

Q1
(a.1) $D = \{(x_1, x_1^2), (x_2, x_2^2)\}$
 $h(x) = ax + b$

$$x_1^2 = ax_1 + b$$

$$x_2^2 = ax_2 + b$$

$$a = \frac{x_1^2 - x_2^2}{x_1 - x_2} = x_1 + x_2$$

$$\text{Substituting, } b = -x_1 x_2$$

$$\text{So, } g(x) = (x_1 + x_2)x - x_1 x_2$$

(a.2)

$$\begin{aligned}\bar{g}(x) &= \int_{-1}^{+1} \int_{-1}^{+1} [(x_1 + x_2)x - x_1 x_2] dx_1 dx_2 \\ &= \int_{-1}^{+1} \left(x_1 + x_2 \right) \frac{x^2}{2} - x_1 x_2 x \Big|_{-1}^{+1} dx_2 \\ &= \int_{-1}^{+1} \left(\frac{(x_1 + x_2)}{2} - x_1 x_2 - \left(\frac{x_1 + x_2}{2} + x_1 x_2 \right) \right) dx_1 \\ &= 0\end{aligned}$$

$$\begin{aligned}
 (d) \quad E_{\text{out}} &= E_x [(g(x) - f(x))^2] \\
 &= E_x [(ax+b - x^2)^2] \\
 &= E_x [x^4] - 2a E_x [x^3] + (a^2 - 2b) E_x [x^2] \\
 &\quad + 2ab E_x [x] + b^2 \\
 &= \frac{1}{2} \int_{-1}^1 x^4 dx - 2a \cdot \frac{1}{2} \int_{-1}^1 x^3 dx + (a^2 - 2b) \frac{1}{2} \int_{-1}^1 x^2 dx \\
 &\quad + 2ab \cdot \frac{1}{2} \int_{-1}^1 x dx + b^2 \\
 &= \frac{1}{5} + \frac{(a^2 - 2b)}{3} + b^2
 \end{aligned}$$

Now,

$$\begin{aligned}
 E_D [E_{\text{out}}] &= \frac{1}{5} + \frac{1}{3} E_D [a^2 - 2b] + E_D [b^2] \\
 &= \frac{1}{5} + \frac{1}{3} E_D [(x_1 + x_2)^2 + 2x_1 x_2] + E_D [x_1^2 x_2^2] \\
 &= \frac{1}{5} + \frac{1}{3} \cdot \frac{1}{4} \int_{-1}^1 \int_{-1}^1 (x_1^2 + x_2^2 + 2x_1 x_2) dx_1 dx_2 \\
 &\quad + \frac{1}{3} \int_{-1}^1 \int_{-1}^1 x_1^2 x_2^2 dx_1 dx_2 \\
 &= \frac{1}{5} + \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{8}{3} + \frac{1}{3} \cdot \frac{4}{9} \\
 &= \frac{8}{15} \\
 &= 0.53
 \end{aligned}$$

$$\begin{aligned}\text{Bias}(x) &= (\bar{g}(x) - f(x))^2 \\ &= f(x)^2 \\ &= x^4\end{aligned}$$

$$\begin{aligned}\text{Bias} &= E_x(x^4) \\ &= \frac{1}{2} \int_{-1}^1 x^4 dx = \frac{1}{5} = 0.2\end{aligned}$$

$$\begin{aligned}\text{Var}(x) &= E_D[(g(x) - \bar{g}(x))^2] \\ &= E_D[a^2x^2 + 2abx + b^2] \\ &= E_D[a^2] \cdot x^2 + 2E_D[ab]x + E_D[b^2] \\ &= E_D[(x_1+x_2)^2] \cdot x^2 - 2E_D[(x_1+x_2)x_1x_2] \cdot x \\ &\quad + E_D[x_1^2x_2^2] \\ &= \frac{1}{4} \int_{-1}^1 \int_{-1}^1 (x_1^2 + 2x_1x_2 + x_2^2) dx_1 dx_2 \cdot x^2 \\ &\quad - \frac{2}{9} \int_{-1}^1 \int_{-1}^1 (x_1^2x_2 + x_1x_2^2) dx_1 dx_2 \cdot x \\ &\quad + \frac{1}{4} \int_{-1}^1 \int_{-1}^1 x_1^2x_2^2 dx_1 dx_2 \\ &= \frac{1}{4} \left(\frac{5}{3} + 0 + \frac{1}{3} \right) \cdot x^2 - 0 \cdot x + \frac{1}{4} \cdot \frac{4}{9} \\ &= \frac{2}{3}x^2 + \frac{1}{9}\end{aligned}$$

