NYCU Pattern Recognition, Homework 4

311552034, 林柏翰

Part. 1, Coding (50%):

1. (10%) Implement K-fold data partitioning.

```
def cross_validation(x_train, y_train, k=5):
    num_samples=len(x_train)
    indice=np.arange(num_samples)
    np.random.shuffle(indice)
    fold_sizes=np.full(k,num_samples//k)
    fold_sizes[:num_samples%k]+=1
    current=0
    k_folds=[]

    for fold_size in fold_sizes:
        validation_indice=indice[current:fold_size]
        training_indice=np.concatenate([indice[:current],indice[current+fold_size:]])
        k_folds.append([training_indice,validation_indice])
        current+=fold_size

    return k_folds
```

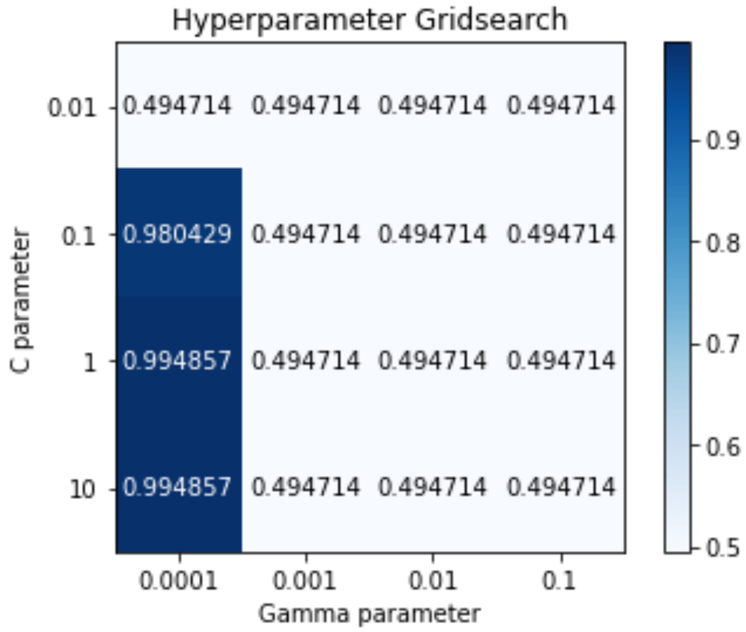
2. (10%) Set the kernel parameter to 'rbf' and do grid search on the hyperparameters **C** and. **gamma** to find the best values through cross-validation. Print the best hyperparameters y ou found. Note that we suggest using K=5 for the cross-validation.

```
best_c, best_gamma = None, None
kfold_data = cross_validation(x_train, y_train, k=5)
c_list=[0.01,0.1,1,10]
gamma_list=[0.0001,0.001,0.01,0.1]
acc_score=np.zeros((len(c_list),len(gamma_list)))
best_acc=0

for i in range(len(c_list)):
    for j in range(len(gamma_list)):
        score=0
        for train_idx ,val_idx in kfold_data:
           clf=SVC(C=c_list[i],gamma=gamma_list[j],kernel='rbf')
           clf.fit(x_train[train_idx],y_train[train_idx])
           score+=clf.score(x_train[val_idx],y_train[val_idx])

        acc_score[i][j]=score/5
        if acc_score[i][j]>best_acc:
            best_c=c_list[i]
            best_gamma=gamma_list[j]
            best_acc=acc_score[i][j]
# TODO HERE
# k-Flod Cross Validation and Grid Search
best_parameters=(best_c, best_gamma)
```

```
print("(best_c, best_gamma) is ", best_parameters)
```

```
(best_c, best_gamma) is  (1, 0.0001)
```

3. (10%) Plot the results of your SVM's grid search. Use "gamma" and "C" as the x and. y axes, respectively, and represent the average validation score with color. Below image is just for reference.



4. (20%) Train your SVM model using the best hyperparameters found in Q2 on the entire training dataset, then evaluate its performance on the test set. Print your testing accuracy.

```
# Do Not Modify Below

best_model = SVC(C=best_parameters[0], gamma=best_parameters[1], kernel='rbf')
best_model.fit(x_train, y_train)

y_pred = best_model.predict(x_test)

print("Accuracy score: ", accuracy_score(y_pred, y_test))

# If your accuracy here > 0.9 then you will get full credit (20 points).
```

```
Accuracy score:  0.995
```

Part. 2, Questions (50%):

1. (10%) Show that the kernel matrix $K = [k(x_n, x_m)]_{nm}$ should be positive semidefinite is the necessary and sufficient condition for $k(x, x')$ to be a valid kernel.

2.

