NCTU Pattern Recognition, Homework 4

Deadline: May 25, 23:59

Part. 1, Coding (50%):

1. (10%) K-fold data partition:

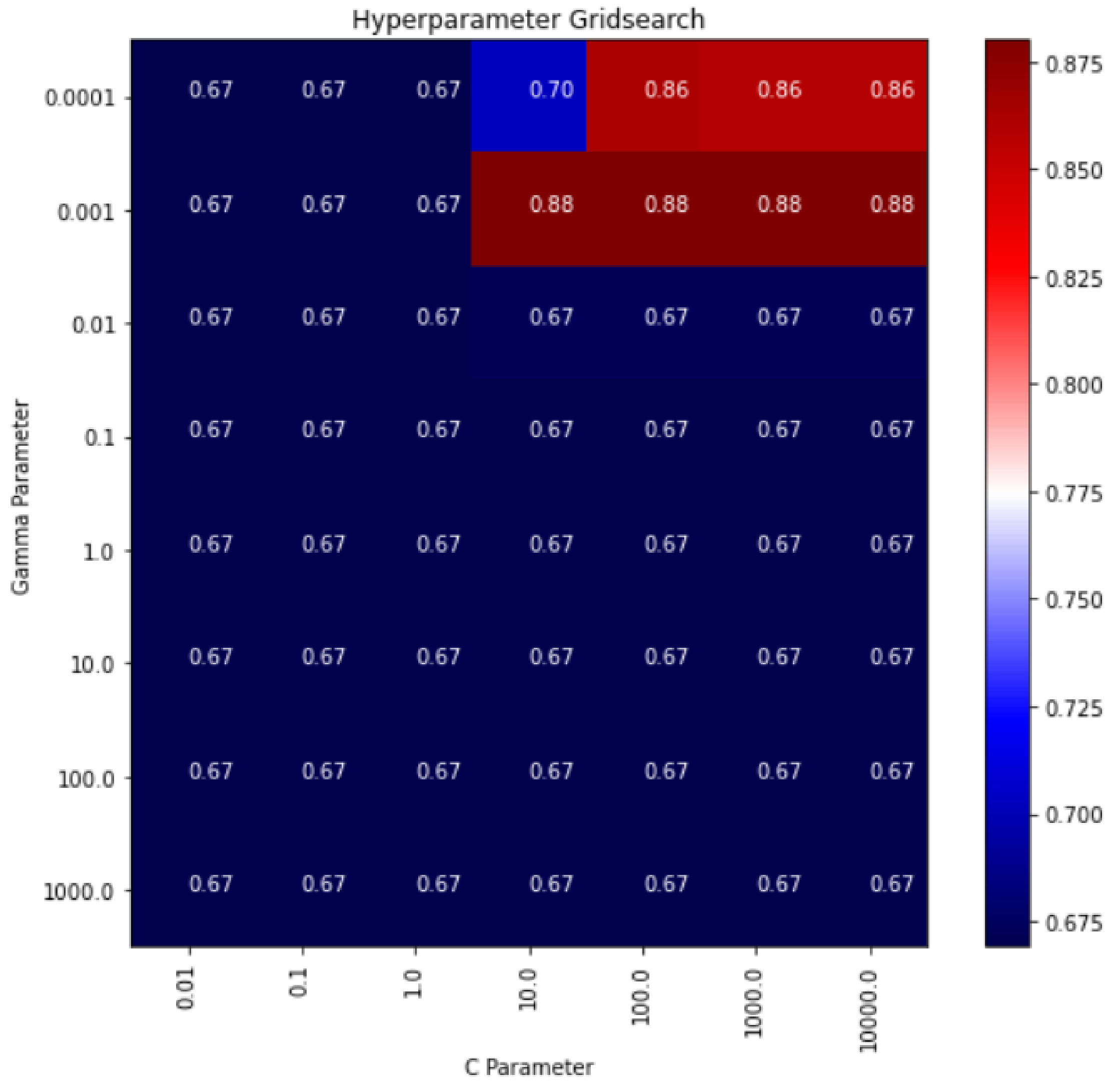
```
def cross_validation(x_train, y_train, k=5):
    x_train = np.asarray(x_train)
    y_train = np.asarray(y_train).reshape(-1,1)
    kfold_data = []
    training_data = np.concatenate((x_train,y_train), axis = 1)
    for i in range(k):
        kf = x_train.shape[0] - x_train.shape[0]//k
        split_num = x_train.shape[0]//k
        #print(split_num)
        split_list = [z for z in range(len(x_train))]  # [0,1,2,3,4,5,6,7....]
        split_list = random.sample(split_list, split_num) # [choose from split_list]
        #print(split_list)
        Validation = training_data[split_list, :]
        #print(Validation.shape)
        #print(Validation)
        Training = np.delete(training_data, split_list, axis = 0)
        Training = Training.reshape(-1, 301)
        #print(Training.shape)
        kfold_data.append([Training, Validation])
    return kfold_data
```

2. (20%) Grid Search & Cross-validation:

```
print(best_parameters)
all_parameters = all_parameters.reshape(-1, 7)
print(best_score)
```

```
{'gamma': 0.001, 'C': 10.0}
0.8799999999999999
```

3. (10%) Plot the grid search results of your SVM.