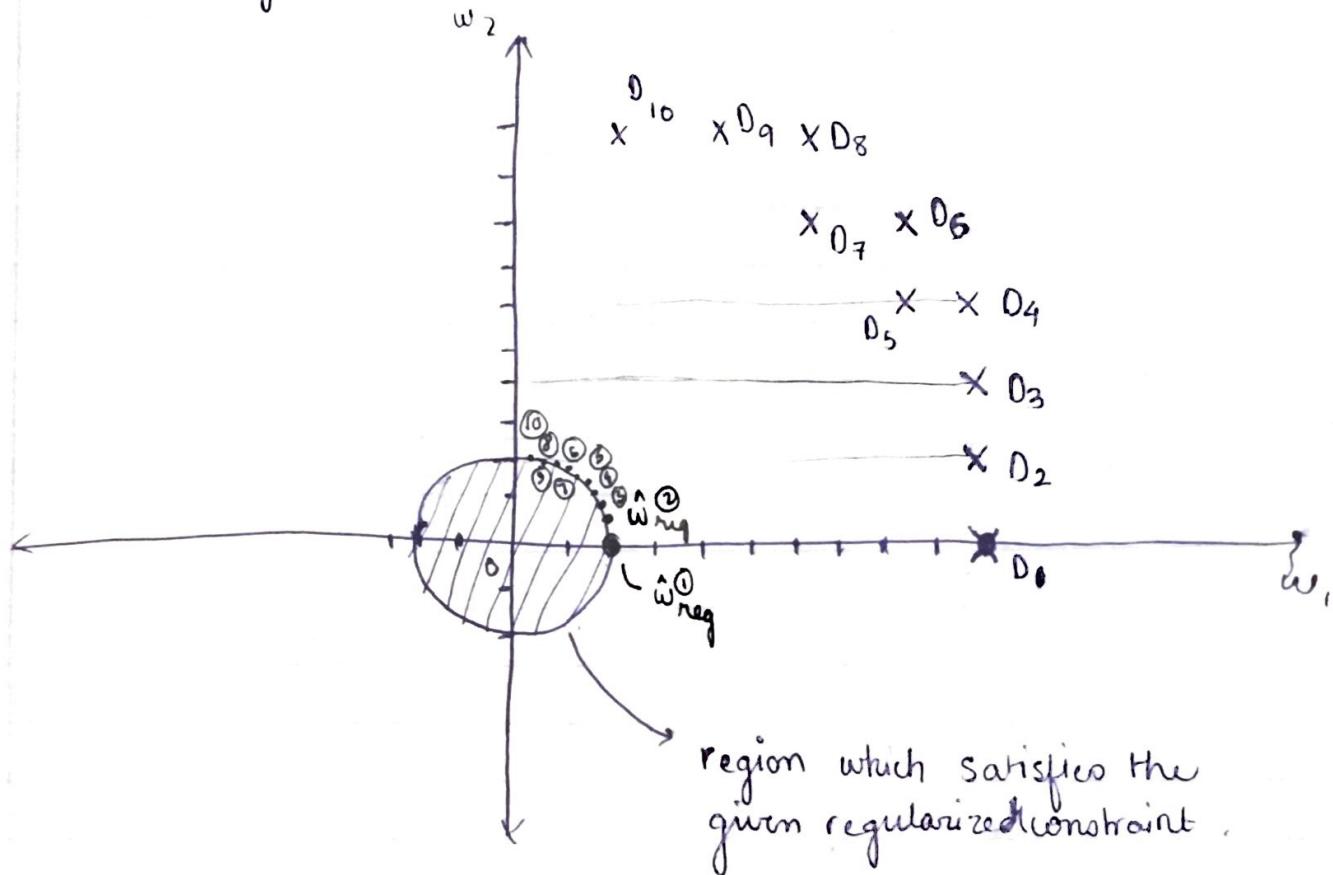
$$\therefore \text{Var}(x) = \frac{2}{3}x^2 + \frac{1}{9}$$

So,

$$\begin{aligned}\text{var} &= E_x \left[\frac{2}{3} x^2 + \frac{1}{9} \right] \\ &= \frac{2}{3} \cdot \frac{1}{2} \int_{-1}^1 x^2 dx + \frac{1}{9} \\ &= \frac{1}{3} \\ &= 0.33\end{aligned}$$

We see that the analytical results are very close to the numerically obtained ones.

Q2.

(a) L2 regularization : $\mathcal{L}(\underline{\omega}) = \|\underline{\omega}\|_2^2$, $C = 2^2$ 

- for $D_1 : (10, 0)$
- for $D_2 : (10, 2)$

$$\hat{\underline{\omega}}_{\text{reg}}^{(1)} : (2, 0)$$

$$\frac{2-y}{2} = \frac{10-x}{10}$$

$$\therefore 5y = x$$

$$x^2 + y^2 = 4$$

$$\therefore 26y^2 = 4$$

$$\therefore y = \sqrt{\frac{2}{13}}, x = 5\sqrt{\frac{2}{13}}$$

So,

$$\hat{\underline{\omega}}_{\text{reg}}^{(2)} = \left(5\sqrt{\frac{2}{13}}, \sqrt{\frac{2}{13}} \right)$$

Here, we compare the slope of the line joining the origin and centre. And substitute the value into the equation of circle to get the points $\hat{\underline{\omega}}_{\text{lin}}$.

