Q1

(a.1) $D = \{(x_1, x_1^2), (x_2, x_2^2)\}$

$h(x) = ax + b$

$x_1^2 = ax_1 + b$

$x_2^2 = ax_2 + b$

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

substituting, $b = -x_1 x_2$

so, $g(x) = (x_1 + x_2)x - x_1 x_2$

(a.2)

$$\bar{g}(x) = \int_{-1}^{+1}\int_{-1}^{+1} \left[(x_1+x_2)x - x_1x_2\right] dx_1\, dx_2$$

$$= \int_{-1}^{+1} (x_1+x_2)\frac{x^2}{2} - x_1x_2x \Big|_{-1}^{+1} dx_2$$

$$= \int_{-1}^{+1} \left(\frac{(x_1+x_2)}{2} - x_1x_2 - \frac{(x_1+x_2)}{2} + x_1x_2\right) dx_1$$

$$= 0$$

Q1

b) The experiment to run is as shown below:

```
# -*- coding: utf-8 -*-
"""
Created on Fri Oct 19 08:38:45 2018

@author: tchat
"""

import numpy as np
import matplotlib.pyplot as plt

# Generate dataset
x1 = np.random.uniform(-1,1,1000)
x2 = np.random.uniform(-1,1,1000)
y1 = x1**2
y2 = x2**2

a = x1+x2
b = -x1*x2

# Plot g_bar vs f
for i in range(0,1000):
    x = [x1[i],x2[i]]
    y = [y1[i],y2[i]]
    plt.plot(x,y,'g')

x_value = np.arange(-1,1,0.01)
y_value = x_value**2
```

```
plt.plot(x_value,y_value,'r')
plt.title('Plot of g_(x) and f(x) together')
plt.xlabel('Input values between -1 and 1')
plt.ylabel('Output values of the function')

# Calculate g_bar
x = np.random.uniform(-1,1,1000)

a_gbar = np.mean(x1+x2)
b_gbar = -np.mean(x1)*np.mean(x2)
g_bar = a_gbar *x + b_gbar

plt.plot(x,g_bar, 'b')

# Calculate bias
f_x = x**2
bias = np.mean((g_bar - f_x)**2)
print('Value of Bias is : ', bias)

# Calculate variance
g_x = a*x + b
var = np.mean((g_x - g_bar)**2)
print('Value of Variance is : ', var)

# Calculate Eout
eout = np.mean((g_x - f_x)**2)
print('Value of Eout is : ', eout)
print('Value of E[Eout] is : ', np.mean(eout))
```

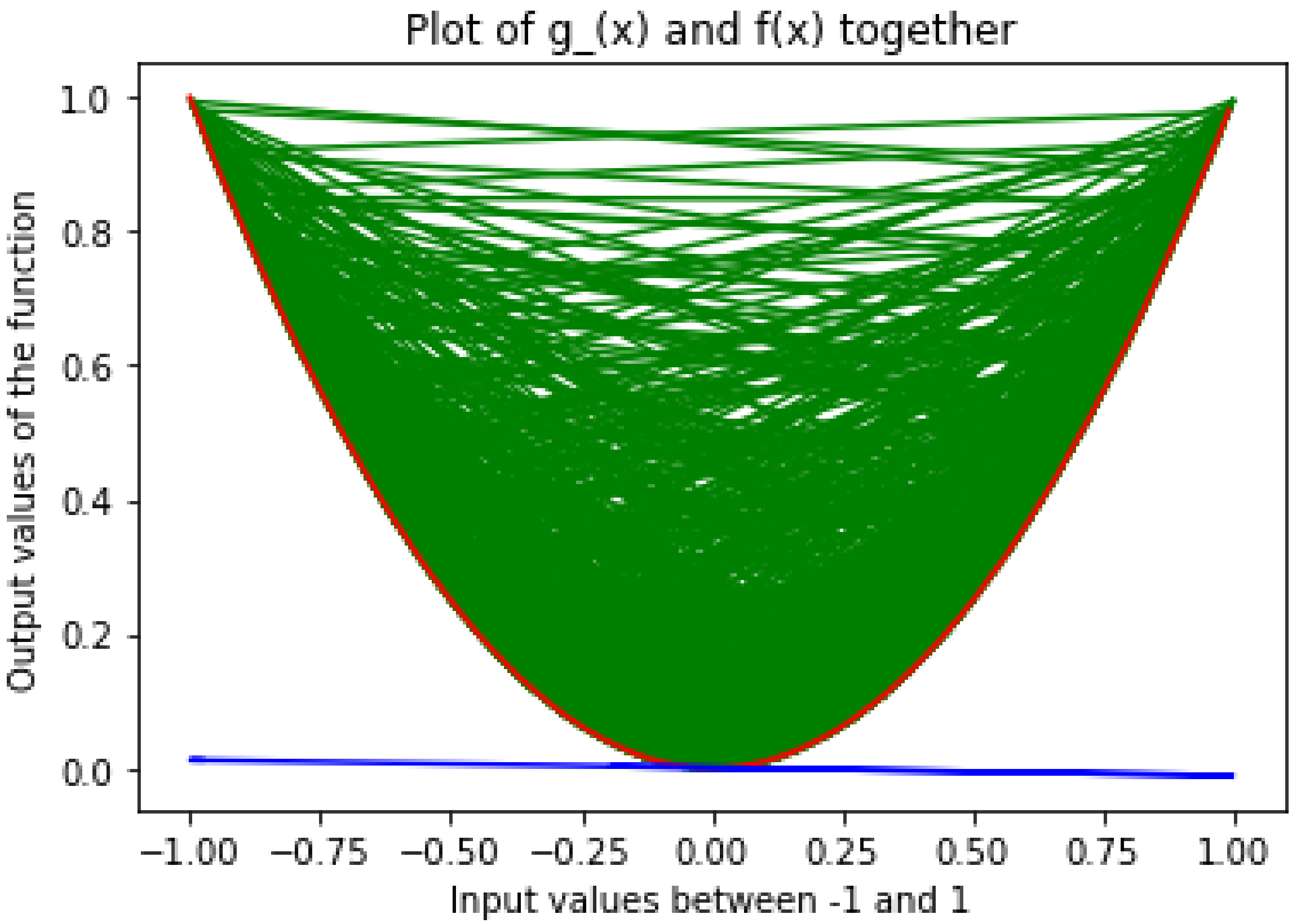
c) The results of the experiment are:

