①

ASSIGNMENT 2
– ISHMEET KAUR
2015042

Ques 2

Let the only 2 data points from each of the two classes be x_1 and x_2
i.e. $x_1 \in C_1 (t_1 = 1)$, $x_2 \in C_2 (t_2 = -1)$
We are given

$$y(x_1) = w^T x_1 + b = 1$$
$$y(x_2) = w^T x_2 + b = -1$$

We have to determine the location of the minimum margin hyperplane using these.
i.e.

$$\max_{\alpha} \hat{L}(\alpha) = \max_{\alpha} \left(\sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{n} \sum_{m=1}^{m} a_n a_m t_n t_m x_n^T x_m \right)$$

$N = 2$ and $\alpha_i \geq 0$ and $\sum_{n=1}^{N} \alpha_n t_n = 0$

$\Rightarrow \alpha_1 - \alpha_2 = 0$ —— ①

$\alpha_1 \, \alpha_2 \geq 0$

$$\hat{L}(\alpha) = \alpha_1 + \alpha_2 - \frac{1}{2} \left(\alpha_1^2 t_1^2 x_1^T x_1 - \alpha_1 \alpha_2 t_1 t_2 x_1^T x_2 + \alpha_2 \alpha_1 t_2 t_1 x_2^T x_1 + \alpha_2^2 t_2^2 x_2^T x_2 \right)$$

$$= \alpha_1 + \alpha_2 - \frac{1}{2} \left(\alpha_1^2 x_1^T x_1 - 2\alpha_1 \alpha_2 x_1^T x_2 + \alpha_2^2 x_2^T x_2 \right)$$

$$= \alpha_1 + \alpha_2 - \frac{1}{2} \alpha_1^2 x_1^T x_1 + \alpha_1 \alpha_2 x_1^T x_2 - \frac{1}{2} \alpha_2^2 x_2^T x_2$$

Let kernel : $x_i^T x_j = k(x_i, x_j) = k_{ij}$

$$\hat{L}(\alpha) = \alpha_1 + \alpha_2 - \frac{1}{2} \alpha_1^2 k_{11} + \alpha_1 \alpha_2 k_{12} - \frac{1}{2} \alpha_2^2 k_{22}$$

$$\frac{\partial \hat{L}(\alpha)}{\partial \alpha_1} = 1 - \alpha_1 k_{11} + \alpha_2 k_{12} = 0.$$

(2)

$$\frac{\partial \tilde{L}(\alpha)}{\partial(\alpha_2)} = 1 - \alpha_2 k_{22} + \alpha_1 k_{12} = 0 \quad \text{——} \quad (2)$$

so, we get

$$1 - \alpha_1 k_{11} + \alpha_2 k_{12} = 1 - \alpha_2 k_{22} + \alpha_1 k_{12}$$

$$\Rightarrow \quad k_{12}(\alpha_2 - \alpha_1) = \alpha_1 k_{11} - \alpha_2 k_{22}$$

from (1), $\alpha_2 - \alpha_1 = 0$.

$$\Rightarrow \quad \alpha_1 k_{11} = \alpha_2 k_{22}$$

$$\Rightarrow \quad \alpha_1 = \frac{\alpha_2 k_{22}}{k_{11}}$$

Putting this in (1)

$$1 - \alpha_2 k_{22} + \frac{\alpha_2 k_{22} k_{12}}{k_{11}} = 0.$$

$$\Rightarrow \quad -\frac{\alpha_2 k_{22}(k_{11} - k_{22})}{k_{11}} = -1$$

$$\Rightarrow \quad \alpha_1 = \frac{1}{k_{11} - k_{22}}$$

$$w = \sum_{n=1}^{N} a_n t_n \phi(x_n) = \frac{1}{k_{11} - k_{12}} x_1 - \frac{1}{k_{22}} \frac{k_{11}}{k_{11} - k_{12}} x_2$$

$$\Rightarrow w = \frac{1}{k_{11} - k_{12}} \left(x_1 - \frac{k_{11}}{k_{22}} \cdot x_1 \right)$$

$$b = \frac{1}{N_s} \sum_{n \in S} \left(t_n - \sum_{m \in S} a_m t_m k_{nm} \right) \text{——}$$