- for $D_3 : (10, 4)$

$$\frac{4-y}{4} = \frac{10-x}{10}$$

$$\therefore x = \frac{5y}{2}$$

$$x^2 + y^2 = 4$$

$$\therefore 25y^2 = 16$$

$$\therefore y = \frac{4}{\sqrt{29}}, x = \frac{10}{\sqrt{29}}$$

So,

$$\hat{\omega}_{\text{reg}}^{(3)} = \left(\frac{10}{\sqrt{29}}, \frac{4}{\sqrt{29}} \right)$$

- for $D_4 : (10, 6)$

$$\frac{6-y}{6} = \frac{10-x}{10}$$

$$\therefore x = \frac{5y}{3}$$

$$x^2 + y^2 = 4$$

$$\therefore 25y^2 = 36$$

$$\therefore y = 3\sqrt{\frac{2}{17}}, x = 5\sqrt{\frac{2}{17}}$$

So,

$$\hat{\omega}_{\text{reg}}^{(4)} = \left(5\sqrt{\frac{2}{17}}, 3\sqrt{\frac{2}{17}} \right)$$

- for $D_5 : (8, 6)$

$$\frac{6-y}{6} = \frac{8-x}{8}$$

$$\therefore x = \frac{4y}{3}$$

$$\hat{\omega}_{\text{reg}}^{(5)} = \left(\frac{8}{5}, \frac{6}{5} \right)$$

$$x^2 + y^2 = 4$$

$$\therefore 25y^2 = 36$$

$$\therefore y = \frac{6}{5}, x = \frac{8}{5}$$

Similarly,

- for $D_6 : (8, 8)$
- for $D_7 : (6, 8)$
- for $D_8 : (6, 10)$
- for $D_9 : (4, 10)$
- for $D_{10} : (2, 10)$

$$\hat{\omega}_{\text{reg}}^{(6)} = (\sqrt{2}, \sqrt{2})$$

$$\hat{\omega}_{\text{reg}}^{(7)} = \left(\frac{6}{5}, \frac{8}{5} \right)$$

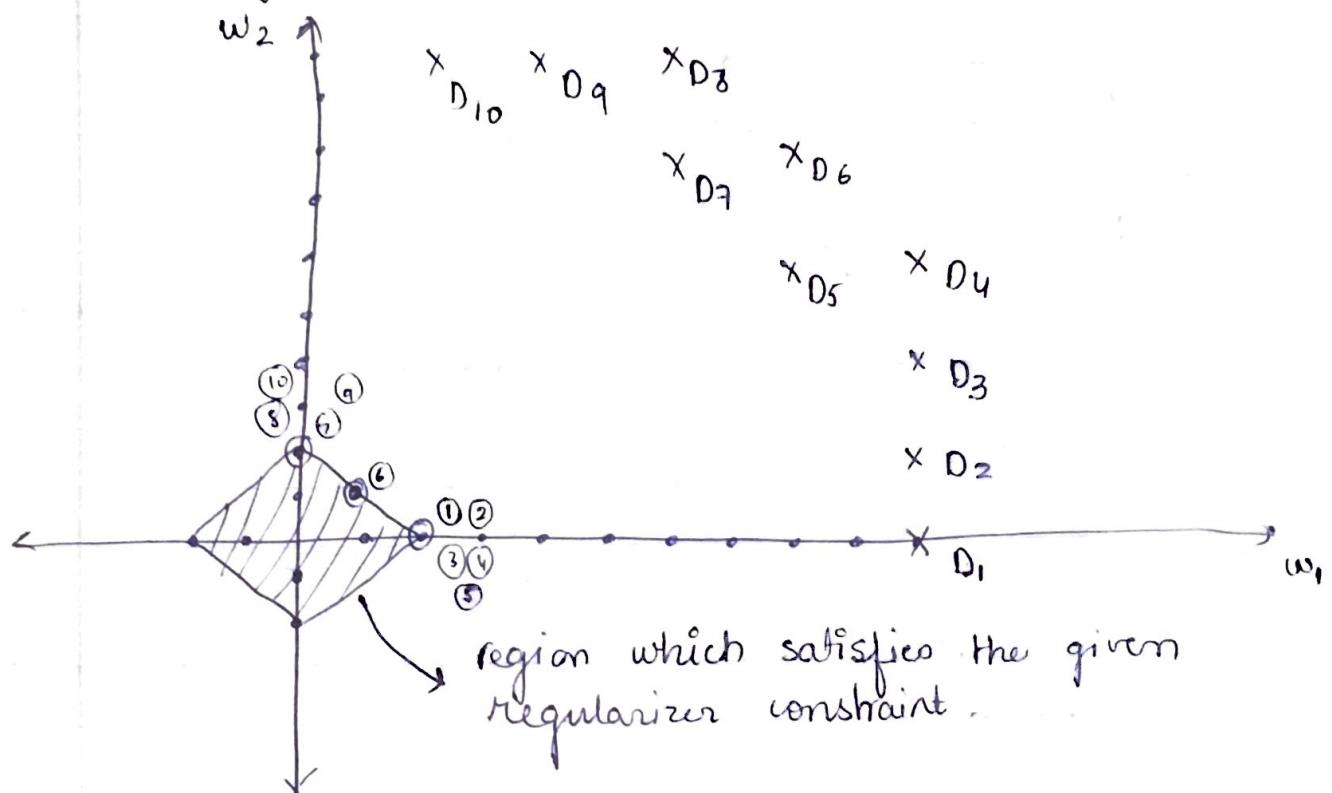
$$\hat{\omega}_{\text{reg}}^{(8)} = \left(3\sqrt{\frac{2}{17}}, 5\sqrt{\frac{2}{17}} \right)$$

