# NCTU Introduction to Machine Learning, Homework 4

**Deadline: Nov. 29, 23:59**

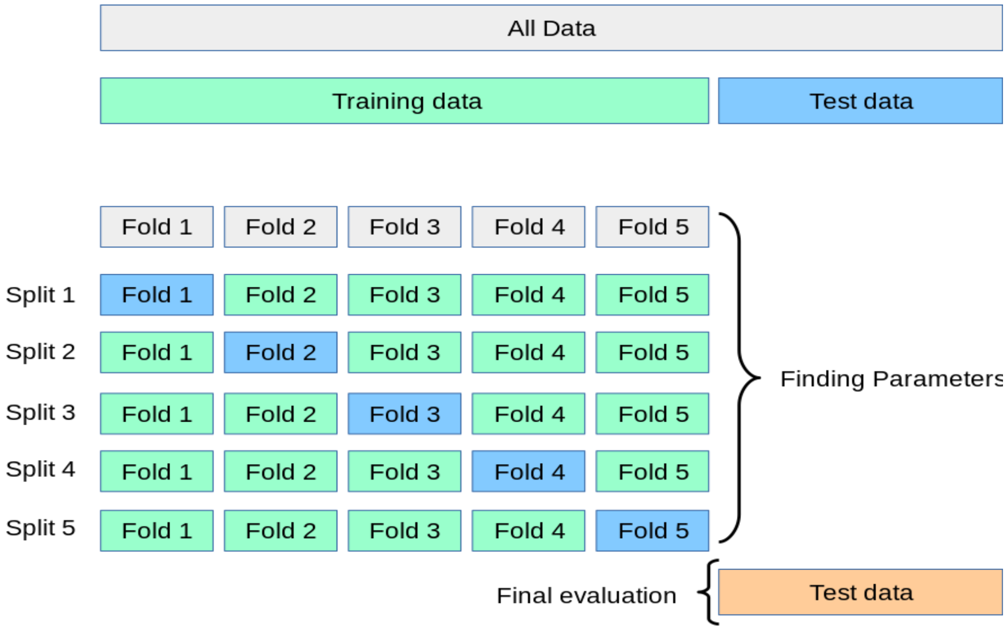
## Part. 1, Coding (50%):