Value of Bias is : 0.184952445711

Value of Variance is : 0.338417840272

Value of Eout is : 0.540673960157

Value of E[Eout] is : 0.540673960157



As we can see from our results,

Expected value of Eout is almost equal to the value of variance + bias experimentally.

(d) $E_{out} = E_x[(g(x) - f(x))^2]$

$= E_x[(ax+b-x^2)^2]$

$= E_x[x^4] - 2a\,E_x[x^3] + (a^2-2b)\,E_x[x^2] + 2ab\,E_x[x] + b^2$

$= \frac{1}{2}\int_{-1}^{1} x^4\,dx - 2a\cdot\frac{1}{2}\int_{-1}^{1} x^3\,dx + (a^2-2b)\frac{1}{2}\int_{-1}^{1} x^2\,dx + 2ab\cdot\frac{1}{2}\int_{-1}^{1} x\,dx + b^2$

$= \frac{1}{5} + \frac{(a^2-2b)}{3} + b^2$

Now,

$E_D[E_{out}] = \frac{1}{5} + \frac{1}{3}E_D[a^2-2b] + E_D[b^2]$

$= \frac{1}{5} + \frac{1}{3}E_D[(x_1+x_2)^2 + 2x_1x_2] + E_D[x_1^2x_2^2]$

$= \frac{1}{5} + \frac{1}{3}\cdot\frac{1}{4}\int_{-1}^{1}\int_{-1}^{1}(x_1^2+x_2^2+4x_1x_2)\,dx_1\,dx_2 + \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} x_1^2x_2^2\,dx_1\,dx_2$

$= \frac{1}{5} + \frac{1}{3}\cdot\frac{1}{4}\cdot\frac{8}{3} + \frac{1}{4}\cdot\frac{4}{9}$

