NYCU Pattern Recognition, HW4

310551079 梁友誠

Part. 1, Coding:

1. K-fold data partition:

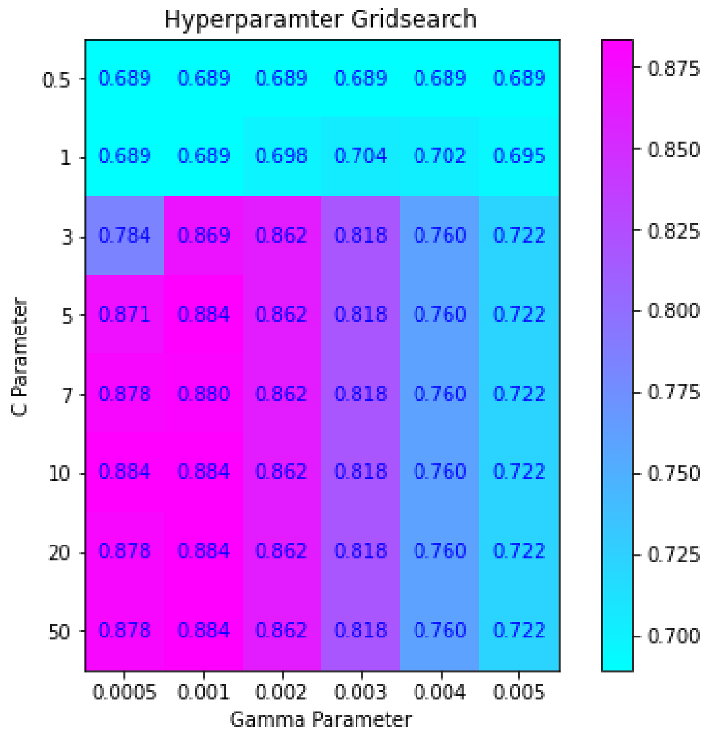
```
# test the cross validation with small data
X = np.arange(20)
_kfold_data = cross_validation(X, X, k=5)
for i, (traing_idx, validation_idx) in enumerate(_kfold_data):
    print("Split: %s, Training index: %s, Validation index: %s" % (i+1, traing_idx, validation_idx))
```

```
Split: 1, Training index: [ 9 19 15  5 10 14 16  3 13 17  6  8  2 12  1  7], Validation index: [18  4  0 11]
Split: 2, Training index: [18  4  0 11 10 14 16  3 13 17  6  8  2 12  1  7], Validation index: [ 9 19 15  5]
Split: 3, Training index: [18  4  0 11  9 19 15  5 13 17  6  8  2 12  1  7], Validation index: [10 14 16  3]
Split: 4, Training index: [18  4  0 11  9 19 15  5 10 14 16  3  2 12  1  7], Validation index: [13 17  6  8]
Split: 5, Training index: [18  4  0 11  9 19 15  5 10 14 16  3 13 17  6  8], Validation index: [ 2 12  1  7]
```

2. Grid Search & Cross-validation:

```
Best C=5, gamma=0.001
score: 0.8836363636363636
```

3. Plot the grid search results of your SVM.



4. Train your SVM model by the best hyperparameters.

```
# By several experiments, I find that use C=5 and gamma=0.002 can get the best result
best_model = SVC(C=5, kernel='rbf', gamma=0.002)
best_model.fit(x_train, y_train)

y_pred = best_model.predict(x_test)
print("Accuracy score: ", accuracy_score(y_pred, y_test))
```

```
Accuracy score:  0.9166666666666666
```

Part. 2, Questions:

Q1.

a. (1) According to $\underline{k(x, x') = k_1(x, x')\, k_2(x, x')}$ (6.18)

$k(x, x')$ is a valid kernel when $k_1(x, x')$ and $k_2(x, x')$ are valid kernels, so we can derive that

$(k_1(x, x'))^2$ is a valid kernel.

(2) According to $\underline{k(x, x') = k_1(x, x') + k_2(x, x')}$ (6.17)

$k(x, x')$ is a valid kernel when $k_1(x, x')$ and $k_2(x, x')$ are valid kernels, and any positive constant is a valid kernel, so we can derive that

$k_2(x, x') + 1$ is a valid kernel.

