NCTU Introduction to Machine Learning, Homework 4

Deadline: Nov. 29, 23:59

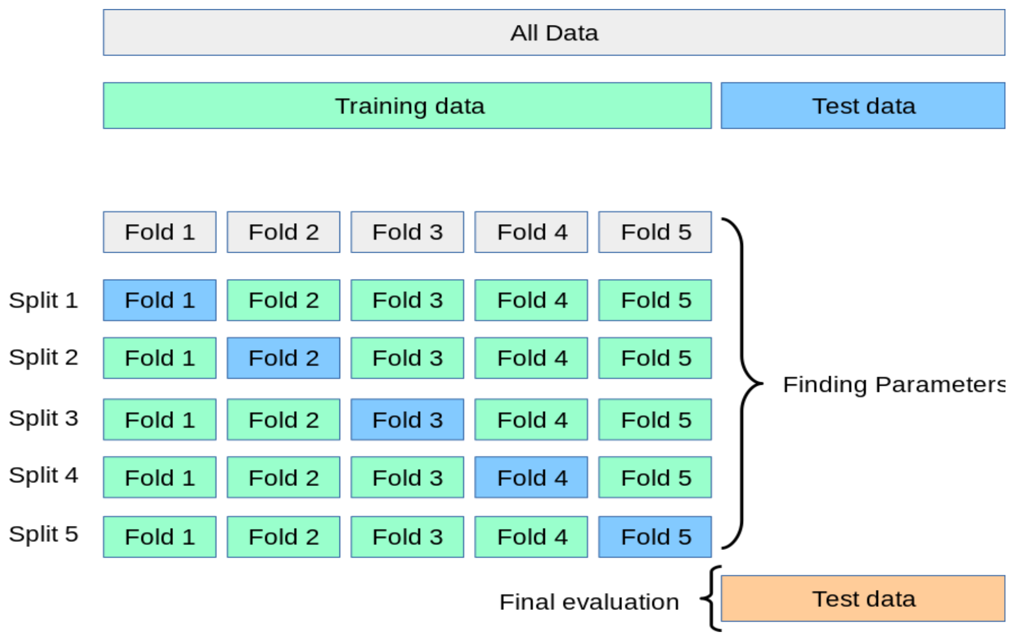
Part. 1, Coding (50%):

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. Find the sample code and data on the GitHub page https://github.com/NCTU-VRDL/CS_AT0828/tree/main/HW4

Please note that only NumPy can be used to implement cross-validation and grid search. You will get no points by simply calling sklearn.model_selection.GridSearchCV.

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)

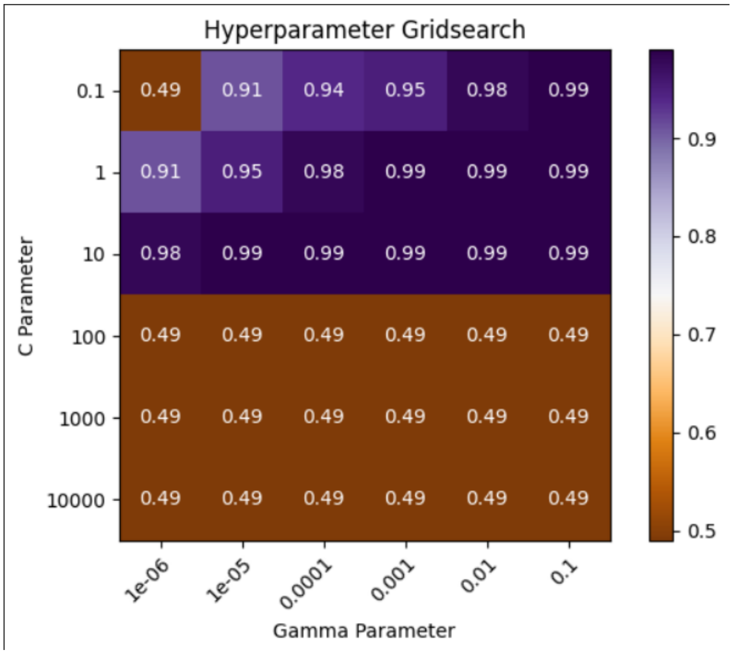
   Note: You need to handle if the sample size is not divisible by K. Using the strategy from sklearn. The first `n_samples % n_splits` folds have size `n_samples // n_splits + 1`, other folds have size `n_samples // n_splits`, where `n_samples` is the number of samples, `n_splits` is K, `%` stands for modulus, `//` stands for integer division. See this post for more details
   Note: Each of the samples should be used **exactly once** as the validation data
   Note: Please **shuffle** your data before partition



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

   ```
   [0.0001, 1]
   ```

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.

設 K = valid kernel matrix 對應 feature map ϕ

$\Rightarrow K(x, x') = K(\phi(x_i), \phi(x_j))$

$\Rightarrow K_{ij} = K(x^{(i)}, x^{(j)}) = \phi(x(i))^T \phi(x(j)) = \phi(x(j))^T \phi(x(i))$

$= K(x^{(j)}, x^{(i)}) = K_{ji} \Rightarrow K$: symmetric matrix.

let $\phi_k(x)$ = k^{th} element in vector $\phi(x)$, α = 任意向量.

$$\alpha^T K \alpha = \sum_i \sum_j \alpha_i K_{ij} \alpha_j = \sum_i \sum_j \alpha_i \underbrace{\phi(x(i))^T \phi(x(j))}_{= \sum_k \phi_k(x(i)) \phi_k(x(j))} \alpha_j$$

$$= \sum_k \sum_i \sum_j \alpha_i \phi_k(x(i)) \phi(x(j)) \alpha_j$$

metric)

$$\sum_k \overbrace{\left(\sum_i \alpha_i \phi_k(x(i))\right)^2}^{\geq 0} \geq 0.$$

、K is symmetric $\Rightarrow K$ is positive semidefinite.

ary & sufficient condition for $K(x, x')$

valid kernel.

