R07922003 劉濬慶 ML HW1

1

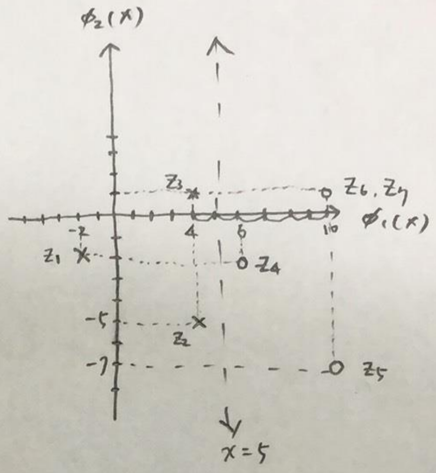
1. $Z = (\phi_1(x), \phi_2(x))$:

$Z_1 = (-2,-2)$ $Z_5 = (10,-7)$

$Z_2 = (4,-5)$ $Z_6 = (10,1)$

$Z_3 = (4,-1)$ $Z_7 = (10,1)$

$Z_4 = (6,-2)$



由圖可知 $\phi_1(x)=5$ 可最佳分割

A: $\phi_1(x)=5$ in the Z space

2

In [4]:
```python
#第2題
from sklearn import svm

x = np.array([[1,0],[0,1],[0,-1],[-1,0],[0,2],[0,-2],[-2,0]])
y = np.array([-1,-1,-1,1,1,1,1])

clf = svm.SVC(kernel='poly',degree=2,coef0=1,gamma=1,C=1e10)
clf.fit(x,y)
print(clf.support_vectors_,clf.support_)
print(y[clf.support_]*clf.dual_coef_[0])
```

```
[[ 0.  1.]
 [ 0. -1.]
 [-1.  0.]
 [ 0.  2.]
 [ 0. -2.]] [1 2 3 4 5]
[0.59647182 0.81065085 0.8887034  0.20566488 0.31275439]
```

2. $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7) = (0, 0.59647182, 0.81065085, 0.8887034, 0.20566488, 0.31275439, 0)$

$\alpha_k \neq 0$ 的為 SV: x_2, x_3, x_4, x_5, x_6

3

In [6]:
```python
#第3題
def g(x):
    r =np.sqrt(2)
    return np.array([1,r*x[0],r*x[1],x[0]**2,x[0]*x[1],x[1]*x[0],x[1]**2])
support = clf.support_
coef = clf.dual_coef_[0]
x4 = np.array([g(i) for i in x])

s = support[2]
b = y[s] - coef.dot(x4[support].dot(x4[s]))
b
```

Out[6]: -1.6661102048266958

3.

$$b = y_s - \sum_{\alpha_k > 0} \alpha_k y_k K(x_k, x_s)$$

(x_s, y_s) 為 support vector

$$g_{svm}(x) = \text{sign}\left(\sum_{\alpha_k>0} \alpha_k y_k K(x_k, x) + b \right) \; ; \; x = (x_1, x_2)$$

step: 由 3: ② 已算出 α_k, $1 \le k \le 7$ 出來

① 算 optimal b, 取 $x_4(-1,0)$, $y_4 = +1$ 套入得

$$b = -1.6661102048266958$$

② 算 $g_{svm}(x) = \text{sign}\left(\sum_{\alpha_k>0} \alpha_k y_k K(x_k, x) + b \right)$; $x = (x_1, x_2)$

SV: $x_2(0,1)$, $x_3(0,-1)$, $x_4(-1,0)$, $x_5(0,2)$, $x_6(0,-2)$

$$\begin{cases} K(x_2, x) = (1 + [0 \;\; 1]^T \cdot [x_1 \;\; x_2])^2 = (1+x_2)^2 \\ K(x_3, x) = (1 + [0 \;\; -1]^T [x_1 \;\; x_2])^2 = (1-x_2)^2 \\ K(x_4, x) = (1 + [-1 \;\; 0]^T [x_1 \;\; x_2])^2 = (1-x_1)^2 \\ K(x_5, x) = (1+2x_2)^2 \\ K(x_6, x) = (1-2x_2)^2 \end{cases}$$

