NCTU Introduction to Machine Learning, Homework 4

109550100 陳宇駿

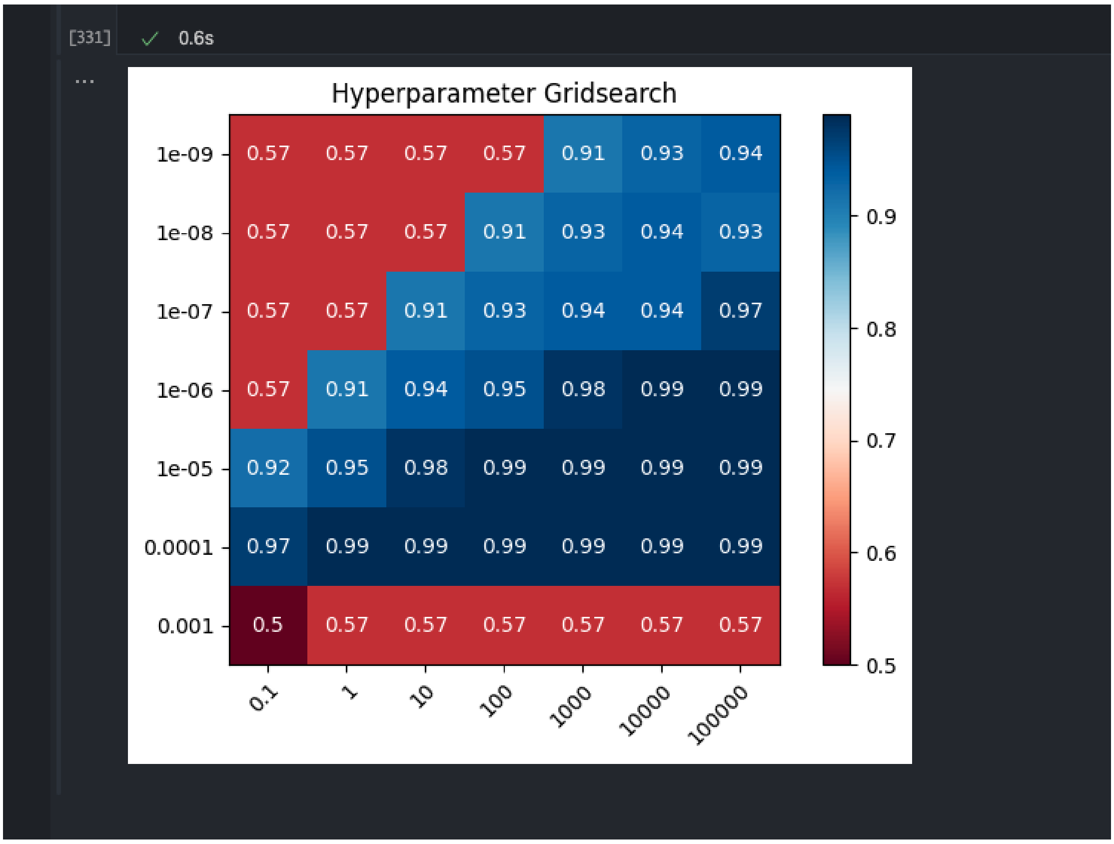
Part. 1, Coding (50%):

1. (10%) K-fold data partition: Implement the K-fold cross-validation function. Your function should take K as an argument and return a list of lists (*len(list) should equal to K*), which contains K elements. Each element is a list containing two parts, the first part contains the index of all training folds (index_x_train, index_y_train), e.g., Fold 2 to Fold 5 in split 1. The second part contains the index of the validation fold, e.g., Fold 1 in split 1 (index_x_val, index_y_val)

2. (20%) Grid Search & Cross-validation: using sklearn.svm.SVC to train a classifier on the provided train set and conduct the grid search of “C” and “gamma,” “kernel’=’rbf’ to find the best hyperparameters by cross-validation. Print the best hyperparameters you found.
Note: We suggest using K=5

```
best_parameters = [bestC, bestgamma]
print(bestscore)
print(best_parameters)
```
```
[310]  ✓ 0.9s
```
```
0.9934285714285714
[1, 0.0001]
```

3. (10%) Plot the grid search results of your SVM. The x and y represent "gamma" and "C" hyperparameters, respectively. And the color represents the average score of validation folds.



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.

Accuracy	**Your scores**
acc > 0.9	**10points**
0.85 <= acc <= 0.9	**5 points**
acc < 0.85	**0 points**

Part. 2, Questions (50%):

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

1. $K = [k(x_n, x_m)]_{n\times m}$

Since K is symmetric, we have $K = V \Lambda V^T$,

where V is an orehogonal matrix and Λ is a diagonal matrix containing K's eigenvalues.