$= \frac{8}{15}$

$= 0.53$

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

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

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

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

So,

$$\text{var} = E_x\left[\frac{2}{3}x^2 + \frac{1}{9}\right]$$

$$= \frac{2}{3}\cdot\frac{1}{2}\int_{-1}^{1} x^2\,dx + \frac{1}{9}$$

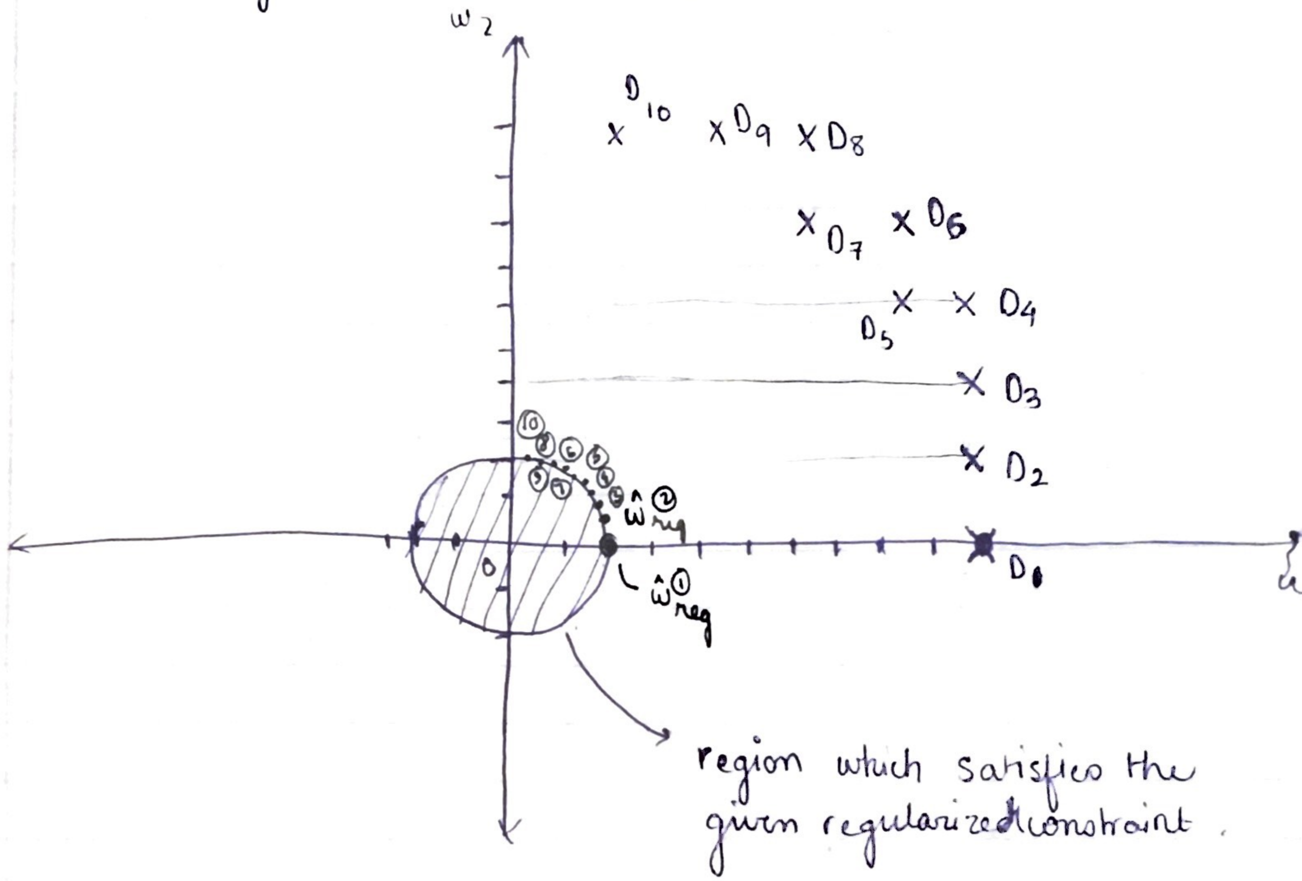
$$= \frac{1}{3}$$

$$= 0.33$$

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

Q2.

(a) L2 regularization : $\Omega(\underline{w}) = \|\underline{w}\|_2^2$, $C = 2^2$



- for D_1 : $(10, 0)$ $\qquad \hat{w}^{①}_{reg}$: $(2, 0)$
- for D_2 : $(10, 2)$

$$\frac{2-y}{2} = \frac{10-x}{10} \qquad x^2 + y^2 = 4$$

$$\therefore 5y = x \qquad \therefore 26y^2 = 4$$

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

So,

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

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 $\hat{w}_{lin}$

- for D_3 : $(10,4)$

$$\frac{4-y}{4} = \frac{10-x}{10} \qquad x^2+y^2=4$$

$$\therefore\ 29y^2 = 16$$

$$\therefore\ x = \frac{5y}{2} \qquad \therefore\ y = \frac{4}{\sqrt{29}},\quad x = \frac{10}{\sqrt{29}}$$