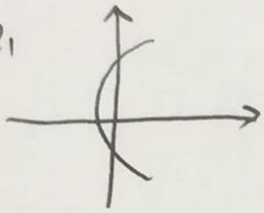
$$g_{svm} = -\alpha_2(1+x_2)^2 - \alpha_3(1-x_2)^2 + \alpha_4(1-x_1)^2 + \alpha_5(1+2x_2)^2 + \alpha_6(1-2x_2)^2 + b$$

4

4. 不一樣，一个為一次式，另一个為二次式

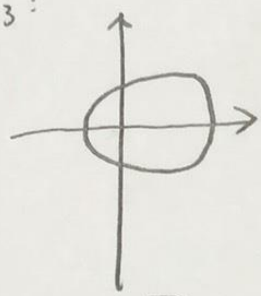
如果都在原本空間：

Q_1



拋物線

Q_3：



橢圓

所以也不同

5.

$$\exp(-x^2) = e^{-x^2}$$

$$= \frac{1}{e^{x^2}}$$

$$= \frac{1}{1+\frac{x^2}{1!}+\frac{x^4}{2!}+\frac{x^6}{3!}+\cdots+\frac{x^{2k}}{k!}+\cdots}$$

$$\left[\exp(-x^2)\right]^2 = \frac{1}{1+\frac{x^2}{1!}+\frac{x^4}{2!}+\frac{x^6}{3!}+\cdots+\frac{x^{2k}}{k!}+\cdots} \times \frac{1}{1+\frac{x^2}{1!}+\frac{x^4}{2!}+\frac{x^6}{3!}+\cdots+\frac{x^{2k}}{k!}+\cdots}$$

$$= \frac{1}{1+\frac{2x^2}{1!}+\left(\frac{2x^4}{2!}+\frac{x^4}{1!}\right)+\left(\frac{2x^6}{3!}+\frac{2x^6}{2!}\right)+\left(\frac{2x^8}{4!}+\frac{2x^8}{3!}+\frac{x^8}{2!2!}\right)+\cdots}$$

$$= \frac{1}{1+\frac{2x^2}{1!}+\frac{4x^4}{2!}+\frac{8x^6}{3!}+\frac{16x^8}{4!}+\cdots+\frac{2^k x^{2k}}{k!}+\cdots}$$

$$\|\tilde{\phi}(x)\|^2 = 1+\frac{2x^2}{1!}+\frac{2^2x^4}{2!}+\cdots+\frac{2^k x^{2k}}{k!}+\cdots$$

$$\exp(-x^2) = \sqrt{\frac{1}{\|\tilde{\phi}(x)\|^2}}$$

$$= \frac{1}{\|\tilde{\phi}(x)\|}$$ —— 得證

6

6. $k_{ij} = \cos(x_i, x_j)$

$$= \begin{bmatrix} 1 & \frac{\vec{x_1}\cdot\vec{x_2}}{|x_1||x_2|} & \frac{\vec{x_1}\cdot\vec{x_3}}{|x_1||x_3|} & \cdots & \frac{\vec{x_1}\cdot\vec{x_N}}{|x_1||x_N|} \\ \frac{\vec{x_2}\cdot\vec{x_1}}{|x_1||x_2|} & 1 & \frac{\vec{x_2}\cdot\vec{x_3}}{|x_2||x_3|} & \cdots & \frac{\vec{x_2}\cdot\vec{x_N}}{|x_2||x_N|} \\ \frac{\vec{x_3}\cdot\vec{x_1}}{|x_3||x_1|} & \frac{\vec{x_3}\cdot\vec{x_2}}{|x_3||x_2|} & 1 & \cdots & \frac{\vec{x_3}\cdot\vec{x_N}}{|x_3||x_N|} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\vec{x_N}\cdot\vec{x_1}}{|x_N||x_1|} & \frac{\vec{x_N}\cdot\vec{x_2}}{|x_N||x_2|} & \frac{\vec{x_N}\cdot\vec{x_3}}{|x_N||x_3|} & \cdots & 1 \end{bmatrix} = A$$

