NCTU Introduction to Machine Learning, Homework 4

Part. 1, Coding (50%):

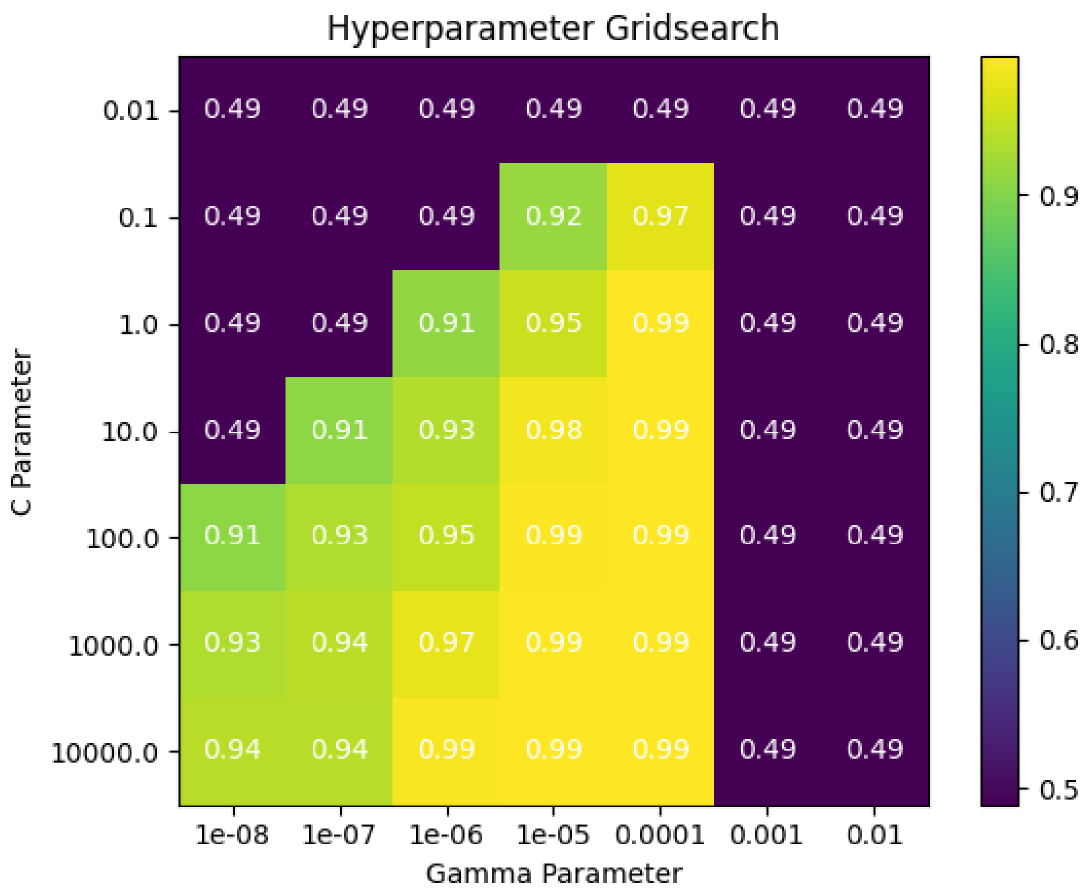
Q1

No output

Q2

```
print(f"best_parameters C = {best_parameters[0]}, gamma = {best_parameters[1]}")
✓ 0.3s
best_parameters C = 1.0, gamma = 0.0001
```

Q3



Part. 2, Questions (50%):

1. For a matrix to be positive semidefinite, all of its eigenvalues are non-negative,
We already know that K is symmetric, $K = V \Lambda V^T$, where V is an orthonormal matrix V_t and the diagonal matrix Λ contains the eigenvalues λ_t of K.
In order to be a valid kernel, the eigenvalues must all be positive so that $\phi: x_i = (\sqrt{\lambda_t} V_{ti})_{t=1}^{n} \in R^n \Rightarrow \phi(x_i)^T \phi(x_j) = \sum_{t=1}^{n} \lambda_t V_{ti} V_{tj} = (V \Lambda V_j^T) = k_{ij}$
Thus, kernel matrix $K = [k(x_n, x_m)]_{nm}$ being positive semidefinite is the necessary and sufficient condition for $k(x, x')$ to be a valid kernel.

