Hw4 report

Part 1. Coding

1. K-fold

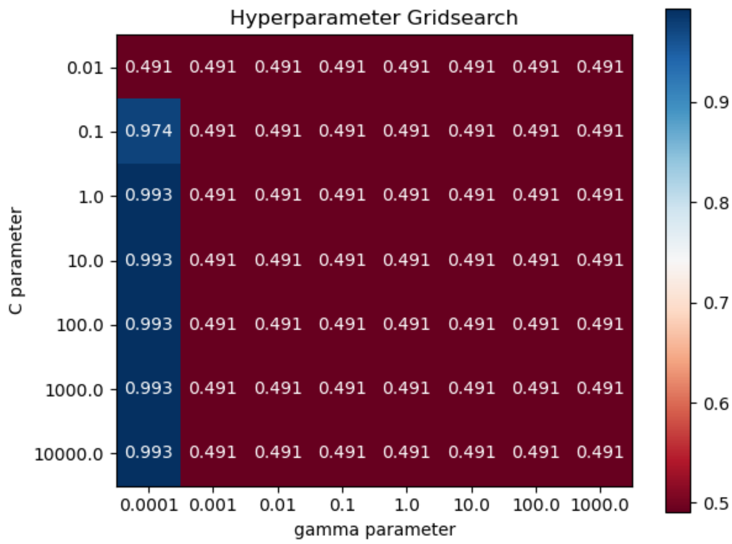
```
In 27
def cross_validation(x_train_, y_train_, k=5):
    rng = np.random.default_rng()
    n_samples = x_train_.shape[0]
    result = []
    index = np.arange(n_samples)
    rng.shuffle(index)
    idx = 0
    splits = []
    for _ in range(n_samples % k):
        step = n_samples // k + 1
        splits.append(index[idx:idx+step])
        idx += step
    for _ in range(k - n_samples % k):
        step = n_samples // k
        splits.append(index[idx:idx+step])
        idx += step
    for ii in range(k):
        result.append([np.concatenate([z for j, z in enumerate(splits) if j != ii]), splits[ii]])
    return result
```

```
In 28
kfold_data = cross_validation(x_train, y_train, k=10)
assert len(kfold_data) == 10 # should contain 10 folds of data
assert len(kfold_data[0]) == 2 # each element should contain train fold and validation fold
assert kfold_data[0][1].shape[0] == 700 # The number of data in each validation fold should equal to training data divided by K
```

2. C=1.0, gamma = 0.0001

```
In 45
arg = np.argmax(results)
best_parameters = [Cs[arg // len(gammas)], gammas[arg % len(gammas)]]
print(best_parameters)

[1.0, 0.0001]
```



3.

Part 2. Questions

1.

$\mathbf{K}$ is positive semidefinite → k is a valid kernel:

1. $\mathbf{K}$ is symmetric, so we have $\mathbf{K} = \mathbf{V\Lambda V^T}$.

Where $\mathbf{V}$ is an orthonormal matrix $\mathbf{v}_t$

, and $\mathbf{\Lambda}$ is a diagonal matrix contains the eigenvalues λ_t of $\mathbf{K}$.

2. $\because$ $\mathbf{K}$ is positive definite, $\therefore$ all eigenvalues $\lambda_t \geq 0$

3. consider the feature map $\phi: \mathbf{x}_i \mapsto \left(\sqrt{\lambda_t} v_{ti}\right)_{t=1}^{n} \in \mathbf{R}^n$

$$\Rightarrow \text{We can get } \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = \sum_{t=1}^{n} \lambda_t v_{ti} v_{tj} = (\mathbf{V\Lambda V^T})_{ij} = \mathbf{K}_{ij} = k(\mathbf{x}_\mathrm{i}, \mathbf{x}_j)$$

k is a valid kernel → $\mathbf{K}$ is positive semidefinite:

$$k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$$

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n)$

, then $\mathbf{K} = \mathbf{X}^\mathrm{T}\mathbf{X}$

For any $\mathbf{v} \in \mathbf{R}^n$:

$$\mathbf{v}^T \mathbf{K} \mathbf{v} = \mathbf{v}^T \mathbf{X}^T \mathbf{X} \mathbf{v} = (\mathbf{X}\mathbf{v})^T \mathbf{X}\mathbf{v} \geq 0$$

$\Rightarrow$ $\mathbf{K}$ is positive semidefinite.

Both → and ← are proven,

$\mathbf{K}$ is positive semidefinite ↔ k is a valid kernel

2.