$A = A^T$ 為对稱矩陣 (symmetric) √

正半定 ⇒ 令 $z_i = \left[\frac{\vec{x_i}}{\sqrt{|x_1||x_i|}} \cdots \frac{\vec{x_N}}{\sqrt{|x_N||x_i|}}\right]$ ⇒ $z^T z$ $[z_1\ z_2 \cdots z_N]^T [z_1\ z_2 \cdots z_N]$ √

7

7. (P) $\min R^2$

subject to $|z_n - c|^2 \le R^2$ for $n = 1, 2, \ldots, N$

↳ $|z_n - c|^2 - R^2 \le 0$ for $n = 1, 2, \ldots, N$

$$L(R, c, \lambda) = R^2 + \sum_{n=1}^{N} \lambda_n \left(|z_n - c|^2 - R^2\right)$$

8

8.

$$\min_{R\in\mathbb{R},\, c\in\mathbb{R}^d}\left(\max_{\lambda_n\geq 0} L(R,c,\lambda)\right) \geq \min_{R\in\mathbb{R},\, c\in\mathbb{R}^d} L(R,c,\lambda')$$

$$\Rightarrow \min_{R\in\mathbb{R},\, c\in\mathbb{R}^d}\left(\max_{\lambda_n\geq 0} L(R,c,\lambda)\right) \geq \max_{\text{all } \lambda'\geq 0}\ \min_{R\in\mathbb{R},\, c\in\mathbb{R}^d} L(b,w,\lambda')$$

$\Rightarrow$ 解 $\displaystyle\max_{\text{all } \lambda'\geq 0}\left[\min_{R\in\mathbb{R},\, c\in\mathbb{R}^d} R^2 + \sum_{n=1}^{N}\lambda_n\left(|z_n-c|^2 - R^2\right)\right]$

对 R 及 C 求導數：

$$\frac{\partial L(R,c,\lambda)}{\partial R} = 2R - 2\sum_{n=1}^{N}\lambda_n R$$

$$\frac{\partial L(R,c,\lambda)}{\partial c} = \frac{\partial\left(R^2 + \sum_{n=1}^{N}\lambda_n\left(|z_n-c|^2 - R^2\right)\right)}{\partial c}$$

$$= \frac{\partial\left(\sum_{n=1}^{N}\lambda_n\left(z_n^T z_n - 2z_n^T c + c^T c\right)\right)}{\partial c}$$

$$= \sum_{n=1}^{N}\lambda_n(2c - 2z_n)$$

令 $\dfrac{\partial L(R,c,\lambda)}{\partial c} = 0$；求 optimal c

$$\Rightarrow \sum_{n=1}^{N}\lambda_n(2c - 2z_n) = 0$$

$$\Rightarrow c\sum_{n=1}^{N}\lambda_n - \sum_{n=1}^{N}\lambda_n z_n = 0$$

$$c = \frac{\sum_{n=1}^{N}\lambda_n z_n}{\sum_{n=1}^{N}\lambda_n}$$

令 $\dfrac{\partial L(R,c,\lambda)}{\partial R} = 0$

$$\Rightarrow 2R - 2\sum_{n=1}^{N}\lambda_n R = 0$$

$$\Rightarrow R = R\sum_{n=1}^{N}\lambda_n$$

$$\Rightarrow R \neq 0;\ 1 = \sum_{n=1}^{N}\lambda_n$$

KKT：

1. $|z_n - c|^2 \leq R^2$ for $n = 1, 2, \ldots, N$
2. $\lambda_n \geq 0$
3. $R\left(1 - \sum_{n=1}^{N}\lambda_n\right) = 0$, $c_\lambda\sum_{n=1}^{N}\lambda_n = \sum_{n=1}^{N}\lambda_n x_n$
4. $\lambda_n\left(|z_n - c|^2 - R^2\right) = 0$

