109550069 林宇柔

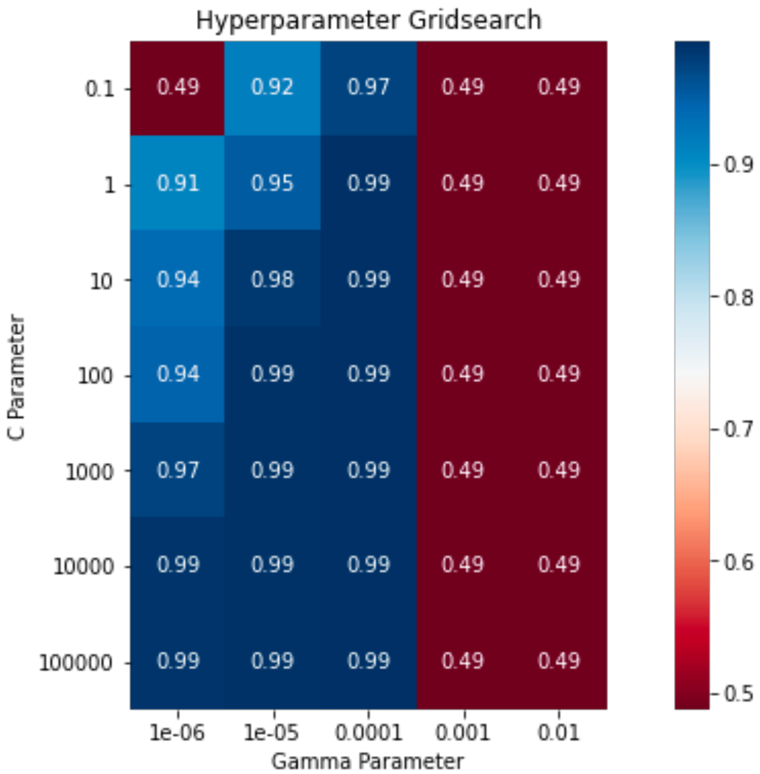
HW4 Report

Part1

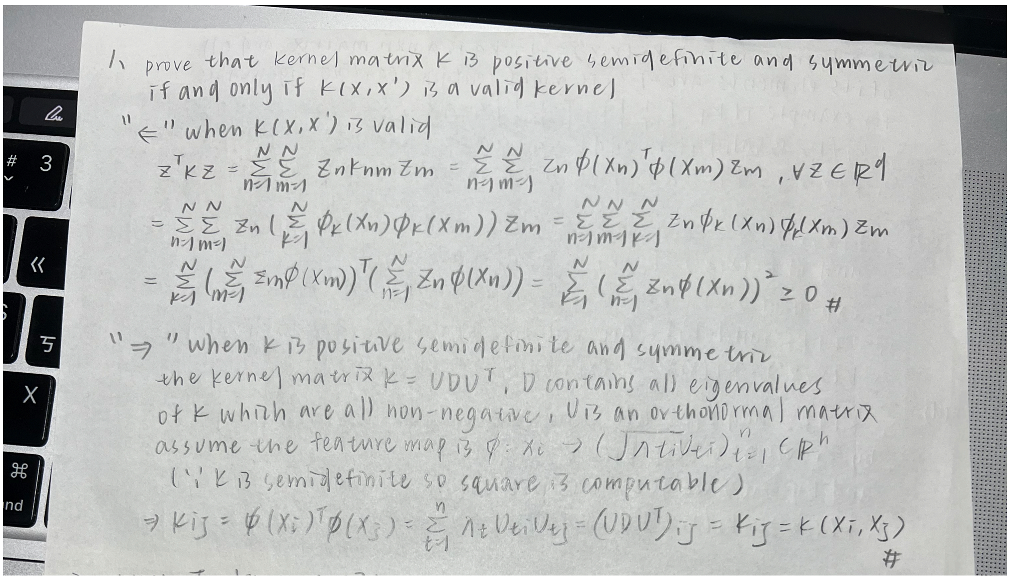
2. The best parameters

```
{'C': 1, 'gamma': 0.0001}
```

3. The grid search results



Part2



1. prove that kernel matrix K is positive semidefinite and symmetric if and only if $K(x, x')$ is a valid kernel

"$\Leftarrow$" when $K(x, x')$ is valid

$$z^T K z = \sum_{n=1}^{N}\sum_{m=1}^{N} z_n k_{nm} z_m = \sum_{n=1}^{N}\sum_{m=1}^{N} z_n \phi(x_n)^T \phi(x_m) z_m \, , \forall z \in \mathbb{R}^d$$

$$= \sum_{n=1}^{N}\sum_{m=1}^{N} z_n \Big(\sum_{k=1}^{N} \phi_k(x_n)\phi_k(x_m)\Big) z_m = \sum_{n=1}^{N}\sum_{m=1}^{N}\sum_{k=1}^{N} z_n \phi_k(x_n)\phi_k(x_m) z_m$$

$$= \sum_{k=1}^{N} \Big(\sum_{m=1}^{N} z_m \phi(x_m)\Big)^T \Big(\sum_{n=1}^{N} z_n \phi(x_n)\Big) = \sum_{k=1}^{N} \Big(\sum_{n=1}^{N} z_n \phi(x_n)\Big)^2 \geq 0 \; \#$$