The Maclaurin series of e^x is $\sum_{n=0}^{\infty}\frac{x^n}{n!}$

So $\exp(k_1(\mathbf{x},\mathbf{x}'))$ equals to $\sum_{n=0}^{\infty}\frac{k_1(\mathbf{x},\mathbf{x}')^n}{n!}$

$$= 1 + k_1(\mathbf{x},\mathbf{x}') + \frac{1}{2}k_1(\mathbf{x},\mathbf{x}')^2 + \frac{1}{6}k_1(\mathbf{x},\mathbf{x}')^3 + \cdots$$

According to $\begin{Bmatrix} k(\mathbf{x},\mathbf{x}') = ck_1(\mathbf{x},\mathbf{x}') \text{ for } c > 0 \\ \mathrm{k}(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}') + k_2(\mathbf{x},\mathbf{x}') \\ k(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}')k_2(\mathbf{x},\mathbf{x}') \end{Bmatrix}$

$\exp\left(k_1(\mathbf{x},\mathbf{x}')\right)$ is a valid kernel.

3.

a. valid

$$k(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}') + 1 = \phi_1(\mathbf{x})^T\phi_1(\mathbf{x}') + 1 = \begin{bmatrix}\phi_1(\mathbf{x}) \\ 1\end{bmatrix}^T \begin{bmatrix}\phi_1(\mathbf{x}') \\ 1\end{bmatrix}$$

feature map $\phi\colon \mathbf{x} \mapsto \begin{bmatrix}\phi_1(\mathbf{x}) \\ 1\end{bmatrix}$

b. invalid

Let $k_1(\mathbf{x},\mathbf{x}') = \mathbf{x}^T\mathbf{x}'$, $k(\mathbf{x},\mathbf{x}') = \mathbf{x}^T\mathbf{x}' - 1$

Let $\mathbf{x} = (1,0), \mathbf{x}' = (0,-1)$, then $\mathbf{K} = \begin{bmatrix}0 & -1 \\ -1 & 0\end{bmatrix}$

Eigenvalues of $\mathbf{K} = 1, -1$, so $\mathbf{K}$ is not positive semidefinite.

c. valid

Let $k_2(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}')^2$

and its feature map ϕ_2

, k_2 is a valid kernel by multiplication of two valid kernel.

$$k(\mathbf{x},\mathbf{x}') = k_2(\mathbf{x},\mathbf{x}') + \exp(|\mathbf{x}|^2) + \exp(|\mathbf{x}'|^2)$$
$$= \phi_1(\mathbf{x})^T\phi_1(\mathbf{x}') + \exp(|\mathbf{x}|^2) * \exp(|\mathbf{x}'|^2)$$
$$= \begin{bmatrix}\phi_2(\mathbf{x}) \\ \exp(|\mathbf{x}|^2)\end{bmatrix}^T \begin{bmatrix}\phi_2(\mathbf{x}') \\ \exp(|\mathbf{x}'|^2)\end{bmatrix}$$

feature map $\phi\colon \mathbf{x} \mapsto \begin{bmatrix}\phi_2(\mathbf{x}) \\ \exp(|\mathbf{x}|^2)\end{bmatrix}$

d. valid

Let $k_2(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}')^2$, and its feature map ϕ_2

$$k(\mathbf{x},\mathbf{x}') = k_2(\mathbf{x},\mathbf{x}') + \exp\big(k_1(\mathbf{x},\mathbf{x}')\big) - 1 = k_2(\mathbf{x},\mathbf{x}') + \sum_{n=0}^{\infty} \frac{k_1(\mathbf{x},\mathbf{x}')^n}{n!} - 1$$

$$= k_2(\mathbf{x},\mathbf{x}') + \sum_{n=1}^{\infty} \frac{k_1(\mathbf{x},\mathbf{x}')^n}{n!}$$

k is a valid kernel by addition between valid kernels
and multiplication with a positive constant.

4.

$$minimize\ (x-2)^2$$
$$\text{s}.t\ (x+3)(x-1) \leq 3$$
is equivalent to $x^2 - 2x - 6 \leq 0$
$$L(x,\lambda) = (x-2)^2 + \lambda(x^2 + 2x - 6),\ \lambda \geq 0$$
$$L(x,\lambda) = (1+\lambda)x^2 + (2\lambda - 4)x - 6\lambda + 4$$
$$\frac{\partial L(x,\lambda)}{\partial x} = 0$$
$$\Rightarrow (2+2\lambda)x + 2\lambda - 4 = 0$$
$$\Rightarrow x = \frac{2-\lambda}{1+\lambda}$$

Dual problem:

$$maximize\ \tilde{L}(\lambda) = \frac{\lambda^2 - 4\lambda + 4}{\lambda + 1} + \frac{2-\lambda}{1+\lambda}(2\lambda - 4) - 6\lambda + 4$$
$$= \frac{-\lambda^2 + 4\lambda - 4}{1+\lambda} - 6\lambda + 4$$
$$s.t.\ \lambda \geq 0$$