}_1)^T$$

To get the direction of w we can see $S_B w$ is always in the direction of $(\bar{m}_2 - \bar{m}_1)$

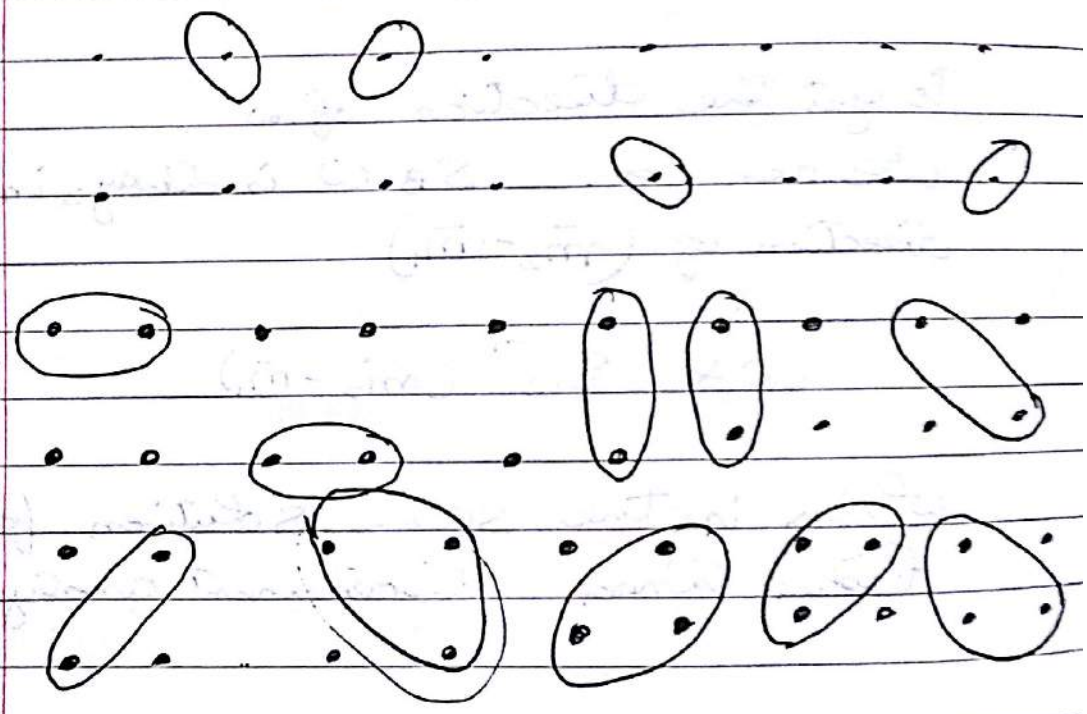
$$w \propto S_w^{-1} (\bar{m}_2 - \bar{m}_1)$$

which is the same solution for fisher linear discriminant analysis

Q5 Data points $x_1 (-1,-1)$ $x_2 (-1,+1)$
$x_3 (+1,-1)$ $x_4 (+1,+1)$

a) Linear kernal cannot shatter this

b) Polynomial kernal of degree 2 can shatter this. We take here ~~a parabola~~ an ellipse to show shattering



c Gaussian kernal can ~~of~~ shatter it