So,

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

- for D_4 : $(10,6)$

$$\frac{6-y}{6} = \frac{10-x}{10} \qquad x^2+y^2=4$$

$$\therefore\ 34y^2 = 36$$

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

So,

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

- for D_5 : $(8,6)$

$$\frac{6-y}{6} = \frac{8-x}{8} \qquad x^2+y^2=4$$

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

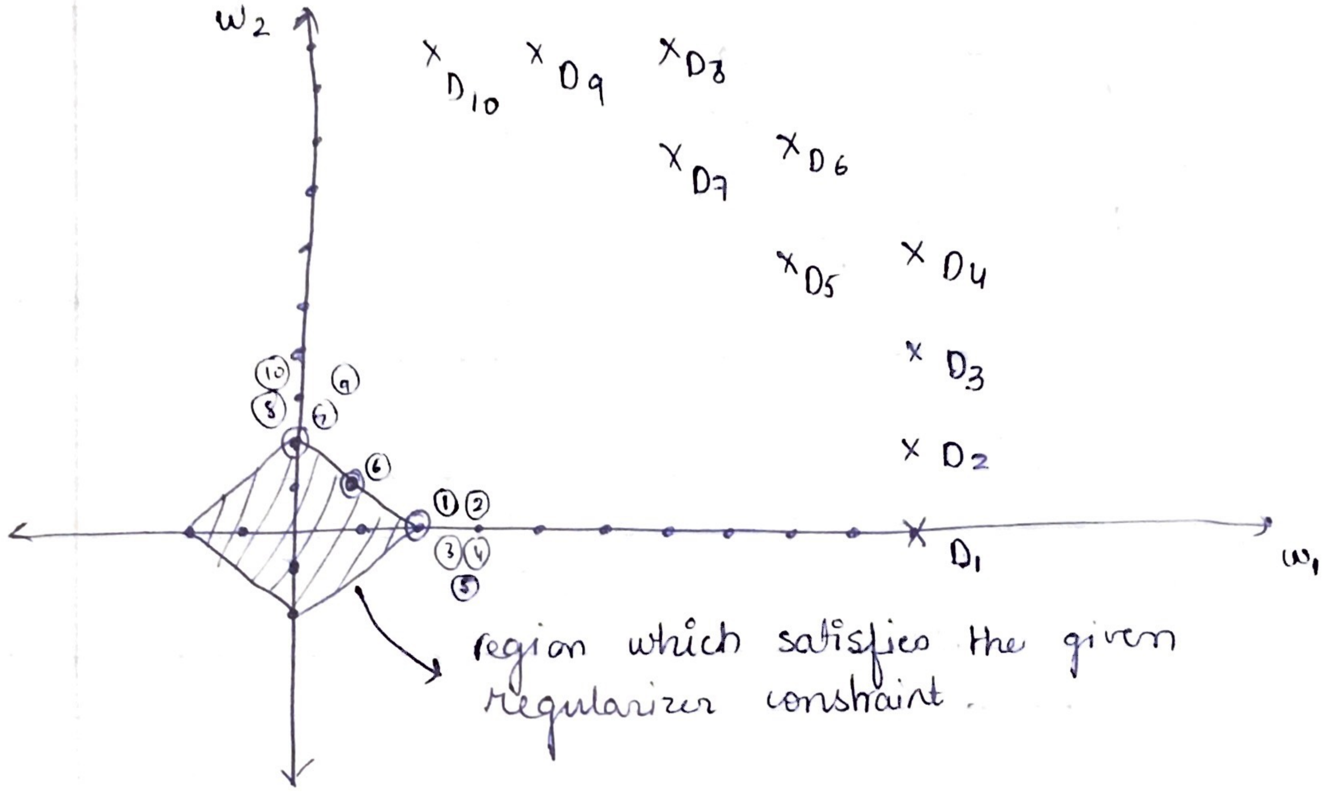
$$\therefore\ x = \frac{4}{3}y \qquad \hat{\omega}^{(5)}_{reg} = \left(\frac{8}{5}, \frac{6}{5}\right) \qquad \therefore\ y = \frac{6}{5},\quad x = \frac{8}{5}$$

Similarly,

- for D_6 : $(8,8)$ $\quad \hat{\omega}^{(6)}_{reg} = (\sqrt{2}, \sqrt{2})$
- for D_7 : $(6,8)$ $\quad \hat{\omega}^{(7)}_{reg} = (6/5, 8/5)$
- for D_8 : $(6,10)$ $\quad \hat{\omega}^{(8)}_{reg} = \left(3\sqrt{2/17}, 5\sqrt{2/17}\right)$
- for D_9 : $(4,10)$ $\quad \hat{\omega}^{(9)}_{reg} = \left(4/\sqrt{29}, 10/\sqrt{29}\right)$
- for D_{10} : $(2,10)$ $\quad \hat{\omega}^{(10)}_{reg} = \left(\sqrt{\frac{2}{13}}, 5\sqrt{\frac{2}{13}}\right)$