$$\cancel{b = \frac{1}{N_S} \sum_{n \in S} \left(t_n - \sum_{m \in S} a_m t_m k_{nm} \right)}$$

$$b = \frac{1}{2} \left(1 - (a_1 k_{11} - a_2 k_{12}) - 1 - (a_1 k_{21} - a_2 k_{22}) \right)$$

$$= \frac{1}{2} \left(-a_1 k_{11} + a_2 k_{12} - a_1 k_{21} + a_2 k_{22} \right).$$

$$= \frac{1}{2(k_{11} - k_{12})} \cdot (k_{11} - k_{21}) = 0 = b$$

Scanned by CamScanner

we have found the location of the maximum margin hyperplane with only 2 data points. Since a general dimension x_1, x_2 is used, it is also irrespective of the dimensionality of the data space.

4) XOR:

A	B	output
0	0	0
0	1	1
1	0	1
1	1	0

$\Longrightarrow$ inputs into vectors

(1, 1)

(−1, +1)

(1, −1)

(−1, −1)

Input vector	output
(−1, −1)	−1
(−1, 1)	1
(1, −1)	1
(1, 1)	−1

take a polynomial kernel of degree 2

ie. $k(x, x_i) = (x^T x_i + 1)^2$

this kernel corresponds to the mapping

$\phi(x) = [1 \; x_1^2 \; \sqrt{2}x_1 x_2 \; x_2^2 \; \sqrt{2}x_1 \; \sqrt{2}x_2]^T$

maximise :-

$L_D(\alpha) = \sum_{i=1}^{4} \alpha_i - \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4} \alpha_i \alpha_j z_i z_j (x_i^t x_j + 1)^2$

$0 \leq \alpha_i$

$\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0$

$\Rightarrow L_D(\alpha) = \sum_{i=1}^{4} \alpha_i - \frac{1}{2}\alpha^t H \alpha$

$\alpha = [\alpha_1 \alpha_2 \alpha_3 \alpha_4]^t$

$$H = \begin{bmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{bmatrix}$$

4

objective function for dual form :-

$$Q(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}(9\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 + 2\alpha_1\alpha_4 + 9\alpha_2^2 + 2\alpha_2\alpha_3 - 2\alpha_2\alpha_4 + 9\alpha_3^2 + 9\alpha_4^2 - 2\alpha_3\alpha_4)$$

we get (optimizing above eqn)

$$9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1 \quad \text{—— (1)}$$

$$-\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4 = 1 \quad \text{—— (2)}$$

$$-\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1 \quad \text{—— (3)}$$

$$\alpha_4 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1 \quad \text{—— (4)}$$

optimized values for langrange multipliers = $\alpha_{01}, \alpha_{02}, \alpha_{03}, \alpha_{04}$

from (1), (2), (3), (4)

$$\alpha_{01} = \alpha_{02} = \alpha_{03} = \alpha_{04} = \frac{1}{8}$$

All input vectors are support vectors

$$Q_0(\alpha) = 1/4$$

$$\frac{1}{2}\|w_0\|^2 = \frac{1}{4} \Rightarrow w_0 = 1/\sqrt{2}$$

Optimum weight vector, $w_0 = \frac{1}{8}\left[-\phi(x_1) + \phi(x_2) + \phi(x_3) - \phi(x_4)\right]$

$$w_0 = \frac{1}{8}\left(-\begin{bmatrix}1\\1\\\sqrt{2}\\1\\-\sqrt{2}\\-\sqrt{2}\end{bmatrix} + \begin{bmatrix}1\\1\\-\sqrt{2}\\1\\-\sqrt{2}\\\sqrt{2}\end{bmatrix} + \begin{bmatrix}1\\1\\-\sqrt{2}\\1\\\sqrt{2}\\-\sqrt{2}\end{bmatrix} - \begin{bmatrix}1\\1\\\sqrt{2}\\1\\\sqrt{2}\\\sqrt{2}\end{bmatrix}\right)$$

$$w_0 = \begin{bmatrix}0\\0\\-1/\sqrt{2}\\0\\0\\0\end{bmatrix}$$ hyperplane : $w_0^T\phi(x) = 0$

