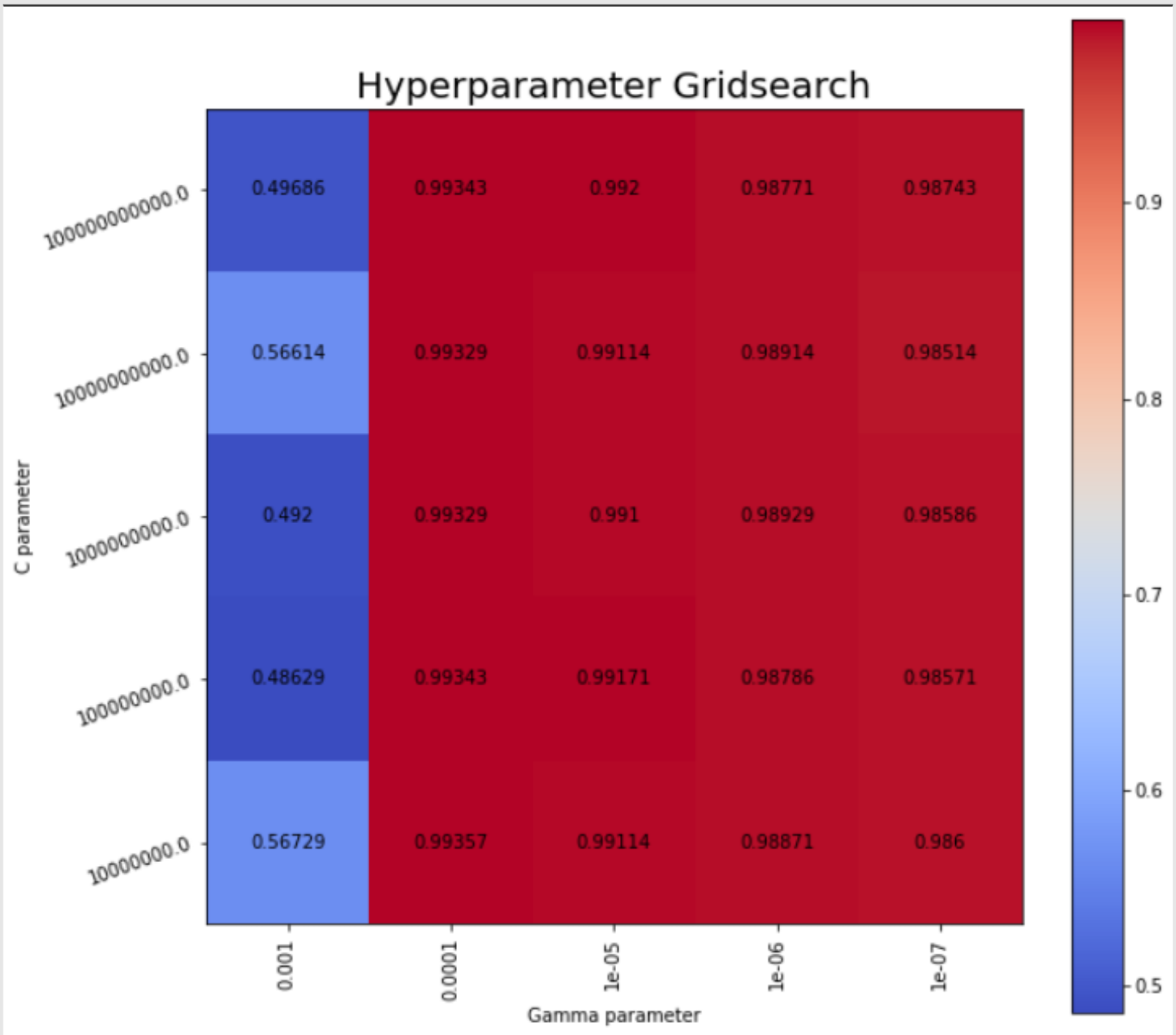
Part1. Coding

1.best parameter:

```
[ ] print(f"best_parameters : parameter for C is {best_C_para} , parameter for Gamma is {best_gamma_para} , with best average accuracy : {best_avg_acc}")

best_parameters : parameter for C is 10000000.0 , parameter for Gamma is 0.0001 , with best average accuracy : 0.9935714285714285
```

2.Hyperparamrter Gridsearch result:



Part 2.

1、

A matrix is positive semidefinite means that all of its eigenvalues are non-negative.

($K_{ij} = K_{ji}$ = 第i跟第j筆資料的內積)

Since K (Gram matrix) is symmetric, so by Spectral Thm., we have $K = V \Lambda V^T$, where V is an orthonormal matrix $\underline{V_t}$ and the diagnoal matrix
↳ each row of V

Λ contains the eigenvalues λ_t of K, and λ_t are all non-negative.

∴

We consider the feature map: $\phi : x_i \mapsto \left(\sqrt{\lambda_t}\, V_{ti}\right)_{t=1}^{n} \in \mathbb{R}^n$

We find that $\phi(x_i)^T \phi(x_j) = \sum_{t=1}^{n} \lambda_t V_{ti} V_{tj}$

$= (V \Lambda V^T)_{ij}$

$= K_{ij} = k(x_i, x_j)$ #

2、

Taylor expansion of e^x:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \frac{x^n}{n!} \quad (n \text{ up to } \infty)$$

又∵講義投影片 P15 的 3 個定理：

Given valid kernel $k_1(x,x')$, $k_2(x,x')$, the following new kernels will also be valid:

(6.13) $k(x,x') = c \cdot k_1(x,x')$, where $c > 0$ is a constant.