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

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



- For D_1 : $(10, 0)$ $\qquad \hat{w}^{(1)}_{reg} = (2, 0)$
- For D_2 : $(10, 2)$ $\qquad \hat{w}^{(2)}_{reg} = (2, 0)$

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 thats the $\hat{w}^{(2)}_{reg}$

- For D_3 : $(10, 4)$ $\qquad \hat{w}^{(3)}_{reg} = (2, 0)$
- For D_4 : $(10, 6)$ $\qquad \hat{w}^{(4)}_{reg} = (2, 0)$
- For D_5 : $(8, 6)$ $\qquad \hat{w}^{(5)}_{reg} = (2, 0)$

- for D6 : (8,8) $\hat{\omega}^{⑥}_{reg} = (1,1)$

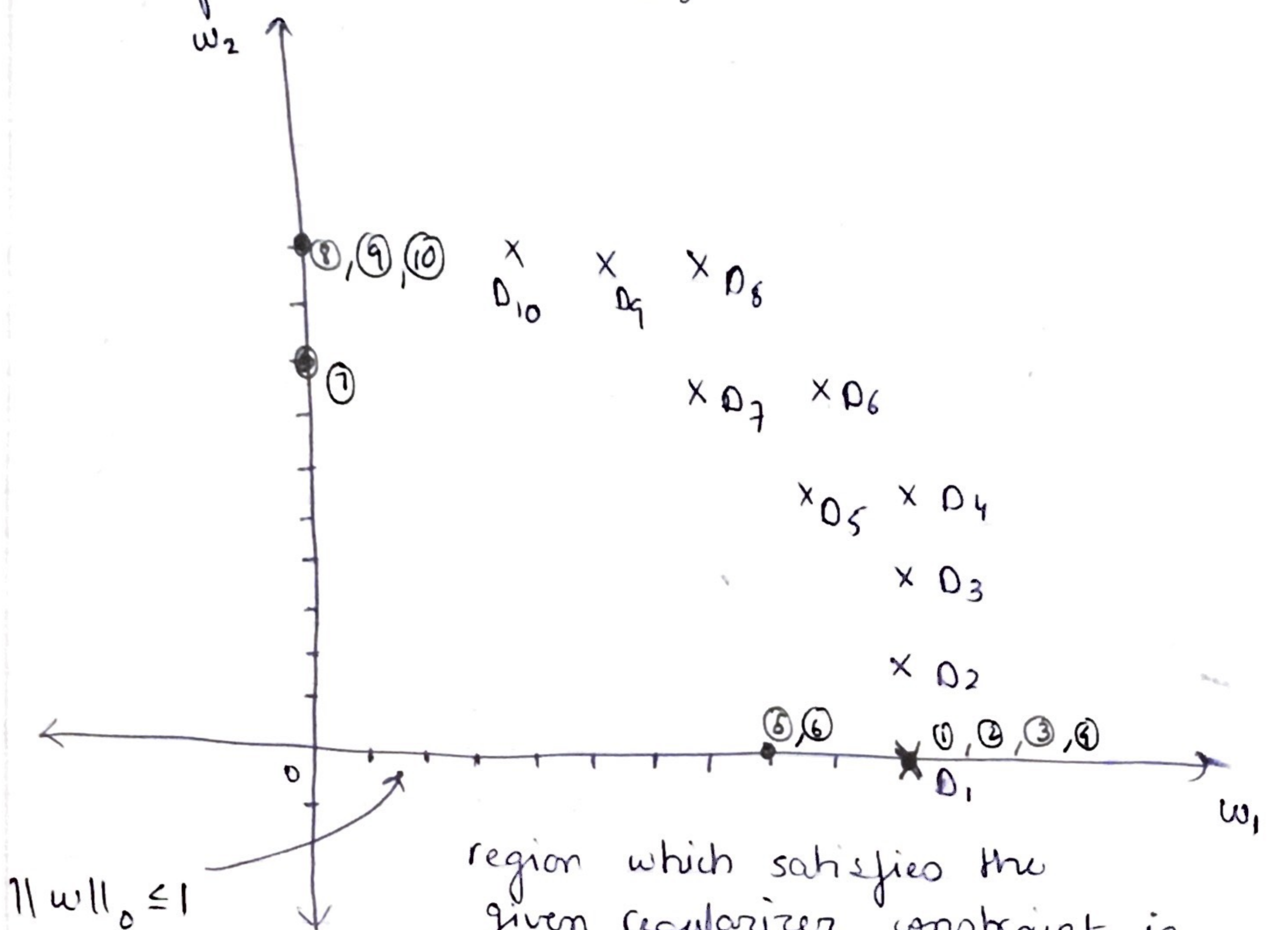
- for D7 : (6,8) $\hat{\omega}^{⑦}_{reg} = (0,2)$
- for D8 : (6,10) $\hat{\omega}^{⑧}_{reg} = (0,2)$
- for D9 : (4,10) $\hat{\omega}^{⑨}_{reg} = (0,2)$
- for D10 : (2,10) $\hat{\omega}^{⑩}_{reg} = (0,2)$