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



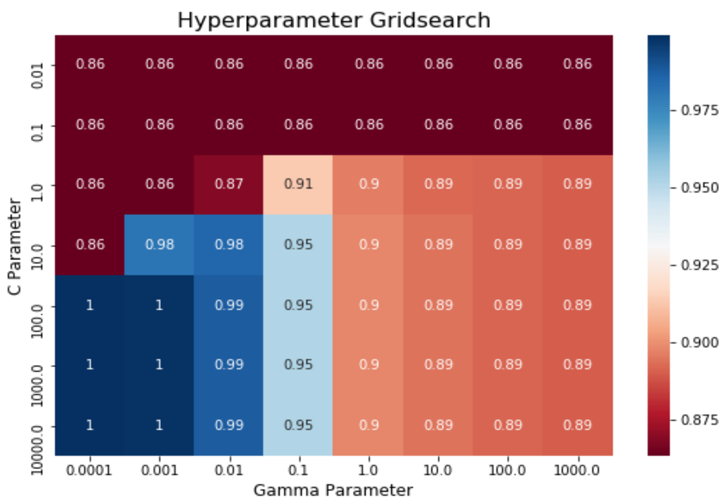

Example output of split 1:

```
Split:  1 , Training index:  [1335  581 1076 ... 1292 5464 2667] , Validation index:  [4718 3622 1068 2199 6720 1470 318
 5288  248 4545 5172 4289 3035 2595 6474  670 3991 1330 2775 1267 1293
 5768 2690 6272 6669 5630 1553 6489 1218 6950 4646 6536 2538 5563 5098
 6719 3471 3005 5974 5552 5461  276 3414 1567 6265 5978 4917 1676 1254
 4410 5984 4123 4006 3038    0 4196 3589 1990 4723 3557  626 1074 1184
 6176 5421 1660 6687  780  350 3074 1600 1019 1642 3522 4195 6982  412
 5904  758  124 3283 6549  971 4331 4089  728 3804 1445 2159 3924 2583
 4923 6563 5574 6154 5784 2711 2006 4068 4075  568 6081 2603 4756 3957
 3043 3155 6092 4845 2036 6061 3312 3266  558 5575 5690 4217 1223 2627
 1562 4587 3349 4193 6449  366 5254  178  782 6050  848 6887 2088 2075
  831 1697 4054 3773 6465 1442 2571 5140 5151 2961 3738 3526  436 2330
 5150 1175 6986 4645   43   74    1 2121 5301 2084  945 3374 4314 4370
 5188 1348 3139 1409 2513 4678 6035 1390  865 6786 3411  440   89 3534
 6844 6591 5027 5660 2676 6108 3028 6442 2793 3281  425 1572 4226  904
 3791 1483  960 1198 2457 4862 3978 3932 5682 1959 1266  795 3423  582
 1685 5505 3596 2897 1228 3122  847 2204 3549 2382   84  675 1941 6869
 5787 4153 4988 5798 5566 2678  355 2185 3172 2896  852  246 3058 1367
 5041 6230 6245  559 2964 1115 5933 4007   20 6855 3849 1849 1314 4496
 6703 6464 5705  345 2847 4475   30 2654 2955 2710 6989 2834 1796 4766
```

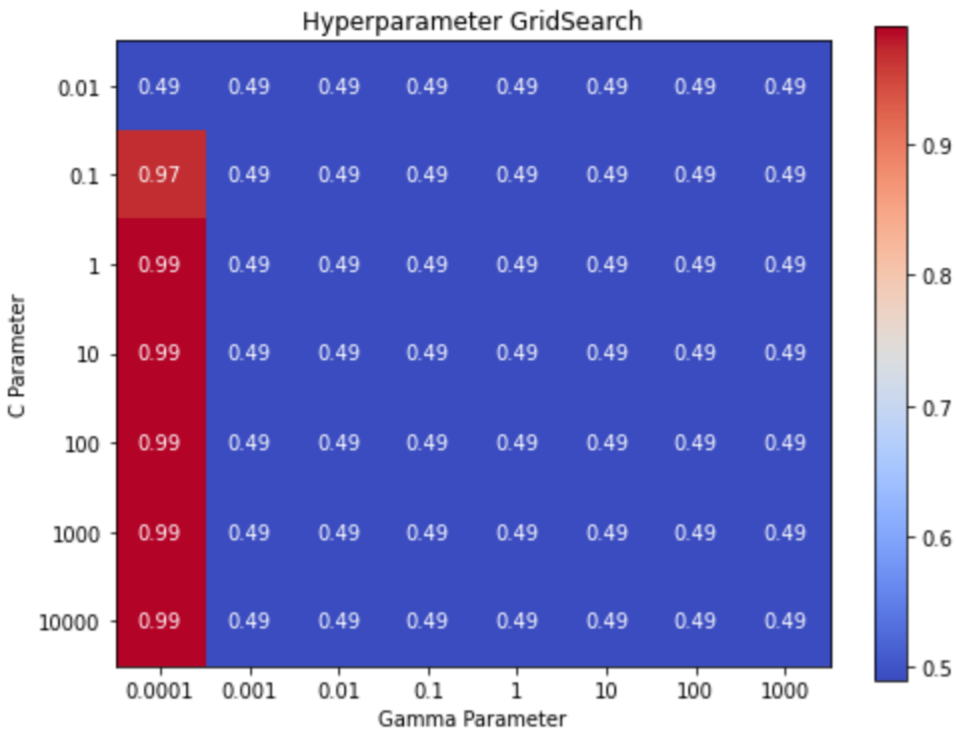
2. (20%) Grid Search & Cross-validation: using [sklearn.svm.SVC]() to train a classifier on the provided train set and conduct the grid search of “C” and “gamma,” “kernel’=’rbf’ to find the best hyperparameters by cross-validation. Print the best hyperparameters you found.
   Note: We suggest using K=5
   best_parameters:  {'kernel': 'rbf', 'gamma': 0.0001, 'C': 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.