9

9. 根據題意: $1=\sum_{n=1}^{N}\lambda_n$, $c=\dfrac{\sum_{n=1}^{N}\lambda_n z_n}{\sum_{n=1}^{N}\lambda_n}=\sum_{n=1}^{N}\lambda_n z_n$

$$L(R,c,\lambda)=R^2+\sum_{n=1}^{N}\lambda_n\left(|z_n-c|^2-R^2\right)$$

$$=\sum_{n=1}^{N}\lambda_n|z_n-c|^2+\underbrace{R^2\left(1-\sum_{n=1}^{N}\lambda_n\right)}_{}=0$$

$$=\sum_{n=1}^{N}\lambda_n\left(z_n^Tz_n-2z_n^Tc+c^Tc\right)$$

$$=\sum_{n=1}^{N}\lambda_n z_n^T z_n-2c\sum_{n=1}^{N}\lambda_n z_n^T+c^Tc\overbrace{\sum_{n=1}^{N}\lambda_n}^{1\,=}$$

$$=\sum_{n=1}^{N}\lambda_n z_n^T z_n-2\left(\sum_{n=1}^{N}\lambda_n z_n\right)\left(\sum_{n=1}^{N}\lambda_n z_n^T\right)+\left(\sum_{n=1}^{N}\lambda_n z_n\right)\left(\sum_{n=1}^{N}\lambda_n z_n^T\right)$$

$$=\sum_{n=1}^{N}\lambda_n z_n^T z_n-\left(\sum_{n=1}^{N}\lambda_n z_n\right)\left(\sum_{n=1}^{N}\lambda_n z_n^T\right)$$

#

10

10.

$$\text{Objective}(\lambda)=\sum_{n=1}^{N}\lambda_n z_n^T z_n-\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m z_n^T z_m$$

$$=\sum_{n=1}^{N}\lambda_n K(x_n,x_n)-\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m K(x_n,x_m)$$

$$\lambda_n\left(|z_n-c|^2-R^2\right)=0$$

$\because\ \lambda_i>0$

$\therefore\ |z_i-c|^2-R^2=0$

$R^2=|z_n-c|^2$, $c=\sum_{n=1}^{N}\lambda_n z_n$

11

11. $C \ge \max_{1 \le n \le N} \alpha_n^*$

hard margin 的跟 soft margin 的 Lagrange Duel 是一樣的

只差在 α_n 的上限需 $\le C$

soft margin lagrange duel :

$$\max_{\alpha_n \ge 0,\, \beta_n \ge 0} \left(\min_{b, w, \xi} \ \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n + \sum_{n=1}^{N} \alpha_n \cdot \left(1 - \xi_n - y_n(w^T z + b)\right) + \sum_{n=1}^{N} \beta_n (-\xi_n) \right)$$

$\frac{\partial L}{\partial \xi_n} = 0 = C - \alpha_n - \beta_n \Rightarrow \beta_n = C - \alpha_n$ 又 $\beta_n \ge 0$

$\therefore \ C - \alpha_n \ge 0$

$\Rightarrow \ \alpha_n \le C$

if $C \ge \max_{1 \le n \le N} \alpha_n^*$, 則 α^* 依然 optimal solution to soft margin SVM

12

12.

$\tilde{K}(x,x') = pK(x,x')$, $p>0$, $\tilde{C} = \frac{C}{p}$

原：

$\min \ \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n\alpha_m y_n y_m K(x_n,x_m) - \sum_{n=1}^{N}\alpha_n$ $\quad g(x) = \text{sign}\left(\sum_{n=1}^{N}\alpha_n y_n K(x_n,x) + b\right)$

subject to $\sum_{n=1}^{N} y_n\alpha_n = 0$, $0 \le \alpha_n \le C$

implicitly $\quad w = \sum_{n=1}^{N}\alpha_n y_n x_n$

$\beta_n = C - \alpha_n$

改：

$\min \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \tilde{K}(x,x') - \sum_{n=1}^{N}\alpha_n$