"$\Rightarrow$" when K is positive semidefinite and symmetric

the kernel matrix $K = UDU^T$, D contains all eigenvalues of K which are all non-negative, U is an orthonormal matrix

assume the feature map is $\phi : x_i \rightarrow (\sqrt{\lambda_t} v_{ti})_{t=1}^{n} \in \mathbb{R}^n$

($\because$ K is semidefinite so square is computable)

$$\Rightarrow k_{ij} = \phi(x_i)^T \phi(x_j) = \sum_{t=1}^{n} \lambda_t v_{ti} v_{tj} = (UDU^T)_{ij} = k_{ij} = k(x_i, x_j)$$

#

109550069 林宇柔

2. With Taylor expansion:

$\exp(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

$\exp(k_1(x,x')) = 1 + k_1(x,x') + \frac{(k_1(x,x'))^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{(k_1(x,x'))^n}{n!}$

$\because$ if $k_1(x,x')$ and $k_2(x,x')$ are valid kernels,
$k_1(x,x') + k_2(x,x')$ and $k_1(x,x')k_2(x,x')$ are also valid kernels,
and all coefficients of Taylor expansion are positive

$\therefore k(x,x') = \exp(k_1(x,x'))$ is a valid kernel #

3.

(a) if a kernel matrix $k_3(x,x') = 1$, k_3 is a $n \times n$ matrix
and all of its elements are 1 so all its eigenvalues
are non-negative

for example if $k_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x = \lambda x \rightarrow \begin{bmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix} x = 0$

$\rightarrow \because \det = 0 \therefore (1-\lambda)^2 - 1 = 0$, $\lambda = 1, 0$ (both nonnegative)

$\therefore k_3(x,x') = 1$ is a valid kernel because it's positive semidefinite

$\because$ if k_1 and k_2 are valid kernels, $k_1 + k_2$ is also a valid kernel

$\therefore k(x,x') = k_1(x,x') + 1$ is a valid kernel #

(b) if a kernel matrix $k_4(x,x') = -1$, k_4 is a $n \times n$ matrix and all
of its elements are -1 so it might contain negative eigenvalue

for example if $k_4 = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}$, $\begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} x = \lambda x$, $\lambda = 0, -2$ $(-2 < 0)$

$\because k_1 + k_2$ is valid if k_1 and k_2 are valid, and k_4 is not valid

$\therefore k(x,x') = k_1(x,x') - 1$ is not a valid kernel #

(c) $\because k(x,x') = f(x)k_1(x,x')f(x')$ is a valid kernel if $k_1(x,x')$ is valid
and if $k_1(x,x') = 1$, k_1 is a valid kernal

$\therefore \exp(\|x\|^2) \cdot 1 \cdot \exp(\|x'\|^2)$ is a valid kernal

$\because k_1 + k_2$ and $k_1 k_2$ are valid if k_1 and k_2 are both valid

$\therefore k(x,x') = k_1(x,x')^2 + \exp(\|x\|^2) * \exp(\|x'\|^2)$

(d) $\because k_1 + k_2$ and $k_1 k_2$ are valid if k_1 and k_2 are both valid
by 2, we get $\exp(k_1(x,x'))$ is valid
and by 3.(b) we get -1 is not valid

$\therefore k(x,x') = k_1(x,x')^2 + \exp(k_1(x,x')) - 1$ is not a valid kernel #

4. minimizing $(x-2)^2$ subject to $(x+3)(x-1) \le 3$

by Lagrange function: $L(x,\lambda) = f(x) + \lambda g(x)$

$$= (x-2)^2 + \lambda((x+3)(x-1) - 3)$$
$$= x^2 - 4x + 4 + \lambda(x^2 + 2x - 6)$$
$$= (\lambda+1)x^2 + (2\lambda - 4)x + (4 - 6\lambda)$$

$\frac{L(x,\lambda)}{\partial x} = 2x(\lambda+1) + (2\lambda - 4) = 0$. $x = \frac{2-\lambda}{\lambda+1}$, $\lambda \ge 0$

$L(x,\lambda) = \frac{(2-\lambda)^2}{\lambda+1} - 2\frac{(2-\lambda)^2}{\lambda+1} + 4 - 6\lambda = \frac{-\lambda^2 + 4\lambda - 4 - 6\lambda^2 - 2\lambda + 4}{\lambda+1}$

$= \frac{-7\lambda^2 + 2\lambda}{\lambda+1} = \frac{\lambda(-7\lambda+2)}{\lambda+1}$

$\therefore$ dual problem is optimizing $\frac{-7\lambda^2 + 2\lambda}{\lambda+1}$ where $\lambda \ge 0$ #