$$\hat{\omega}_{\text{reg}}^{(9)} = \left(\frac{4}{\sqrt{29}}, \frac{10}{\sqrt{29}} \right)$$

$$\hat{\omega}_{\text{reg}}^{(10)} = \left(\sqrt{\frac{2}{13}}, 5\sqrt{\frac{2}{13}} \right)$$

Only $\hat{w}_{\text{reg}}^{(1)}$ is sparse (i.e., sparse in case of D_1). This is same as before. So, none are more sparse.

(b) L1 regularization : $\mathcal{L}(\underline{w}) = \|\underline{w}\|_1$, $C = 2$.



- for $D_1 : (10, 0)$

$$\hat{w}_{\text{reg}}^{(1)} = (2, 0)$$

- for $D_2 : (10, 2)$

$$\hat{w}_{\text{reg}}^{(2)} = (2, 0)$$

This is because, for L1 norm regularizer, w_{reg} is the corner point of the area

When we draw circles, it first reaches the point $(2, 0)$. Hence that's the \hat{w}_{reg} .

- for $D_3 : (10, 4)$

$$\hat{w}_{\text{reg}}^{(3)} = (2, 0)$$

- for $D_4 : (10, 6)$

$$\hat{w}_{\text{reg}}^{(4)} = (2, 0)$$

- for $D_5 : (8, 6)$

$$\hat{w}_{\text{reg}}^{(5)} = (2, 0)$$

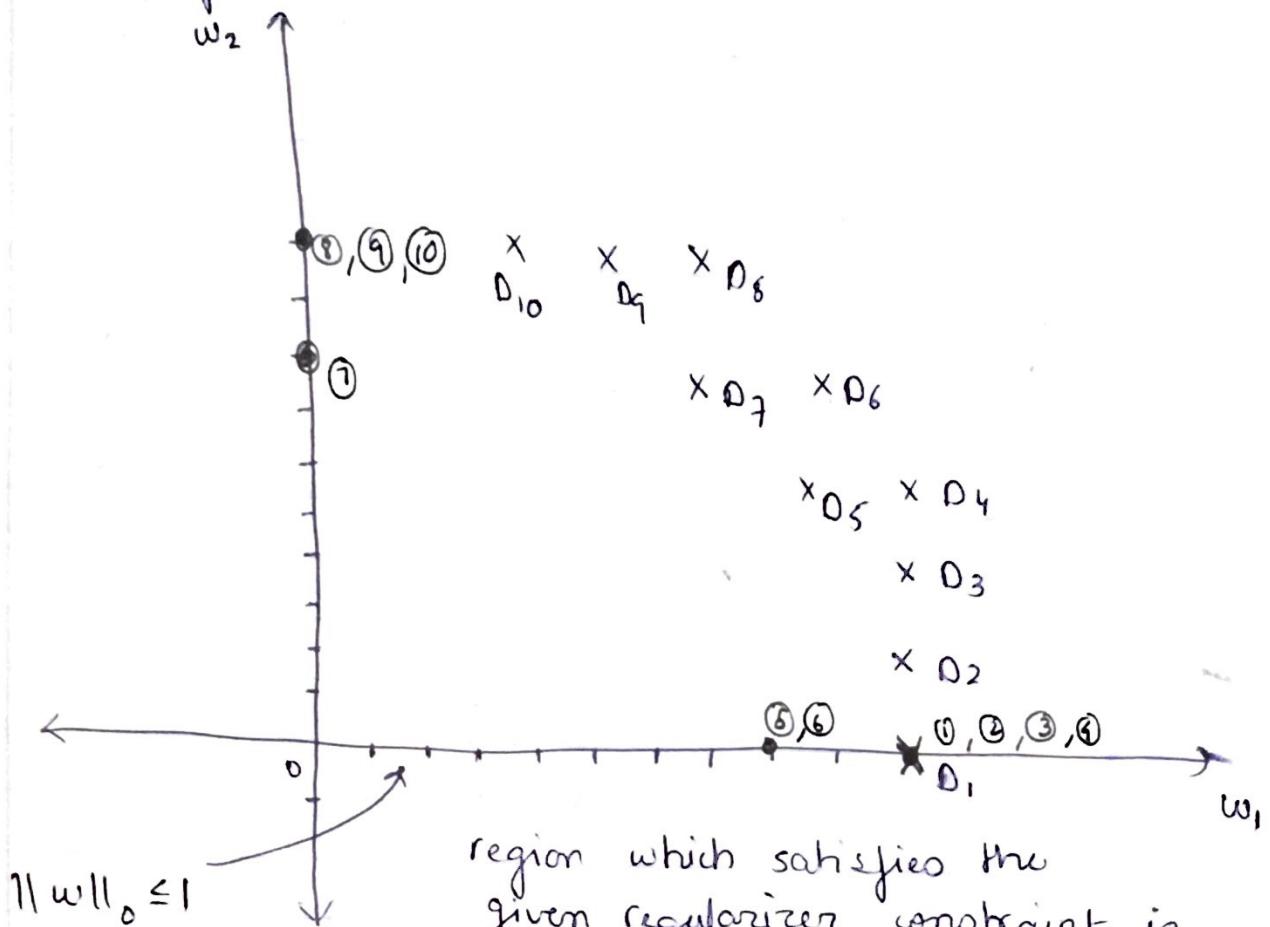
- for D6 : (8,8) $\hat{w}_{reg}^{(6)} = (1,1)$
- for D7 : (6,8) $\hat{w}_{reg}^{(7)} = (0,2)$
- for D8 : (6,10) $\hat{w}_{reg}^{(8)} = (0,2)$
- for D9 : (4,10) $\hat{w}_{reg}^{(9)} = (0,2)$
- for D10 : (2,10) $\hat{w}_{reg}^{(10)} = (0,2)$

Here, D1 was already sparse for \hat{w}_{reg}
 Now, $\hat{w}_{reg}^{(1)}, \hat{w}_{reg}^{(3)}, \hat{w}_{reg}^{(4)}, \hat{w}_{reg}^{(5)}$,
 all are at (2,0) so sparse.

Similarly, $\hat{w}_{reg}^{(7)}, \hat{w}_{reg}^{(8)}, \hat{w}_{reg}^{(9)}, \hat{w}_{reg}^{(10)}$
 are all at (0,2) and hence sparse.

So, eight values become sparse from previous case.