$\Rightarrow$ given that $K = [k(x_n, x_m)]_{nm}$ is positive semidefinite

let $K = V \Lambda V^T$ be the eigenvector decomposition of K

V: eigenvector matrix

Λ: eigenvalue matrix

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

$= K_{ij} = k(x_i, x_j)$

$\Rightarrow k(x_i, x_j)$ is valid if K is positive semidefinite #

$\Leftarrow$ given that $k(x, x')$ is valid

則 $k_{ij} = k(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \phi(x_j)^T \phi(x_i) = k_{ji}$

$\Rightarrow K$ 為 symmetric

取任意向量 z

$$z^T K z = \sum_i \sum_j z_i K_{ij} z_j$$

$$= \sum_i \sum_j z_i \phi(x_i)^T \phi(x_j) z_j$$

$$= \sum_i \sum_j z_i \sum_k \phi(x_i) \phi(x_j) z_j$$

$$= \sum_k \sum_i \sum_j z_i \phi(x_i) \phi(x_j) z_j$$

$$= \sum_k \left[\sum_i z_i \phi_k(x_i) \right]^2 \geq 0$$

∵ z 為任意向量，

∴ K 為 positive semidefinite #

2. (10%) Given a valid kernel $k_1(x, x')$, explain that $k(x, x') = exp(k_1(x, x'))$ is also a valid kernel. (Hint: Your answer may mention some terms like ____ series or ____ expansion.)

2. from Taylor's expansion

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

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

For positive constant 1, it's apparent that it's symmetric as well as PSD, which means that 1 is a valid kernel.

Given $k_1(x, x')$ is valid, so for the remaining part, we can find that they are all valid according to (6.13) and (6.17). Thus, from (6.17), the whole expansion is valid.

$\Rightarrow$ $exp(k_1(x, x'))$ is valid if $k_1(x, x')$ is valid

3. (20%) Given a valid kernel $k_1(x, x')$, prove that the following proposed functions are or are not valid kernels. If one is not a valid kernel, give an example of $k(x, x')$ that the corresponding K is not positive semidefinite and show its eigenvalues.

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

b. $k(x, x') = k_1(x, x') - 1$

3.

(a)

取 $k_1 = x^T x'$ (valid kernel)

$\Rightarrow k = x^T x' + x$ ，取 $x_1 = 0$ $x_2 = -\frac{1}{2}$

$\Rightarrow$ gram matrix $= \begin{bmatrix} 0 & 0 \\ -\frac{1}{2} & \frac{1}{-4} \end{bmatrix}$

$\Rightarrow$ eigenvalue $= 0, -\frac{1}{4}$ $\Rightarrow$ 非 PSD

$\Rightarrow k(x, x') = k_1(x, x') + x$ 非 valid kernel

(b) 取 $k_1(x, x')$ 為 $x^T x'$ $\Rightarrow k(x, x') = x^T x' - 1$

$x_1 = 1$ $x_2 = -1$

gram matrix $= \begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix}$ eigenvalue 為 $2, -2$

非 PSD $\Rightarrow k(x, x') = k_1(x, x')$

非 valid kernel

c. $k(x, x') = k_1(x, x')^2 + exp\left(\|x\|^2\right) * exp(\|x'\|^2)$

3.

(c) $k_1(x, x')$ valid → from 6.18 $k_1(x, x') k_1(x, x')$ valid

so all we have to prove is $exp(\|x\|^2) * exp(\|x'\|^2)$

is also a valid kenel. so that from (6.17)

we can prove that $k_1(x, x')^2 + exp(\|x\|^2) * exp(\|x'\|^2)$

is also valid

according to 6.14: $f(x) k_1(x, x') f(x')$ is valid

now we set $f(x) = exp(\|x\|^2)$.

$\Rightarrow$ $f(x) * 1 * f(x')$ is valid

↳ positive constant is valid

$\Rightarrow$ $exp(\|x\|^2) * exp(\|x'\|^2)$ is valid

$\Rightarrow$ Q.E.D #

d. $k(x, x') = k_1(x, x')^2 + exp\left(k_1(x, x')\right) - 1$

3.

(d) 已知 $k_1(x, x')$ 為 valid kernel

$$k(x, x') = \underbrace{k_1(x, x')}_{①}{}^2 + \underbrace{exp(k_1(x, x')) - 1}_{②}$$

① is valid kernel according to 6.18

② $exp(k_1(x, x')) - 1$

Taylor expansion

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

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

$\because$ $k_1(x, x')$ valid $\Rightarrow$ from 6.13, 6.17, 6.18

② is also valid

$\because$ ①, ② valid $\Rightarrow$ from 6.17, $k(x, x') = ① + ②$ valid #

4. Consider the optimization problem

$$\textit{minimize } (x - 2)^2$$
$$\textit{subject to } (x + 4)(x - 1) \leq 3$$

State the dual problem. (Full points by completing the following equations)

$L(x, \lambda) =$ ___ (x - 2)^2 + λ[(x + 4)(x - 1) – 3]_____________

$\nabla_x L(x, \lambda) =$ _____ 2(x - 2) + λ(2x + 3)________________

when $\nabla_x L(x, \lambda) = 0$,

$x =$ _____ (4 - 3λ)/(2 + 2λ)___________

$L(x, \lambda) = L(\lambda) =$

[(4 - 3λ) / (2 + 2λ) - 2]^2 + λ{[(4 - 3λ) / (2 + 2λ)) + 4][((4 - 3λ) / (2 + 2λ)) – 1] – 3}