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

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

So, eight values become sparse from previous case.

(c) L0 regularization : $\Omega(\underline{w}) = \|w\|_0$, $C = 1$



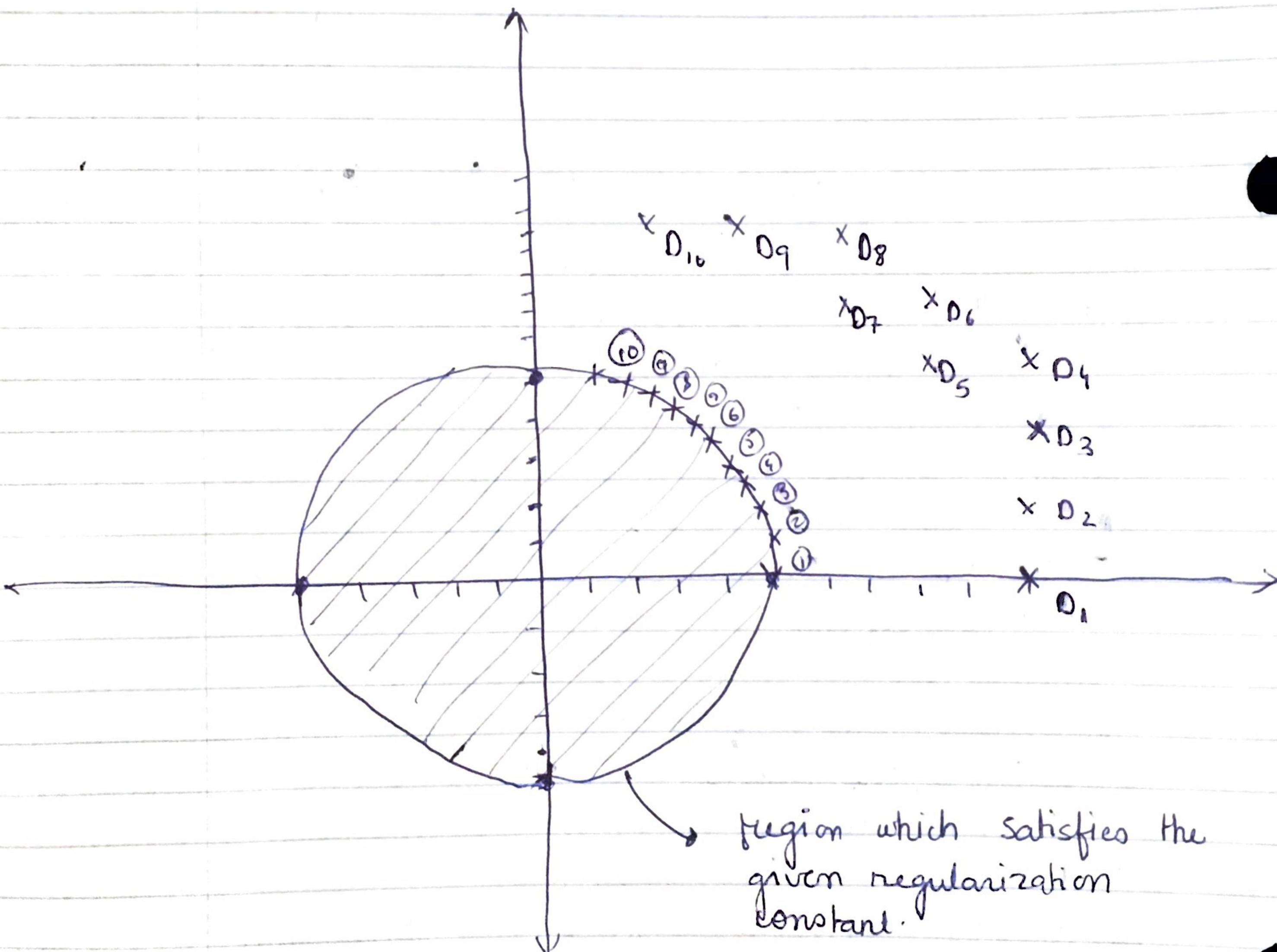
region which satisfies the given regularizer constraint is either x or y axis, given whichever axis comes in contact with the expanding circle first.