$(k_2(x, x') + 1)^2$ is also valid (by (6.18))

$\therefore$ $(k_1(x, x'))^2 + (k_2(x, x') + 1)^2$ is valid (by (6.17))

b. (1) According to (a.), we know $(k_1(x, x'))^2$ is valid.

(2) According to $\underline{k(x, x') = f(x) \exp(k_1(x, x'))\, f(x')}$ (6.14)

$\therefore$ $\underbrace{\exp(\|x\|^2)}_{f(x)} \times \underbrace{1}_{k_1(x, x')} \times \underbrace{\exp(\|x'\|^2)}_{f(x')}$ is a valid kernel

$\therefore$ $(k_1(x, x'))^2 + \exp(\|x\|^2) \times \exp(\|x'\|^2)$ is a valid kernel (by (6.17))

Q2.

Consider a finite input space $X = \{x_1, x_2, \dots x_m\}$ and Gram matrix $K = [K(x_n, x_m)]_{nm}$, K is symmetric and is a positive semidefinite matrix (PSD)

$\because$ K is symmetric, so we can decompose $K = V \Lambda V^T$

where V is an orthonormal matrix V_l and the diagonal matrix Λ contains the eigenvalues λ_l of K

K is positive semidefinite, so all eigenvalues are non-negative

Define a feature mapping into a m-dimensional space where the lth bit in feature expansion for example x_i is

$$\phi_l(x_i) = \sqrt{\lambda_l} (\vec{V_l})_i$$

The inner product is

$$\phi(x_i)^T \cdot \phi(x_j) = \sum_{l=1}^{m} \phi_l(x_i) \phi_l(x_j)$$
$$= \sum_{l=1}^{m} \lambda_l (V_l)_i (V_l)_j$$

We want to show that $K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$

$$\therefore K_{i,j} = [V \Lambda V^T]_{i,j} = [[V\Lambda] V^T]_{i,j}$$
$$[V\Lambda] = (\vec{V_1}, \vec{V_2}, \dots, \vec{V_m}) \Lambda$$
$$[V\Lambda]_{i,l} = (V_l)_i \lambda_l$$
$$[[V\Lambda]V^T]_{i,j} = \sum_{l=1}^{m} (V_l)_i \lambda_l (V_l)_j = \phi(x_i)^T \phi(x_j)$$

So we find that $K_{i,j} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j)$

Q3.

$$a_n = -\frac{1}{\lambda}\{w^T\phi(x_n) - t_n\}$$
$$= -\frac{1}{\lambda}\{w_1\phi_1(x_n) + w_2\phi_2(x_n) + \dots + w_m\phi_m(x_n) - t_n\}$$
$$= -\frac{w_1}{\lambda}\phi_1(x_n) - \frac{w_2}{\lambda}\phi_2(x_n) - \dots \frac{w_m}{\lambda}\phi_m(x_n) + \frac{t_n}{\lambda}$$
$$= (c_n - \frac{w_1}{\lambda})\phi_1(x_n) + (c_n - \frac{w_2}{\lambda})\phi_2(x_n) + \dots + (c_n - \frac{w_m}{\lambda})\phi_m(x_n)$$

where $c_n = \frac{t_n/\lambda}{\phi_1(x_n) + \phi_2(x_n) + \dots + \phi_m(x_m)}$

From what we have derived above, we can see that a_n is a linear combination of $\phi(x_n)$, what's more, we first substitute $K = \Phi\Phi^T$ into $J(a) = \frac{1}{2}a^T KKa - a^T Kt + \frac{1}{2}t^T t + \frac{\lambda}{2}a^T Ka$. and the we will obtain

$$J(a) = \frac{1}{2}a^T\Phi\Phi^T\Phi\Phi^T a - a^T\Phi\Phi^T t + \frac{1}{2}t^T t + \frac{\lambda}{2}a^T\Phi\Phi^T a \quad (6.5)$$

Next we substitute

$$w = -\frac{1}{\lambda}\sum_{n=1}^{N}\{w^T\phi(x_n) - t_n\}\phi(x_n) = \sum_{n=1}^{N} a_n\phi(x_n) = \Phi a$$

into (6.5), we will obtain

$$J(w) = \frac{1}{2}\sum_{n=1}^{N}\{w^T\phi(x_n) - t_n\}^2 + \frac{\lambda}{2}w^T w$$

just as required.

