NYCU Introduction to Machine Learning, Homework 4

109550116 楊傑宇

Part. 1, Coding (50%):

1.

```
5
2
1400
```

```python
def cross_validation(x_train, y_train, k=5):
    #combine
    length = y_train.shape[0]
    y_train = y_train.reshape((length,1))
    all_train = np.hstack([x_train,y_train])
    fold=None
    part_a=list()
    if length%k==0:
        part_a = np.vsplit(all_train,k)
    else:
        a=int(length/k)
        b=int(length%k)
        c=(a+1)*b
        part_aa = np.vsplit(all_train[:c,:],b)
        part_bb = np.vsplit(all_train[c:,:],k-b)
        part_a = part_aa+part_bb
        #print(len(part_aa))
        #print(part_aa[0].shape)
        #print(len(part_bb))
    #print(len(part_a))

    for i in range(k):
        val = part_a[i]
        train = None
        jud=0
        for j in range(k):
            if j==i:
                continue
            else:
```

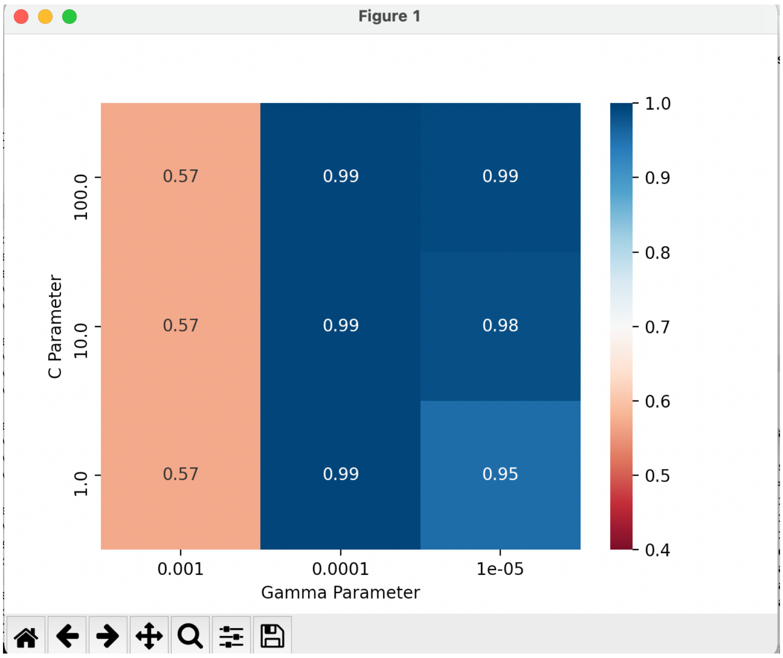
```python
                if jud == 0:
                    train = part_a[j]
                else:
                    train = np.vstack([train,part_a[j]])
                jud+=1

        combine = [train]
        combine.append(val)
        if i == 0:
            fold=[combine]
        else:
            fold.append(combine)
```

2.

```
best_parameters c= 100.0 , g= 0.0001
(base) jerrys_macbook_pro@jerrys-MacBook-Pro Desktop %
```

3.



Part. 2, Questions (50%):

1.

1-

By Mercer's thm. $k(x, x')$ is valid kernel if and only if kernel matrix is PSD.

Proof $\Rightarrow$ if $k(x, x')$ is a valid kernel. it has ϕ such that $k(x, x') = \phi(x) \cdot \phi(x')$

Let $V = [\phi(x_1), \phi(x_2) \cdots \phi(x_n)]$ then K (kernel matrix) $= V^T V$ so. K must be PSD because for any vector z we have $(z^T V^T)(Vz) \geq 0$. #

Proof $\Leftarrow$ since K is PSD it has non-negative eigenvalues and can be factored as $V \Lambda V^T$

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

we can find $\phi(x_i)^T \phi(x_j) = \sum_{t=1}^{n} \lambda V_{ti} V_{tj} = (V \Lambda V^T)_{ij} = K_{ij} = k(x_i, x_j)$

2.