$\Rightarrow \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} p\alpha_n\alpha_m y_n y_m K(x,x') - \sum_{n=1}^{N}\alpha_n$

$\Rightarrow \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\frac{1}{p}(p\alpha_n)(p\alpha_m) y_n y_m K(x,x') - \sum_{n=1}^{N}p\alpha_n \underset{\text{可化簡}}{\Rightarrow} \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}(p\alpha_n)(p\alpha_m)y_n y_m K(x,x') - \sum_{n=1}^{N}p\alpha_n$

顯然地 $\underline{\tilde{\alpha} = \frac{1}{p}\alpha}$

subject to $\sum_{n=1}^{N} y_n\alpha_n = 0 \Rightarrow \sum_{n=1}^{N} p y_n\tilde{\alpha}_n = 0 \Rightarrow \underline{\sum_{n=1}^{N} y_n\tilde{\alpha}_n = 0}$

$0 \le \alpha_n \le C \Rightarrow 0 \le p\tilde{\alpha} \le p\tilde{C} \Rightarrow \underline{0 \le \tilde{\alpha} \le \tilde{C}}$

Implicitly $\quad \tilde{w} = \sum_{n=1}^{N}\tilde{\alpha} y_n x_n = \frac{1}{p}\sum_{n=1}^{N}\alpha y_n x_n = \frac{w}{p}$

$\tilde{\beta}_n = \tilde{C} - \tilde{\alpha}_n = \frac{1}{p}(C - \alpha_n) = \frac{1}{p}\beta_n$

$\tilde{b} = y_m - \sum_{n=1}^{N}\tilde{\alpha}_n y_n \tilde{K}(x_n,x_m) = y_m - \sum_{n=1}^{N}\frac{\alpha_n}{\cancel{p}} y_n \cancel{p} K(x_n,x_m)$

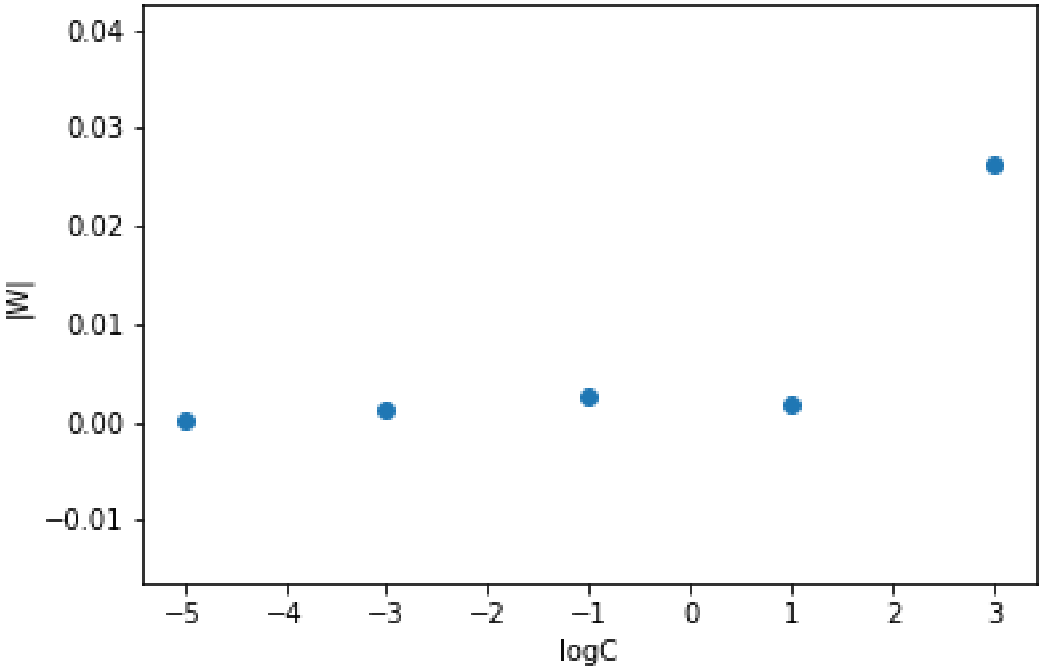
$=$ 原 b

$\tilde{g}(x) = \text{sign}\left(\sum_{n=1}^{N}\tilde{\alpha}_n y_n \tilde{K}(x_n,x) + \tilde{b}\right)$