If K is positive semidefinite, all eigenvalues ≥ 0

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

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

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

2. $k(x, x') = \exp(k_1(x, x'))$

$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ (taylor expansion)

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

by 6.13 ($c k_1(x, x')$ is valid, where c is a const. > 0)

& 6.17 (addition of valid kernels is a valid kernel)

& 6.18 (multiplication of valid kernels is a valid kernel)

$1 + k_1(x, x') + \frac{1}{2} k_1(x, x') k_1(x, x') + \frac{1}{6} k_1(x, x') k_1(x, x') k_1(x, x') \ldots$ is valid

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

(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. $k(x, x') = k_1(x, x') + 1$

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

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

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

3. a. let $k_2(x, x') = 1 \Rightarrow K = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}_{n \times m}$

$\Rightarrow$ eigenvalue = n or $0 \geq 0$

$\Rightarrow K$ is positive semidefinite

$\Rightarrow k_2(x, x')$ is a valid kernel

both $k_1(x, x')$ and $k_2(x, x')$ are valid kernels

$\Rightarrow k_1(x, x') + k_2(x, x')$ is a valid kernel

$\Rightarrow k(x, x') = k_1(x, x') + 1$ is a valid kernel.

b. let $k_2(x, x') = -1 \Rightarrow K = \begin{bmatrix} -1 & \cdots & -1 \\ \vdots & \ddots & \vdots \\ -1 & \cdots & -1 \end{bmatrix}_{n \times m}$

$\Rightarrow$ eigenvalue < 0

$\Rightarrow K$ is not positive semidefinite

$\Rightarrow k_2(x, x')$ is not a valid kernel

$\Rightarrow k(x, x') = k_1(x, x') + k_2(x, x')$ is not a valid kernel

$\Rightarrow k(x, x') = k_1(x, x') - 1$ is not a valid kernel

Counter example:

let $K = \begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}$. K has eigenvalue 5, 1 (positive semidefinite)

$K - 1 \cdot I = \begin{bmatrix} 2 & -3 \\ -3 & 2 \end{bmatrix}$ has eigenvalue 5, -1 (not positive semidefinite)

3. c. let $f(x) = \exp(\|x\|^2)$, $\underbrace{k_2(x,x') = 1}_{\text{is valid (by 3.a)}}$

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

$$= \underbrace{k_1(x,x')k_1(x,x')}_{\text{valid (by Ch.6, 6.18)}} + \underbrace{f(x)k_2(x,x')f(x')}_{\text{valid (by Ch.6 6.14)}}$$

$\therefore$ $k(x,x')$ is a valid kernel

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

$$= \underbrace{k_1(x,x')^2}_{\text{valid (by 6.18)}} + \left(1 + k_1(x,x') + \frac{1}{2}k_1(x,x')^2 + \cdots\right) - 1$$

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

$$= k_1(x,x')^2 + \underbrace{\sum_{n=1}^{\infty} \frac{k_1(x,x')^n}{n!}}_{\text{valid (proven in question 2., by 6.13, 6.17, 6.18)}}$$

$\therefore$ $k(x,x')$ is a valid kernel.

(10%) Consider the optimization problem

$$minimize\ (x - 2)^2$$
$$subject\ to\ (x + 3)(x - 1) \leq 3$$

State the dual problem.

4. let $f_1(x) = (x-2)^2$, $f_2(x) = (x+3)(x-1) - 3$

the original question turns into minimizing $f_1(x)$ subject to $f_2(x) \leq 0$

$$L(x, \lambda) = f_1(x) + \lambda f_2(x)$$

$$= (x-2)^2 + \lambda(x+3)(x-1) - 3\lambda$$

$$\frac{\partial L}{\partial x} = 2(x-2) + \lambda(x-1) + \lambda(x+3)$$

$$= \lambda(2x+2) + 2(x-2)$$

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

$$= (2\lambda+2)x + 2\lambda - 4 = 0$$

$$\Rightarrow \quad x = \frac{2-\lambda}{\lambda+1}$$

$$\phi(x) = \left(\frac{2-\lambda}{\lambda+1}\right)^2 + \lambda\left(\frac{2-\lambda}{\lambda+1} + 3\right)\left(\frac{2-\lambda}{\lambda+1} - 1\right) - 3\lambda$$

$\therefore$ dual problem: maximizing $\left(\frac{2-\lambda}{\lambda+1}\right)^2 + \lambda\left(\frac{2-\lambda}{\lambda+1} + 3\right)\left(\frac{2-\lambda}{\lambda+1} - 1\right) - 3\lambda$

subject to $\lambda \leq 0$