Report of Intro to Machine Learning, Homework 4

109550027 紀竺均

Part. 1, Coding (50%):

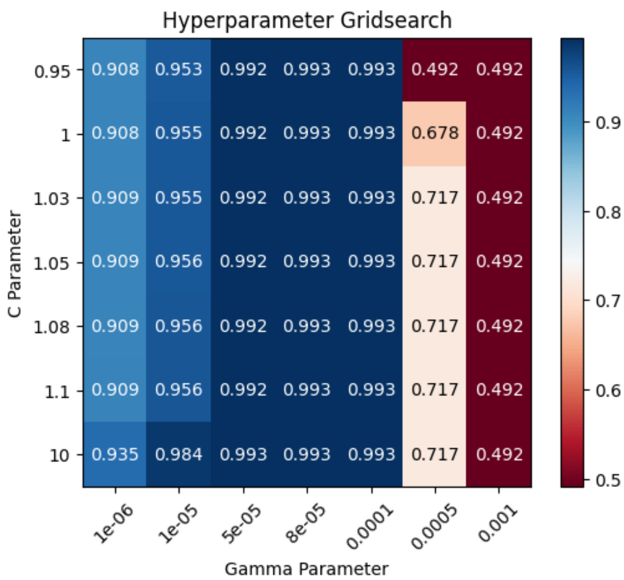
In this coding assignment, you need to implement the cross-validation and grid search using only NumPy, then train the [SVM model from scikit-learn] on the provided dataset and test the performance with testing data.

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.

```
print(best_parameters)
print(bestacc)
```

```
(0.95, 0.0001)
0.9934285714285714
```

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.

Part. 2, Questions (50%):

(10%) Show that the kernel matrix K=kxn,xmnm should be positive semidefinite is the necessary and sufficient condition for k(x,x') to be a valid kernel.

1) K is positive semidefinite $\xrightarrow{①} K = V\Lambda V^T$

$\xrightarrow{②}$ all eigenvalues ≥ 0

we can find the mapping function:

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

mapping function exist $\rightarrow$ kernel is valid.

② prove if $k(x,x')$ is valid $\rightarrow$ K is positive semidefinite

$$z^T k z = \sum_{n=1}^{N}\sum_{m=1}^{N} z_n k_{nm} z_m$$

$$= \sum_{n=1}^{N}\sum_{m=1}^{N} z_n \phi(x_n)^T \phi(x_m) z_m$$

$$= \sum_{n=1}^{N}\sum_{m=1}^{N} z_n \left(\sum_{k=1}^{N} \phi_k(x_n)\phi_k(x_m)\right) z_m$$

$$= \sum_{n=1}^{N}\sum_{m=1}^{N}\sum_{k=1}^{N} z_n \phi_k(x_n)\phi_k(x_m) z_m$$

$$= \sum_{k=1}^{N}\left[\sum_{n=1}^{N} z_n \phi_k(x_n)\right]^2 \geq 0 \rightarrow \text{K is positive semidefinite}$$

(10%) Given a valid kernel $k1(x,x')$, explain that $kx, x' = \exp(k1x,x')$ is also a valid kernel. Your answer may mention some terms like ____ series or ____ expansion.

2. $\exp(x) = \lim_{i\to\infty} \left(1 + x + \dots + \frac{x^i}{i!}\right)$

$k(x,x') = \exp(k_1(x,x')) = \lim_{i\to\infty} 1 + k_1(x,x') + \frac{(k_1(x,x'))^2}{2} + \dots + \frac{(k_1(x,x'))^i}{i!}$

↳ infinite series

→ 第j項 $\frac{(k_1(x,x'))^j}{j!}$ is still valid (from 6.13 6.18)

↳ 每一項都是 valid kernel → the addition of infinit series is still valid (from 6.17)

⇒ the exp of valid kernel k1 is a valid kernel.

(20%) Given a valid kernel k1(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. kx, x'= k1x, x'+1

b. kx, x'= k1x,x'-1

c. kx, x'= k1x,x'2+exp‖x‖2*exp(‖x'‖2)

d. kx, x'= k1x,x'2+expk1x,x'-1

3.(a). Let $k_2(x,x')=1$

$K_{nm}=k(x_n,x_m) \Rightarrow$ every element in $K=1$

$\rightarrow \lambda$ of $K_{n\times n}=0$ or n $\rightarrow$ positive semidefinite

$\rightarrow k_2(x,x')=1$ is a valid kernel

$\rightarrow k(x,x')=k_1(x,x')+1$ is valid (from 6.17)

(b) Let $k_2(x,x')=-1$

$K_{nm}=k(x_n,x_m) \Rightarrow$ every element in $K=-1$

$\rightarrow \lambda$ of $K_{n\times n}=0$ or $-n$ $\rightarrow$ NOT positive semidefinite

$\rightarrow k_2(x,x')=-1$ is NOT a valid kernel

$\rightarrow k(x,x')=k_1(x,x')-1$ is NOT valid.

(c) ① $k_1(x,x')^2=k_1(x,x')\cdot k_1(x,x')$ is a valid kernel (from 6.18)

② Let $k_2(x,x')=\exp(\|x\|^2)\times 1 \times \exp(\|x'\|)^2$

$\boxed{\text{Let } \begin{cases} f(x)=\exp(x^2) \\ k_3(x,x')=1 \end{cases}}$

$=f(x)\cdot k_3(x,x')\cdot f(x')$

$\because k_3$ is valid kernel (from a)

$\therefore k_2$ is valid kernel (from 6.14)

from ①, ②, (6.17) $\rightarrow k(x,x')=k_1(x,x')+k_2(x,x')$ is a valid kernel

(10%) Consider the optimization problem

minimize x- 22

subject to x+3x−1 3

State the dual problem.

(d) ① $k_1(x,x')^2 = k_1(x,x') \cdot k_1(x,x')$ is valid (from 6.18)

② $k_2(x,x') = \exp(k_1(x,x'))$ is valid (from 6.16)

③ Let $k_3(x,x') = -1$ is not semi-definite (from (b))

with $\lambda = 0$ or $-n$ → not a valid kernel

⇒ $k(x,x') = k_1(x,x')^2 + \exp(k_1(x,x')) - 1$ is not valid.

4. $L = (x-2)^2 + \lambda[(x+3)(x-1)-3]$

$= x^2 - 4x + 4 + \lambda x^2 + 2\lambda x - 6\lambda$

$= (\lambda+1)x^2 + (2\lambda-4)x + 4 - 6\lambda$

$\frac{L}{\partial x} = 2(\lambda+1)x + (2\lambda-4) \overset{\text{令}}{=} 0$

$x = \frac{4-2\lambda}{2\lambda+2} = \frac{2-\lambda}{\lambda+1}$ (代入 L)

$L = (\lambda+1)\left(\frac{2-\lambda}{\lambda+1}\right)^2 + (2\lambda-4)\cdot\frac{2-\lambda}{\lambda+1} + 4 - 6\lambda$

$= \frac{4-4\lambda+\lambda^2}{\lambda+1} + \frac{-8+8\lambda-2\lambda^2}{\lambda+1} + \frac{4-2\lambda-6\lambda^2}{\lambda+1} = \frac{-7\lambda^2+2\lambda}{\lambda+1}$

⇒ dual problem: Maximize $\frac{-7\lambda^2+2\lambda}{\lambda+1}$ subject to $\lambda \geq 0$