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

- for D_7 : $(6,8)$ $\quad \hat{w}^{(7)}_{reg} = (0,8)$
- For D_8 : $(6,10)$ $\quad \hat{w}^{(8)}_{reg} = (0,10)$
- for D_9 : $(4,10)$ $\quad \hat{w}^{(9)}_{reg} = (0,10)$
- for D_{10} : $(2,10)$ $\quad \hat{w}^{(10)}_{reg} = (0,10)$

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 : $\Omega(\underline{w}) = \|w\|_0$, $C = 5^2$



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

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

We use $x^2 + y^2 = 25$ instead of $x^2 + y^2 = 4$

- For D_3 : $(10, 4)$ $\quad \hat{w}_{reg}^{(3)} = \left(\frac{25}{\sqrt{29}}, \frac{10}{\sqrt{29}}\right)$
- For D_4 : $(10, 6)$ $\quad \hat{w}_{reg}^{(4)} = \left(\frac{25}{\sqrt{34}}, \frac{15}{\sqrt{34}}\right)$
- For D_5 : $(8, 6)$ $\quad \hat{w}_{reg}^{(5)} = (4, 3)$
- For D_6 : $(8, 8)$ $\quad \hat{w}_{reg}^{(6)} = \left(\frac{5}{\sqrt{2}}, \frac{5}{\sqrt{2}}\right)$
- For D_7 : $(6, 8)$ $\quad \hat{w}_{reg}^{(7)} = (3, 4)$
- For D_8 : $(6, 10)$ $\quad \hat{w}_{reg}^{(8)} = \left(\frac{15}{\sqrt{34}}, \frac{25}{\sqrt{34}}\right)$
- For D_9 : $(4, 10)$ $\quad \hat{w}_{reg}^{(9)} = \left(\frac{10}{\sqrt{29}}, \frac{25}{\sqrt{29}}\right)$
- For D_{10} : $(2, 10)$ $\quad \hat{w}_{reg}^{(10)} = \left(\frac{25}{\sqrt{26}}, \frac{5}{\sqrt{26}}\right)$