$= \text{sign}\left(\sum_{n=1}^{N}\frac{\alpha_n}{\cancel{p}} y_n \cancel{p} K(x_n,x) + b\right)$

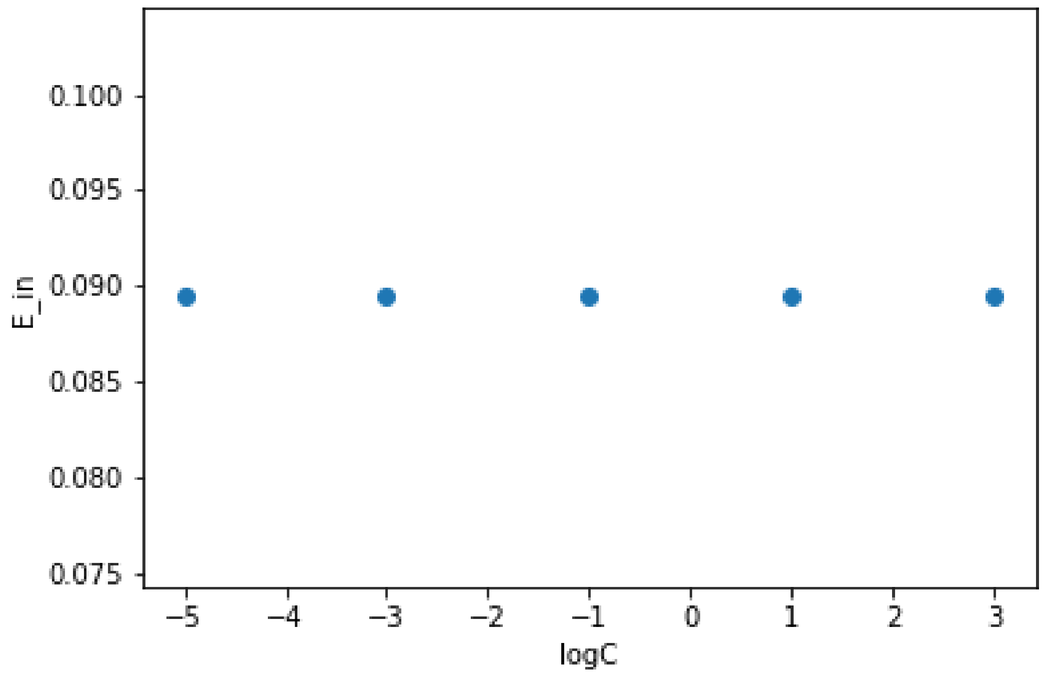
$=$ 原 $g(x)$ — 得證

- 13



Findings : logC=3 時，|w|長度最大