(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.

Taylor expansion around 0:

$$exp(k) = exp(0) + exp(0)k + \frac{exp(0)}{2!}k^2 + \frac{exp(0)}{3!}k^3 + \cdots$$

$$exp(k) = 1 + k + \frac{k^2}{2} + \frac{k^3}{6} + \cdots$$

↳ exponetial of a kernel = polynomial serie with non-negative coefficients.

exp(k)

⇒ known $k(x, x')$ is valid when $k(x, x') = q(k_1(x, x'))$, where $q(\cdot)$ is a polynomial with non-negative coeff

⇒ prove that $k(x, x') = exp(k(\quad'))$ -

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

a. let $q(m) = m+1 \Rightarrow$ polynomial w neg ve coefficient

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

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

$(x\ x') = k(x, x') + 1$ is a valid kernel #

b. f = any real valued function f., $x, y \in \mathbb{R}^n$

$\sum_i \sum_j k(x_i, x_j) c_i c_j = \sum_i \sum_j f(x_i) f(x_j) c_i c_j = \left(\sum_i f(x_i) c_i\right)^2 \geq 0$

$\Rightarrow k(x, x') = f(x) f(x')$ is a valid kernel.

let function $f(m) = m$.

$k_1(x, x') = f(x) f(x') = (x \cdot x') \Rightarrow k(x, x') = \underbrace{(x * x')}_{k_1} - 1$

$S = \sum_{i,j=1}^{n} a_i a_j \underline{k(x_i, x_j)} \geq 0$

$\downarrow$

$S = \sum_{i,j=1}^{n} a_i a_j (x_i x_j - 1)$

let $\begin{cases} a_1 = 1 & a_2 = 1 & a_3 = 1 \\ x_1 = 1 & x_2 = 1 & x_3 = 0 \end{cases}$

$S = a_1^2 + a_2^2 + a_3^2 + 2\underline{a_1 a_3 k(x_1, x_3)} + 2\underline{a_2 a_3 k(x_2, x_3)} + 2\underline{a_1 a_2 k(x_1, x_2)}$

$= a_1^2 + a_2^2 + a_3^2 + \underline{2a_1 a_3 x_1 x_3 - 2a_1 a_3} + \underline{2a_2 a_3 x_2 x_3 - 2a_2 a_3} + \underline{2a_1 a_2 x_1 x_2 - 2a_1 a_2}$

$= 1 + 1 + 1 + 0 - 2 + 0 - 2 + 2 - 2$

$= -1 \rightarrow < 0 \Rightarrow$ 有反例、$k(x, x') = k_1(x, x') - 1$

is not a valid kernel

c. $k_1(x,x')^2 = k_1(x,x')k_1(x,x') \Rightarrow$ valid kernel 相乘

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

$exp(y) = 1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \ldots \Rightarrow$ polynomial with nonnegative coefficient

$\|x\|^2 \geq 0$、$\|x'\|^2 \geq 0 \Rightarrow exp(\|x\|^2) \geq 1$、$exp(\|x'\|^2) \geq 1$

$\Rightarrow exp(\|x\|^2)exp(\|x'\|^2) \geq 1$

$\Rightarrow$ let $c = exp(\|x\|^2)exp(\|x'\|^2) \Rightarrow c \geq 0$

$\rightarrow k(x,x') = \underbrace{k_1(x,x')^2}_{\text{valid}} + c$

$\rightarrow q(m) = (m+c) \rightarrow q =$ polynomial with nonnegative coefficient.

$\rightarrow k(x,x') = q(k_1(x,x'))$

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

d. $k_1(x,x')$ is a valid kernel $\Rightarrow \geq 0$.

$\Rightarrow exp(m) \geq 1$ when $m \geq 0 \Rightarrow exp(k_1(x,x')) \geq 1$

$\Rightarrow exp(k_1(x,x')) - 1 \geq 0$

let $c = exp(k_1(x,x')) - 1 \Rightarrow c \geq 0$

let $q(m) = m^2 + c \Rightarrow q =$ polynomial with nonnegative coefficient.

$\Rightarrow k(x,x') = q(k_1(x,x'))$ ∴ $k(x,x')$ is a valid kernel #

(10%) Consider the optimization problem

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

State the dual problem.

$(x-2)^2 = x^2-4x+4 \qquad -(x+3)(x-1) \geq 3.$

$$\text{Lagrangian} = L(x, a) = (x^2-4x+4) + a(x^2+2x-6)$$

$$\frac{dL}{dx} = 2x-4+2ax+2a = 0 \qquad (2a+2)x = 4-2a$$

$$x = \frac{2-a}{a+1}$$

$\rightarrow$ for $a \neq -1$, min at $x = \dfrac{2-a}{a+1}$

$\Rightarrow$ dual:

$$\text{maximize } L(a) = \frac{-(a-2)^2}{a+1} - 6a + 4$$

$$\text{subject to } a \geq 0.$$