(c) L^0 regularization : $\Omega(\underline{w}) = \|\underline{w}\|_0$, $C = 1$



$$\|\underline{w}\|_0 \leq 1$$

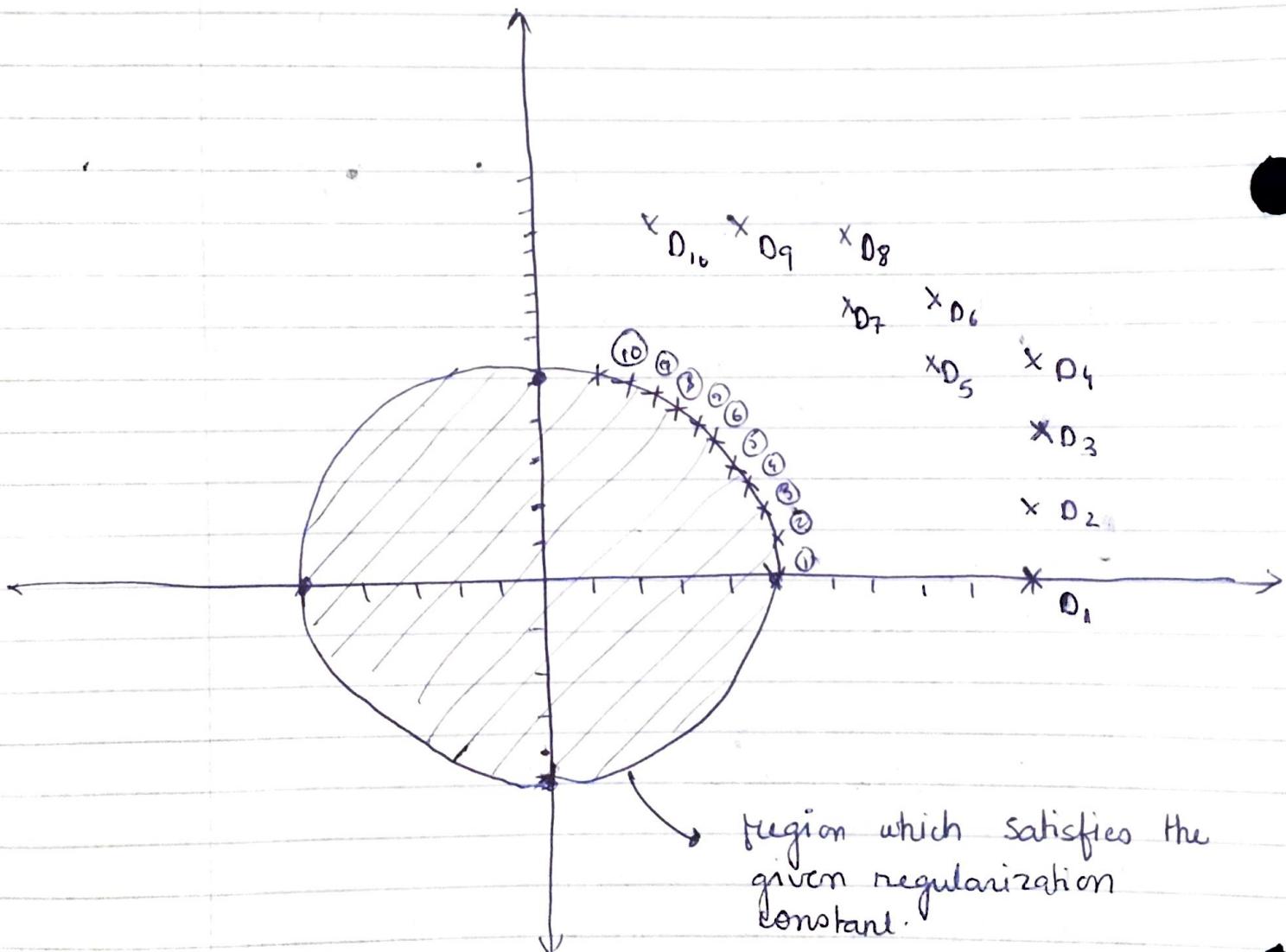
region which satisfies the given regularizer constraint is either x or y axis, given whichever axis comes in contact with the expanding circle first.

- For $D_1 : (10, 0)$ $\hat{\underline{w}}_{\text{reg}}^{(1)} = \hat{\underline{w}}_{\text{lin}} = (10, 0)$
It is regularized.
- For $D_2 : (10, 2)$ $\hat{\underline{w}}_{\text{reg}}^{(2)} = (10, 0)$
As it's the nearest axis, by calculating distance from each axis.
- For $D_3 : (10, 4)$ $\hat{\underline{w}}_{\text{reg}}^{(3)} = (10, 0)$
- For $D_4 : (10, 6)$ $\hat{\underline{w}}_{\text{reg}}^{(4)} = (10, 0)$
- For $D_5 : (8, 6)$ $\hat{\underline{w}}_{\text{reg}}^{(5)} = (8, 0)$
- For $D_6 : (8, 4)$ $\hat{\underline{w}}_{\text{reg}}^{(6)} = (8, 0)$ or $(0, 8)$
It is equidistant from both axis, so can take any value of the two.

- for D_7 : $(6, 8)$ $\hat{w}_{\text{reg}}^{(1)} = (0, 8)$
- For D_8 : $(6, 10)$ $\hat{w}_{\text{reg}}^{(2)} = (0, 10)$
- for D_9 : $(4, 10)$ $\hat{w}_{\text{reg}}^{(3)} = (0, 10)$
- for D_{10} : $(2, 10)$ $\hat{w}_{\text{reg}}^{(4)} = (0, 10)$

Here all the \hat{w}_{reg} are sparse in nature compared to the corresponding \hat{w}_{lin} - So, 9 of them are more sparse.

(d) L2 regularization : $\mathcal{R}(\underline{w}) = \|\underline{w}\|_2$, $C = 5^2$



- for D_1 : $(10, 0)$ $\hat{w}_{\text{reg}}^{(1)} = (5, 0)$
we use the same steps as used in (a).