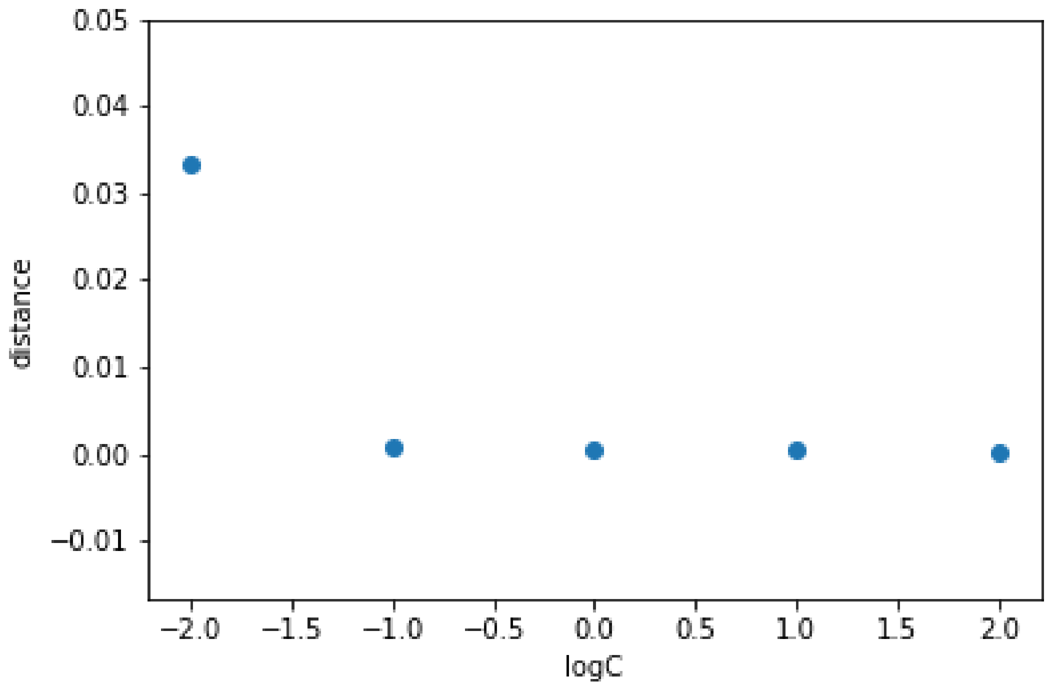
- 14



Findings : E in 都一樣

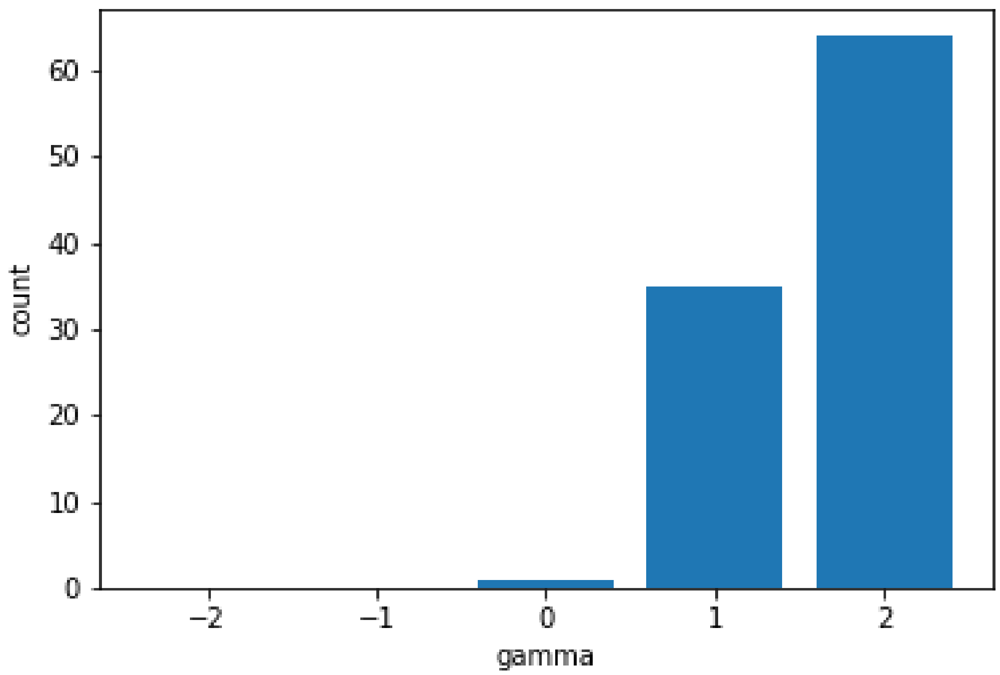
-

- 15



Findings : logC=-2 時，距離最遠

- 16



Findings : gamma 幾乎不會選到 -1 跟 -2