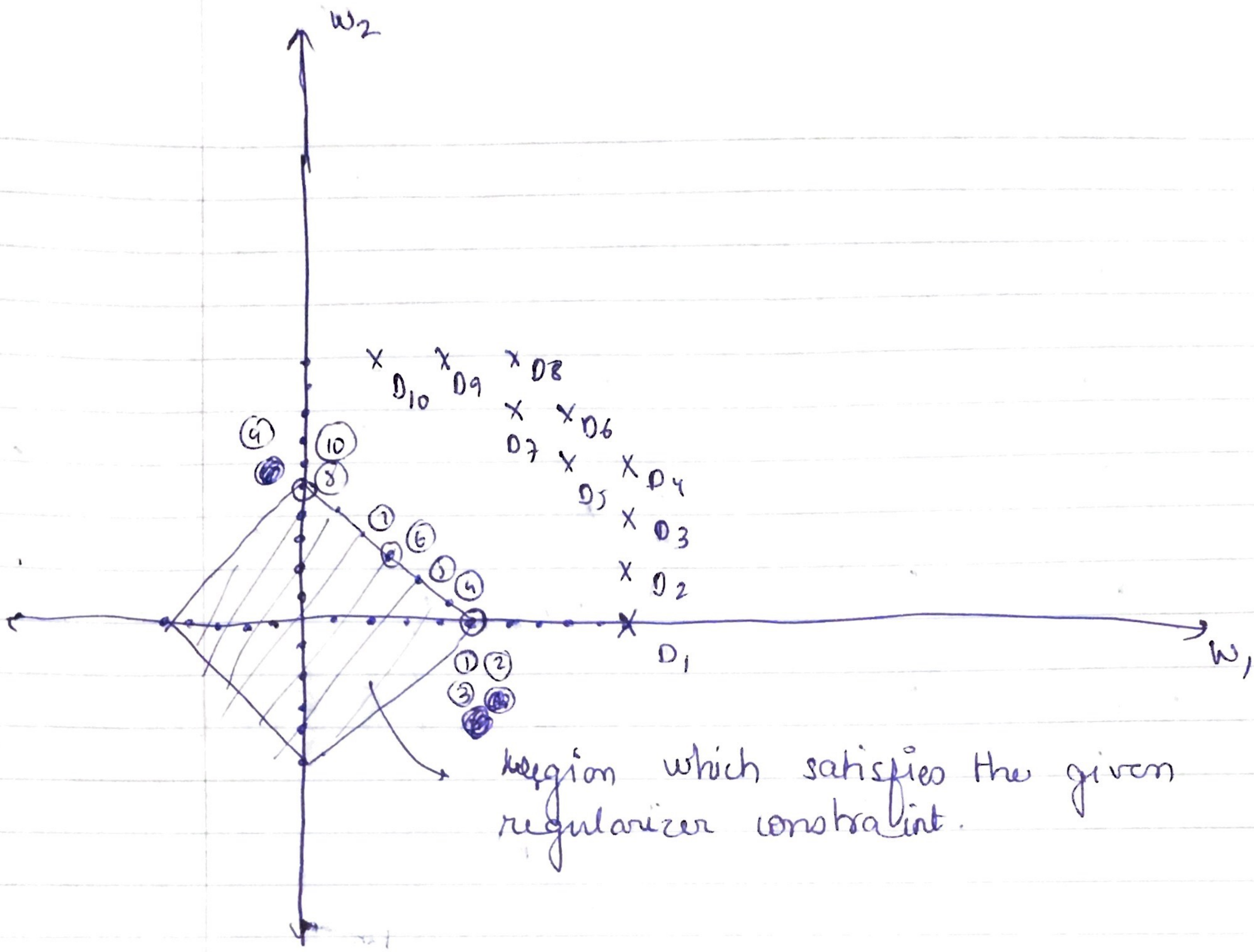
No extra sparsity.

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

Here its the same case as before seen in (b).

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

P.T.O.



- For D_1 : $(10, 0)$ $\qquad \hat{w}^{(1)}_{reg} = (5, 0)$
- For D_2 : $(10, 2)$ $\qquad \hat{w}^{(2)}_{reg} = (5, 0)$
- For D_3 : $(10, 4)$ $\qquad \hat{w}^{(3)}_{reg} = (5, 0)$
- For D_4 : $(10, 6)$ $\qquad \hat{w}^{(4)}_{reg} = (4.5, 0.5)$
- For D_5 : $(8, 6)$ $\qquad \hat{w}^{(5)}_{reg} = (3.5, 1.5)$

- For D_6 : $(8, 8)$ $\qquad \hat{w}^{(6)}_{reg} = (2.5, 2.5)$
- For D_7 : $(6, 8)$ $\qquad \hat{w}^{(7)}_{reg} = (1.5, 3.5)$
- For D_8 : $(6, 10)$ $\qquad \hat{w}^{(8)}_{reg} = (0.5, 4.5)$
- For D_9 : $(4, 10)$ $\qquad \hat{w}^{(9)}_{reg} = (0, 5)$
- For D_{10} : $(2, 10)$ $\qquad \hat{w}^{(10)}_{reg} = (0, 5)$

Here, $\hat{w}^{(1)}_{reg}$ was already sparse.
Now, all the rest except for $\hat{w}^{(6),(4),(5),(7),(8)}_{reg}$ are sparse.
So, four values become sparse from previous case.

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

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

Distance $= 8$

Between (a) - (c) smallest distance is given by (c)

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

Distance $= 8$

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

Distance $= 0$

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

Distance $= 5$

The distance for both cases is same.

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

Distance $= 5$

- D_4 : $(10,6)$

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

Distance $= 9.66$

Among (a) - (c) smallest distance is given by (c).

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

Distance $= 10$

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

Distance $= 6$

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

Distance $= 6.662$

(e) $\hat{\omega}_{reg} = (4.5, 0.5)$

Distance $= 7.778$

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

- $D_6 : (8, 8)$

(a) $\hat{\omega}_{reg} = (\sqrt{2}, \sqrt{2})$
distance $= 9.313$

(b) $\hat{\omega}_{reg} = (1, 1)$
distance $= 9.89$

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

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

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

distance $= 6.313$

(e) $\hat{\omega}_{reg} = (2.5, 2.5)$
distance $= 7.778$

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