- For $D_2 : (10, 2)$ $\hat{w}_{\text{reg}}^{(2)} = \left(\frac{5}{\sqrt{26}}, \frac{25}{\sqrt{26}}\right)$

We use $x^2 + y^2 = 25$ instead of $x^2 + y^2 = 9$

- For $D_3 : (10, 1)$ $\hat{w}_{\text{reg}}^{(3)} = \left(\frac{25}{\sqrt{29}}, \frac{10}{\sqrt{29}}\right)$

- For $D_4 : (10, 1)$ $\hat{w}_{\text{reg}}^{(4)} = \left(\frac{25}{\sqrt{34}}, \frac{15}{\sqrt{34}}\right)$

- For $D_5 : (8, 6)$ $\hat{w}_{\text{reg}}^{(5)} = (4, 3)$

- For $D_6 : (8, 8)$ $\hat{w}_{\text{reg}}^{(6)} = \cancel{\left(\frac{5}{\sqrt{2}}, \frac{5}{\sqrt{2}}\right)}$

- For $D_7 : (6, 8)$ $\hat{w}_{\text{reg}}^{(7)} = (3, 4)$

- For $D_8 : (6, 10)$ $\hat{w}_{\text{reg}}^{(8)} = \left(\frac{15}{\sqrt{34}}, \frac{25}{\sqrt{34}}\right)$

- For $D_9 : (4, 10)$ $\hat{w}_{\text{reg}}^{(9)} = \left(\frac{10}{\sqrt{29}}, \frac{25}{\sqrt{29}}\right)$

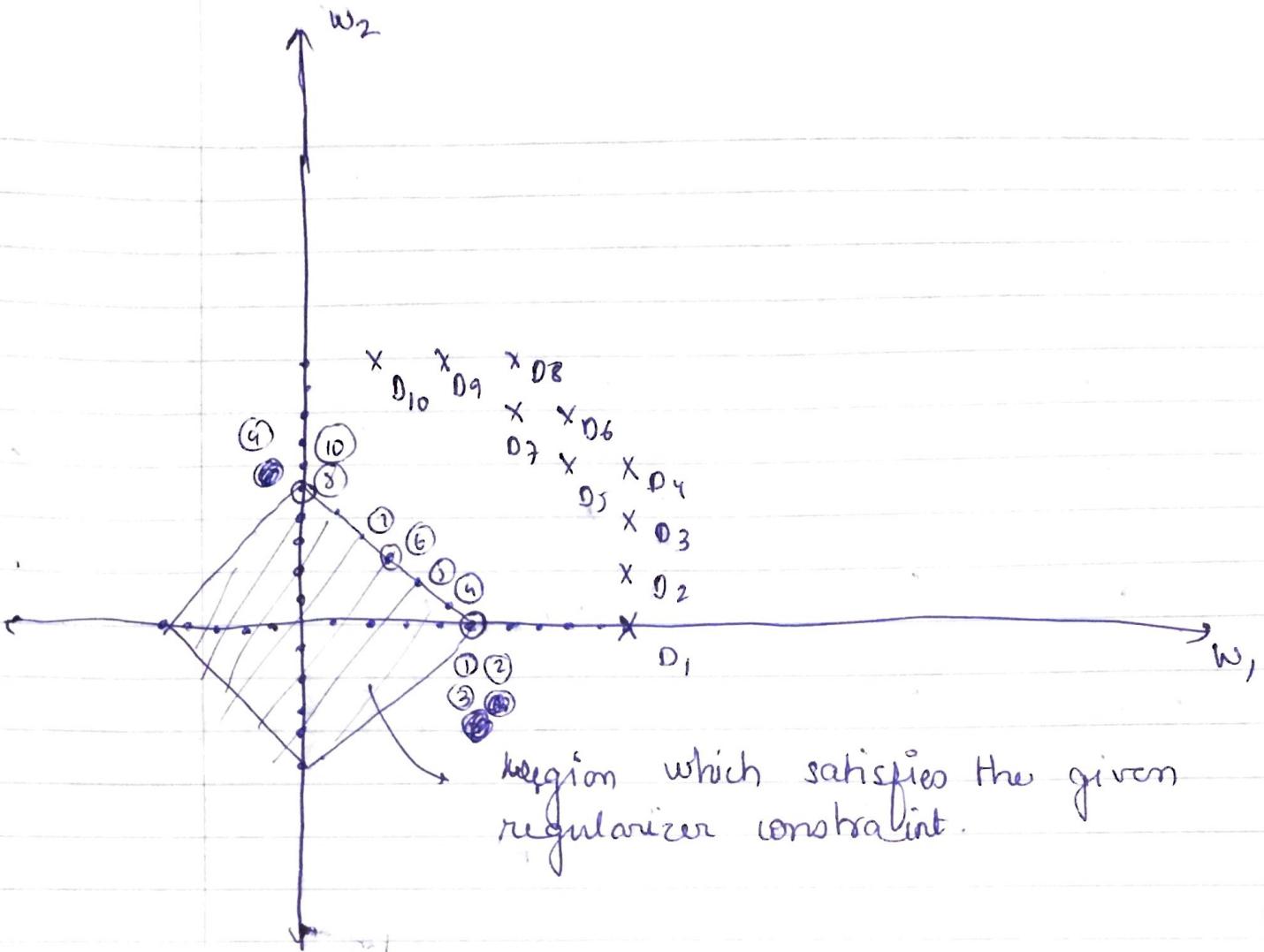
- For $D_{10} : (2, 10)$ $\hat{w}_{\text{reg}}^{(10)} = \left(\frac{25}{\sqrt{26}}, \frac{5}{\sqrt{26}}\right)$

No extra sparsity.