$\because b = 0$

$$w_0^T \phi(x) = \begin{bmatrix} 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ \sqrt{2}x_1x_2 \\ x_2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \end{bmatrix} = 0$$

$$w_0^T \phi(x) = 0 + 0 - x_1x_2 + 0 + 0 + 0$$
$$= -x_1x_2 = 0.$$

for

$$\left.\begin{array}{l} x_1 = x_2 = -1 \\ x_1 = x_2 = 1 \end{array}\right\} \; y = -1$$
$$\left.x_1 \neq x_2 \right\} \text{——} \; y = 1.$$

Hence, the XOR problem is solved

ques 1. This is because apart from maximising the margin the smoothness of the kernel function affects the complexity of the classifier. This in turn affects the risk of overfitting ~~the complexitye of the classifier~~

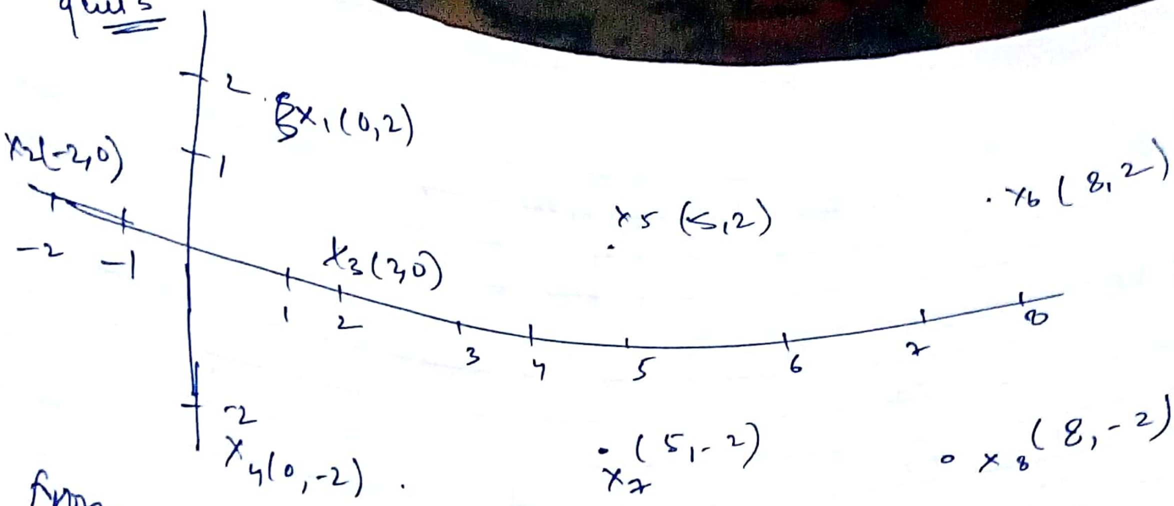
If scale factor = small ⟶ linear SVM.

scale factor = high ⟶ ~~slu~~ overfitting. even after maximising margin

Depending on kernel parameters, we can control the overfitting. the best parameters can be determined using grid search.

(6)

ques 3



from graph, support vectors are $x = 2$ & $x = 5$

decision boundary : $\frac{2+5}{2} = 3.5$

this is equidistant from both sides,

$\therefore$ Max. margin $= \frac{|5-2|}{\cdot} = 3$.

Remove $x_7(5,-2)$. $\Rightarrow$ boundary becomes slant

eqn of line passing through x_5 and x_8 :

L_1 :

$$\frac{y-2}{x-5} = \frac{-2-2}{8-5} = -\frac{4}{3}.$$

$$\Rightarrow y = -\frac{4}{3}x + \frac{26}{3}$$

eqn of line passing through x_3, parallel to L_1 :

$$\frac{y-0}{x-2} = -\frac{4}{3} \Rightarrow y = -\frac{4}{3} + 8/3.$$

distance b/w these 2 parallel lines $= \frac{|c_1 - c_2|}{\sqrt{1+m^2}}$

7

$c_1 = \frac{26}{3}$ $c_2 = 8/3$ $m = -4/3$.

$$d = \frac{|26/3 - 8/3|}{\sqrt{1 + \left(\frac{-4}{3}\right)^2}} = \frac{6}{1.66} = 3.6$$

$\therefore$ on removal of point, maximum margin becomes 3.6