4. (10%) Train your SVM model by the best hyperparameters you found from question 2 on the whole training data and evaluate the performance on the test set.

find the best parameters from the whole dataset

```
best_score = 0
all_parameters = []
best_gamma = 0
best_C = 0
for gamma in [0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0]:
    for C in [0.01, 0.1, 1.0, 10.0, 100.0, 1000.0, 10000.0]:
        svm = SVC(gamma = gamma, kernel = 'rbf', C = C)
        svm.fit(x_train, y_train)
        score = svm.score(x_test, y_test)
        all_parameters.append(score)
        if score > best_score:
            best_score = score
            best_gamma = gamma
            best_C = C
            best_parameters = {'gamma' : best_gamma, 'C' : best_C}

best_model = SVC(C = best_C, kernel = 'rbf', gamma = best_gamma)
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:  0.90625
```

using the parameters found from question 2

```
best_model = SVC(C = 10.0, kernel = 'rbf', gamma = 0.001)
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:  0.8958333333333334
```

Part. 2, Questions (50%):

1. Given $k_1(x,x')$ is valid

(a) prove $k(x,x') = (k_1(x,x'))^2 + (k_1(x,x')+1)^2$

according to $k(x,x') = k_1(x,x')k_2(x,x')$ is valid

$\Rightarrow k(x,x') = k_1(x,x') \cdot k_1(x,x') = (k_1(x,x'))^2$ is valid ①

---

according to $k(x,x') = g(k_1(x,x'))$ is valid

$\Rightarrow k(x,x') = k_1(x,x') + 1$ (where $g(x) = x+1$) is valid

similar with ① $\Rightarrow k_2(x,x') = (k_1(x,x')+1)^2$ is valid ②

---

according to $k(x,x') = k_1(x,x') + k_2(x,x')$ is valid

by ①, ② $= (k_1(x,x'))^2 + (k_1(x,x')+1)^2$ is valid

*

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

according (a)① $\Rightarrow$ we can find that $(k_1(x,x'))^2$ is valid ①

$\begin{cases} k_3(x,x') = \|x\|^2 \\ k_3(x',x') = \|x'\|^2 \end{cases}$

from $k(x,x') = \exp(k_1(x,x'))$ is valid

$\exp(\|x\|^2) = \exp(k_3(x,x'))$ is valid

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

$k(x,x')$ ① $= k_2(x,x') + k_4(x,x')$ is valid

$\Rightarrow k(x,x') = (k_1(x,x'))^2 + \exp(\|x\|^2) \times \exp(\|x\|^2)$ is valid

*

2. Consider the general case of kernel matrix,

let $k(x,x) = G_{ij} = k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, for $i, j \ldots$

$\phi : x_i \rightarrow (\sqrt{\lambda_t} \, V_{ti})_{t=1} \in \mathbb{R}^n$

eigen values ≥ 0 (when kernel matrix is positive

in (-∞), eigen values should ≥ 0

so, kernel matrix is positive semidefinite

3.

$$a_n = -\frac{1}{\lambda}(w^T\phi(x_n) - t_n)$$
$$= \frac{-1}{\lambda}(w_1\phi_1(x_n) + w_2\phi_2(x_n) + \ldots + w_N\phi_N(x_n) - t_n)$$
$$= \frac{-w_1}{\lambda}\phi_1(x_n) - \frac{w_2}{\lambda}\phi_2(x_n) - \ldots - \frac{w_N}{\lambda}\phi_N(x_n) + \frac{t_n}{\lambda}$$
$$= (b_n - \frac{w_1}{\lambda})\phi_1(x_n) + (b_n - \frac{w_2}{\lambda})\phi_2(x_n) + \ldots + (b_n - \frac{w_N}{\lambda})\phi_N(x_n)$$

$$b_n = \frac{t_n/\lambda}{\phi_1(x_n) + \phi_2(x_n) + \ldots + \phi_N(x_n)}$$

From 課本(6.7)

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

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

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

4. ① $k(x,x') = \exp(-\|x-x'\|^2/2\sigma^2) = \phi(x^T \phi(x')$

$|x-x'|^2 = x^T x + (x')^T x' - 2x^T x'$

and $k(x,x') = \exp(-x^T x/2\sigma^2)\exp(x^T x'/\sigma^2)\exp(-(x')^T x'/2\sigma^2)$

and $k(x,x') = \exp(k_1(x,x'))$

$k(x,x') = f(x)\, k_1(x,x') f(x')$

so $k(x,x')$ is valid ※

②

$\exp(-(x-x')^2/2\sigma^2) = e^{\left(\frac{-x^2}{2\sigma^2} - \frac{(x')^2}{2\sigma^2}\right)} \times \left(1 + \frac{2xx'}{1! \cdot 2\sigma^2} + \frac{(2xx')^2}{2! \cdot (2\sigma^2)^2} + \cdots\right)$

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

$= \phi(x)^T \phi(x')$

$\Rightarrow \phi(x) = e^{\left(\frac{-x^2}{2\sigma^2}\right)}\left[1, \frac{\sqrt{2}}{\sqrt{2}\sigma}x, \frac{\sqrt{2^2}}{\sqrt{2! \cdot (2\sigma^2)^2}}x^2, \ldots\right]^T$ ※

5、

$(x+3)(x-1) \le 2 \Rightarrow x^2+2x-3 \le 2 \Rightarrow x^2+2x-7 \le 0$

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

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

$= (1+\lambda)x^2 + (-4+2\lambda)x + (4-5\lambda)$

$\frac{\partial L}{\partial x} = (2+2\lambda)x - 4 + 2\lambda$

$\frac{\partial L}{\partial \lambda} = 0$ when $x = \frac{2-\lambda}{1+\lambda} \longrightarrow L(x,\lambda)$

$g(\lambda) = \frac{-4+4\lambda-\lambda^2}{1+\lambda} + (4-5\lambda)$ $\therefore$ dual problem: $\max \frac{-4+4\lambda-\lambda^2}{1+\lambda} + (4-5\lambda)$

subject to $\lambda \ge 0$ ※