2.

Using Taylor expansion around 0:

$$\exp(k) = \exp(0) + \exp(0)k + \frac{\exp(0)}{2!}k^2 + \frac{\exp(0)}{3!}\cdots$$

$$\exp(k) = 1 + k + \frac{1}{2}k^2 + \cdots$$

we can see that it's just an infinite series of multiplications and additions of k.

From 6.17, 6.18 we know exponential of a kernel is a kernel.

3、4.

3.

(a.)

if $k_1(x,x')$ use feature map $\phi_1(x)$

we can use another mapping $\phi: x \mapsto [\phi_1(x), 1]^T$

$\Rightarrow \langle \phi(x), \phi(x') \rangle = \langle \phi_1(x), \phi_1(x') \rangle + 1 = k_1(x,x') + 1 = k(x,x')$

$\Rightarrow k(x,x')$ is vaild.

(b.)

if $k_1(x,x') = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ \vdots & 0 & 1 & \cdots \\ & \vdots & \vdots & \end{bmatrix} = I \Rightarrow$ PSD, valid.

$k(x,x') = k_1(x,x') - 1 = \begin{bmatrix} 0 & -1 & -1 & \cdots \\ -1 & 0 & -1 & \cdots \\ -1 & -1 & 0 & -1 & \cdots \\ & & -1 & 0 \\ \vdots & \vdots & \vdots & \end{bmatrix}$ is not P.S.D.

$\Rightarrow$ not vaild.

(c.)

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

$$G = \begin{bmatrix} \exp(2|x_1|^2) & \exp(|x_1|^2+|x_2|^2) & \exp(|x_1|^2+|x_3|^2) & \cdots \\ \exp(|x_2|^2+|x_1|^2) & & & \\ \vdots & & \ddots & \end{bmatrix}$$

$\exp(2|x_1|^2) > 0$

$\exp(2|x_1|^2 + 2|x_2|^2) - \exp(|x_1|^2+|x_2|^2)\exp(|x_1|^2+|x_2|^2) = 0$.

∵ symmetric ∴ 合頑擇.

$\Rightarrow$ by Sylvester's criterion, G_2 is P.S.D.

$\Rightarrow$ $k_2(x, x')$ and $k_1(x, x')$ are valid.

$\Rightarrow$ $k_1(x, x')$ is valid.

(d)

$k_1(x, x')$ is valid, K_1 has eigenvalues ≥ 0

$\Rightarrow$ $\exp(k_1(x, x')) \geq 1$ $\Rightarrow$ $\exp(k_1(x, x')) - 1 \geq 0$

$\Rightarrow$ $k(x, x') = k_1(x, x')^2 + \exp(k_1(x, x')) - 1 \geq 0$

$\Rightarrow$ k has all pivots ≥ 0 and symmetric

$\Rightarrow$ k is valid.

4.

Lagrangian function: $L(x, \lambda) = f(x) - \lambda g(x)$

$f(x) = (x-2)^2$, $g(x) = 3 - (x+3)(x-1)$

with constraint: $\frac{\partial L(x,\lambda)}{\partial x} = 0 = 2(x-2) - \lambda(-2x-2)$.

$\Rightarrow$ $x = \frac{2-\lambda}{1+\lambda}$ $\lambda \geq 0$.

代回 $\Rightarrow$ $\frac{(-\lambda+2)^2}{(\lambda+1)} + \frac{(2\lambda-4)(-\lambda+2)}{(\lambda+1)} + \frac{(4-6\lambda)(\lambda+1)}{\lambda+1}$

$\Rightarrow$ $\frac{-7\lambda^2+2\lambda}{\lambda+1} = 0$ $\Rightarrow$ $-21\lambda^2 - 10\lambda + 2 = 0$ $\lambda \approx 0.15$

$x \approx 1.6\ldots$