(e) L1 regularization : $\mathcal{R}(\underline{w}) = \|\underline{w}\|_1$, $c = 5$

Here it's the same case as before seen in (b).

The region which satisfies the given regularizer constraint is similar to (b).



- For D_1 : $(10, 0)$
- For D_2 : $(10, 2)$
- For D_3 : $(10, 4)$
- For D_4 : $(10, 6)$
- For D_5 : $(8, 6)$

$$\begin{aligned}\hat{w}_{\text{reg}}^{\textcircled{1}} &= (5, 0) \\ \hat{w}_{\text{reg}}^{\textcircled{2}} &= (5, 0) \\ \hat{w}_{\text{reg}}^{\textcircled{3}} &= \cancel{(5, 5)} \quad \cancel{(5, 5)} \quad \cancel{(5, 5)} \quad (5, 0) \\ \hat{w}_{\text{reg}}^{\textcircled{4}} &= \cancel{(5, 5)} \quad \cancel{(5, 5)} \quad \cancel{(5, 5)} \quad (4.5, 0.5) \\ \hat{w}_{\text{reg}}^{\textcircled{5}} &= \cancel{(5, 5)} \quad (3.5, 1.5)\end{aligned}$$

- for D_6 : $(8, 8)$
- for D_7 : $(6, 8)$
- for D_8 : $(6, 10)$
- for D_9 : $(4, 10)$
- for D_{10} : $(2, 10)$

$$\begin{aligned}\hat{w}_{\text{reg}}^{\textcircled{6}} &= (2.5, 2.5) \\ \hat{w}_{\text{reg}}^{\textcircled{7}} &= \cancel{(2, 8)} \quad (1.5, 3.5) \\ \hat{w}_{\text{reg}}^{\textcircled{8}} &= (0, 5) \quad 4.5 \\ \hat{w}_{\text{reg}}^{\textcircled{9}} &= (0, 5) \\ \hat{w}_{\text{reg}}^{\textcircled{10}} &= (0, 5)\end{aligned}$$

Here, $\hat{w}_{\text{reg}}^{\textcircled{1}}$ was already sparse.
 Now, all the rest except for $\hat{w}_{\text{reg}}^{\textcircled{1}}$ are sparse
 So, four $\hat{w}_{\text{reg}}^{\textcircled{2-10}}$ values become sparse from previous case.

$$(f) - D_1 : (10, 0)$$

$$(a) \hat{w}_{reg} = (2, 0)$$

$$\text{Distance} = 8$$

$$(b) \hat{w}_{reg} = (2, 0)$$

$$\text{Distance} = 8$$

$$(c) \hat{w}_{reg} = (10, 0)$$

$$\text{Distance} = 0$$

$$(d) \hat{w}_{reg} = (5, 0)$$

$$\text{Distance} = 5$$

Between (a) - (c)

smallest distance
is given by

(c)

The distance for both
cases is same.

$$\hat{w}_{reg} = (5, 0)$$

$$\text{Distance} = 5$$

$$- D_4 : (10, 6)$$

$$(a) \hat{w}_{reg} = \left(5\sqrt{\frac{2}{17}}, 3\sqrt{\frac{2}{17}} \right)$$

$$\text{Distance} = 9.66$$

$$(b) \hat{w}_{reg} = (2, 0)$$

$$\text{Distance} = 10$$

$$(c) \hat{w}_{reg} = (10, 0)$$

$$\text{Distance} = 6$$

Among (a) - (c)

smallest distance is given
by (c).

$$(d) \hat{w}_{reg} = \left(\frac{25}{\sqrt{34}}, \frac{15}{\sqrt{34}} \right)$$

$$\text{Distance} = 6.662$$

$$(e) \hat{\omega}_{reg} = (0.5, 0.5)$$

Among (d) and (e),
(e) is small.

$$\text{Distance} = 7.778$$

- $D_6 : (8, 8)$

$$(a) \hat{\omega}_{reg} = (\sqrt{2}, \sqrt{2})$$

$$\text{distance} = 9.313$$

Among (a) - (c),
(c) is smallest
distance.

$$(c) \hat{\omega}_{reg} = (8, 0)$$

$$\text{distance} = 8$$

$$(d) \hat{\omega}_{reg} = \left(\frac{5}{\sqrt{2}}, \frac{5}{\sqrt{2}} \right)$$

Among (d) and (e),
(d) is small.

$$\text{distance} = 6.313$$

$$(e) \hat{\omega}_{reg} = (2.5, 2.5)$$

$$\text{distance} = 7.778$$