2. From Taylor Series at 0:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

We subsitute x to $k_1(x, x')$

$$e^{k_1(x,x')} = 1 + \frac{k_1(x,x')}{1!} + \frac{k_1(x,x')^2}{2!} + \frac{k_1(x,x')^3}{3!} + \cdots$$

Since $k_1(x,x')$ is a valid kernel and the multiplication of two valid kernel is a valid kernel, $k_1(x,x')^2$ is a valid kernel, We can then prove $k_1(x,x')^n$ where $n \in N$ is valid by mathematical induction
Furthermore, we also know the addition of two valid kernel function is also a valid kernel function. Thus $k(x,x') = \exp(k_1(x,x'))$ is a valid kernel function.

3. Suppose we already know that given valid kernels $k_1(x,x')$ and $k_2(x,x')$, the following $k(x,x')$ will also be valid;

$$k(x,x') = k_1(x,x') + k_2(x,x') \quad \cdots ①$$
$$k(x,x') = k_1(x,x') k_2(x,x') \quad \cdots ②$$
$$k(x,x') = \exp(k_1(x,x')) \quad \cdots ③$$
$$k(x,x') = f(x) k_1(x,x') f(x') \quad \cdots ④$$

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

Suppose that there exist $k_2(x,x') = 1$, K_2 will be a $N \times N$ all ones matrix, where N is the number of data. Since the eigenvalue of K_2 MUST be N and 0, K_2 is positive semidefinite, which means $k_2(x,x') = 1$ is valid,
$k(x,x') = k_1(x,x') + 1 = k_1(x,x') + k_2(x,x')$
By ②, $k(x,x')$ is a valid kernel function #

(b) Suppose that $k_1(x,x') = (x^T x')^2$ and $m = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $n = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
$k_1(m,n) = (m^T n)^2 = 0$
$k_1(m,m) = 1$
$k_1(n,n) = 1$
$\Rightarrow K_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
Since we know that $k(x,x') = k(m,n) - 1$
$K = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ and $\lambda = 1, -1 \Rightarrow$ not positive semidefinite, not valid #

(c) $k(x,x') = k_1(x,x')^2 + e^{\|x\|^2} \times e^{\|x'\|^2}$

$= k_1(x,x')^2 + f(x)\, k_2(x,x')\, f(x')$, where $f(x) = e^{\|x\|^2}$

$f(x') = e^{\|x'\|^2}$

$k_2(x,x') = 1$

By ② we know that $k_1(x,x')^2$ is valid

By (b) we know that $k_2(x,x') = 1$ is valid

By ④ we see $f(x)\, k_2(x,x')\, f(x')$ is valid

And by ① we conclude that $k(x,x')$ is valid ※

(d) By Talyor Series at 0,

$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$

$e^x - 1 = \frac{x}{1!} + \frac{x^2}{2!} + \cdots$

Therefore $e^{k_1(x,x')} - 1 = \frac{k_1(x,x')}{1!} + \frac{k_1(x,x')^2}{2!} + \frac{k_1(x,x')^3}{3!} + \cdots$

By ①,③ and mathematical induction we know that $k_3(x,x') = e^{k_1(x,x')} - 1$ is a valid kernel.

By ①, $k(x,x') = k_1(x,x')^2 + e^{k_1(x,x')} - 1 = k_1(x,x')^2 + k_3(x,x')$ is valid ※

4. $L(x,a) = (x-2)^2 - a(3 - (x+3)(x-1))$, $a \geq 0$

$L(x,a) = x^2 - 4x + 4 - a(-x^2 - 2x + 6)$, $a \geq 0$

$\frac{\partial L(x,a)}{\partial x} = 0 \Rightarrow 0 = 2x - 4 + 2ax + 2a \Rightarrow x(2+2a) = 4 - 2a \Rightarrow x = \frac{2-a}{1+a}$

Dual problem

$\Rightarrow$ maximize $\left(\frac{2-a}{1+a} - 2\right)^2 - a\left(-\left(\frac{2-a}{1+a}\right)^2 - 2\left(\frac{2-a}{1+a}\right) + 6\right)$

subject to $a \geq 0$