Q 4.

(1) $k(x, x') = \exp(-\|x - x'\|^2 / 2\sigma^2)$

expand $\|x - x'\|^2$, we get $\|x - x'\|^2 = x^T x + (x')^T x' - 2x^T x'$

$\therefore k(x, x') = \exp\left(\frac{-x^T x}{2\sigma^2}\right) \exp\left(\frac{x^T x'}{\sigma^2}\right) \exp\left(\frac{-(x')^T x'}{2\sigma^2}\right)$

According to $k(x, x') = \exp(k_1(x, x'))$ (6.16)

$\therefore \exp\left(\frac{x^T x'}{\sigma^2}\right)$ is valid kernel

and $k(x, x') = \underbrace{f(x)}_{\exp\left(\frac{-x^T x}{2\sigma^2}\right)} k_1(x, x') \underbrace{f(x')}_{\exp\left(\frac{-(x')^T x'}{2\sigma^2}\right)}$ (6.14)

So, $k(x, x') = \exp(-\|x - x'\|^2 / 2\sigma^2)$ is also a valid kernel

(2) We know that $k(x, x') = \exp\left(\frac{-x^2}{2\sigma^2}\right) \exp\left(\frac{-(x')^2}{2\sigma^2}\right) \exp\left(\frac{x \cdot x'}{\sigma}\right)$

Use taylor series to expand the $\exp\left(\frac{x \cdot x'}{\sigma}\right)$

So we get

$$k(x, x') = e^{\frac{-x^2}{2\sigma^2}} \cdot e^{\frac{-(x')^2}{2\sigma^2}} \cdot \left(1 + \sqrt{\frac{(\frac{1}{2\sigma^2})}{1!}}\, x \sqrt{\frac{(\frac{1}{2\sigma^2})}{1!}}\, x' + \sqrt{\frac{(\frac{1}{2\sigma^2})^2}{2!}}\, x^2 \sqrt{\frac{(\frac{1}{2\sigma^2})^2}{2!}}\, (x')^2 \cdots\right)$$

$$= e^{\frac{-x^2}{2\sigma^2}} \begin{bmatrix} 1 \\ \sqrt{\frac{(\frac{1}{2\sigma^2})}{1!}}\, x \\ \vdots \end{bmatrix}^T e^{\frac{-(x')^2}{2\sigma^2}} \begin{bmatrix} 1 \\ \sqrt{\frac{(\frac{1}{2\sigma^2})}{1!}}\, x' \end{bmatrix}^T = \phi(x)^T \phi(x')$$

$$\therefore \phi(x) = e^{\frac{-x^2}{2\sigma^2}} \left(1, \sqrt{\frac{(\frac{1}{2\sigma^2})}{1!}}\, x, \sqrt{\frac{(\frac{1}{2\sigma^2})^2}{2!}}\, x^2, \sqrt{\frac{(\frac{1}{2\sigma^2})^3}{3!}}\, x^3, \cdots\right)$$

where $x \in R^1$

Q 5.

We can rewrite the original problem to Lagrangian function to solve the optimization problem.

$L(x, \lambda) = f(x) + \lambda g(x)$, $g(x)$ is a constraint.

the constraint $g(x)$ is $(x+3)(x-1) \leq 2$

$\Rightarrow x^2 + 2x - 3 - 2 \leq 0$

$\therefore L(x, \lambda) = (x-2)^2 + \lambda(x^2 + 2x - 5)$

$= x^2 - 4x + 4 + \lambda x^2 + \lambda 2x - 5\lambda$

Now, we calculate the gradient of $L(x, \lambda)$

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

$\frac{\partial L(x,\lambda)}{\lambda} = x^2 + 2x - 5 = 0$ ... original constraint

Thus, we can rewrite $L(x,\lambda)$ to $g(\lambda)$ with substitute λ for x

$\therefore g(\lambda) = \frac{-\lambda^2 - 4\lambda + 12}{1+\lambda} + 4 - 5\lambda$ ... Lagranian dual function

We want to minimize the original problem, which means we can maximize the corresponding dual problem

So the dual problem is :

$$\text{maximize} \quad \frac{-\lambda^2 - 4\lambda + 12}{1+\lambda} + 4 - 5\lambda$$
$$\text{subject} \quad \lambda \geq 0$$

#