(6.18) $k(x,x') = k_1(x,x')\,k_2(x,x')$

(Note that k_1, k_2 不一定要不同)

(6.17) $k(x,x') = k_1(x,x') + k_2(x,x')$

$\Rightarrow$ ∵ every term in $e^{k_1(x,x')}$ is a valid kernel

≒ Taylor expansion (by 6.13 + 6.18)

, sum these terms up is also a valid kernel (by 6.17)

$\Rightarrow$ Thus $e^{k_1(x,x')}$ is a valid kernel function. #

3、

a. Valid:

We know $k_1(x, x')$ is valid, $\therefore \exists \phi$, s.t.

$$\phi(x)^T \phi(x') = k_1(x, x')$$

$\Rightarrow$ now consider the other mapping $\phi'(x)$,

$\phi'(x) = \langle \phi(x), 1 \rangle$, then $\phi'(x)^T \phi'(x')$

若ϕ有D維，
第k維以ϕ_k記，
則$\phi' = \langle \phi, 1 \rangle$
$= \langle \phi_1, \phi_2, \dots \phi_k, 1 \rangle$
此表示法以此
類推到各題

$$= \langle \phi(x), 1 \rangle^T \langle \phi(x'), 1 \rangle$$

$$= \phi(x)^T \phi(x') + 1$$

$$= k_1(x, x') + 1$$

$\Rightarrow$ so a. is a valid kernel #

b. not valid:

反例: consider 2筆資料 $(-\frac{1}{2}, \frac{1}{2})$; $(\frac{1}{2}, \frac{-1}{2})$,

valid kernel $k_1(x, x') = (x^T x')^2$

(it is the example given in 講義 P13)

$$\Rightarrow k(x, x') = k_1(x, x') - 1 = (x^T x')^2 - 1$$

$\Rightarrow$ computing K using $k(x, x')$: $\begin{bmatrix} -\frac{3}{4} & -\frac{3}{4} \\ -\frac{3}{4} & -\frac{3}{4} \end{bmatrix}$

K 作 eigenvalue decomposition:

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\frac{3}{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$\Rightarrow$ exists negative eigenvalue

c. valid:

consider a mapping: $\phi' : x_i \to (e^{x_i^T x_i}, 0)$

$\Rightarrow \phi'(x)^T \phi'(x')$ (令它叫 $k_2(x, x')$)

$= (e^{x^T x}, 0) \cdot (e^{x'^T x'}, 0)$

$= e^{\|x\|^2} \cdot e^{\|x'\|^2}$

又∵ $k_1(x, x')^2$ 是 valid kernel function

(by 2. 裡面寫到的 6.18)

$\Rightarrow$ ∴原式可以寫成 $k_1(x, x')^2 + k_2(x, x')$

$\Rightarrow$ by 2. 裡面寫到的 6.17, valid kernel 相加仍是 valid kernel

$\Rightarrow$ ∴ c. is valid kernel.

d. valid:

$$k(x,x') = k_1(x,x')^2 + e^{k_1(x,x')} - 1$$

$$= k_1(x,x')^2 \cancel{-1} + \underbrace{\cancel{1} + k_1(x,x') + \frac{k_1(x,x')^2}{2!} + \frac{k_1(x,x')^3}{3!} + \cdots + \frac{k_1(x,x')^n}{n!}}_{\text{Taylor expansion of } e^{k_1(x,x')}} \quad (n \text{ up to } \infty)$$

$\Rightarrow$ by 2. 裡面寫到的 (6.13), (6.18),

$k(x,x')$ 每個 term 都是 valid kernel

$\Rightarrow$ Sum these valid kernels, and it is also valid # (by 2. 裡面寫到的 (6.17)

4.

Corresponding lagrangian function:

$$L(x) = (x-2)^2 + \lambda[(x+3)(x-1)-3]$$

where $\lambda \geq 0$ (lagrange multiplier)

We need to find corresponding x, λ,

s.t. $L(x)$ is maximized, subjected to $\lambda \geq 0$:

$$\frac{dL(x)}{d\lambda} = 0: \quad x^2+2x-6=0$$

$$\Rightarrow x = -1 \pm \sqrt{7}$$

$$2(x-2)+\lambda(2x+2)$$

$$\frac{dL(x)}{dx}=0: \; 2(x-2)+\lambda(2x+2)=0$$

$$\Rightarrow \lambda=\frac{2-x}{x+1}$$

$\Rightarrow$ 分別代入 $x=-1+\sqrt{7}$ 及 $x=-1-\sqrt{7}$

$$x=-1+\sqrt{7}: \; \lambda=\frac{3-\sqrt{7}}{\sqrt{7}}\geq 0 \Rightarrow (\checkmark)$$

$$x=-1-\sqrt{7}: \; \lambda=\frac{3+\sqrt{7}}{-\sqrt{7}}=<0 \Rightarrow \text{discard}$$

($\because$ 設 $\lambda\geq 0$)

$$\Rightarrow x=-1+\sqrt{7}\;;\; \lambda=\frac{3-\sqrt{7}}{\sqrt{7}} \quad \#$$

*補 state dual problem

from $\frac{dL(x)}{dx}=0$: $x=\frac{2-\lambda}{1+\lambda}$ => 代入 $L(x)$

$$\left[\left(\frac{2-\lambda}{1+\lambda}\right)-2\right]^2+$$

$$\lambda\left(\left(\frac{2-\lambda}{1+\lambda}\right)^2+2\left(\frac{2-\lambda}{1+\lambda}\right)-6\right)$$

$$=L(\lambda)$$

$$=\frac{-\lambda^2+4\lambda-4}{1+\lambda}+4-6\lambda=\frac{-7\lambda^2+2\lambda}{1+\lambda}$$

we need to maximize $L(\lambda)$,

subjected to $\lambda\geq 0$