   *Note: This image is for reference, not the answer*
   *Note: [matplotlib]() is allowed to use*




My heatmap result:




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

1. let Gram matrix K 的維度是 $N \times N$

Gram matrix is symmetric, positive semidefinite 的 matrix

if and only if $k_{nm} = \varphi(x_n)^T \varphi(x_m)$, $k_{nm} = k_{mn}$ 且 $z \in R^d$

positive semidefinite $\Leftrightarrow$ $k(x,x')$ is valid kernel.

"$\Leftarrow$" suppose k is valid.

$$
\begin{aligned}
z^T K z &= \sum_{n=1}^{N} \sum_{m=1}^{N} z_n K_{nm} z_m \\
&= \sum_{n=1}^{N} \sum_{m=1}^{N} z_n \varphi(x_n)^T \varphi(x_m) z_m \\
&= \sum_{n=1}^{N} \sum_{m=1}^{N} z_n \left( \sum_{k=1}^{N} \varphi_k(x_n)^T \varphi_k(x_m) \right) z_m \\
&= \sum_{n=1}^{N} \sum_{m=1}^{N} \sum_{k=1}^{N} z_n \varphi_k(x_n) \varphi_k(x_m) z_m \\
&= \sum_{k=1}^{N} \left( \sum_{n=1}^{N} z_n \varphi_k(x_n) \right)^2 \geq 0
\end{aligned}
$$

"$\Rightarrow$" suppose positive semidefinite is given

let $K = UDU^T$ = eigenvectors decomposition of K

$$
\begin{aligned}
K_{ij} &= (\sqrt{D} u_i)^T (\sqrt{D} u_j), \quad \varphi(x_i) = \sqrt{D} u_i \\
&= \varphi(x_i)^T \cdot \varphi(x_j)
\end{aligned}
$$

Thus, if K is positive semidefinite, then $k(x,x')$ is 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.

2. consider using Taylor expansion of $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

$k(x, x') = \exp(k_1(x, x'))$ can be written as

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

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}') \quad (6.17)$$
$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}') \quad (6.18)$$
$$k(\mathbf{x}, \mathbf{x}') = ck_1(\mathbf{x}, \mathbf{x}') \quad (6.13)$$

since all coefficient of Taylor expansion are positive and by 6.17, 6.18, 6.13

we can prove that $k(x, x') = \exp(k_1(x, x'))$ 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. $k(x, x') = k_1(x, x')^2 + exp\big(k_1(x, x')\big) - 1$

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

if $k_c(x,x') = 1$ is a valid kernel, then $k(x,x')$ is a valid kernel

to prove that $k_c(x,x') = 1$ is a valid kernel.

positive semidefinite $\Rightarrow F^T K_c F \geq 0, \forall F \in R^d$

$\Rightarrow \sum_{i,j} f_i k_c(i,j) f_j = k_c(i,j) \sum_{i,j} f_i f_j$

$\Rightarrow c\|F\| \geq 0 \Rightarrow$ valid kernel

according to $k(x,x') = k_1(x,x') + k_2(x,x')$ - (6.17)

$k(x,x')$ is a valid kernel since $k_1(x,x')$ and 1 are both valid kernel

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

if $0 \leq k_1(x,x') < 1$, then $k(x,x') < 0$

thus $k(x,x')$ is not a valid kernel.

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

① $k_1(x,x')^2 = k_1(x,x') \cdot k_1(x,x')$ is a valid kernel by $k(x,x') = k_1(x,x') \cdot k_2(x,x')$ - 6.18

② let $\varphi(x) = exp(\|x\|^2)$, then $exp(\|x'\|^2) = \varphi(x')$

$exp(\|x\|^2) \cdot exp(\|x'\|^2) = \varphi(x) \cdot \varphi(x') = k_c(x,x')$ is a valid kernel by definition

③ $k(x,x') = k_1(x,x')^2 + exp(\|x\|^2) \cdot exp(\|x'\|^2)$ is a valid kernel by $k(x,x') = k_1(x,x') + k_2(x,x')$ - 6.17

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

① $k_1(x,x')^2 = k_1(x,x') \cdot k_1(x,x')$ is a valid kernel by $k(x,x') = k_1(x,x') \cdot k_2(x,x')$ - 6.18

② $exp(k_1(x,x'))$ is a valid kernel by 6.16.

③ $k_2(x,x') = k_1(x,x')^2 + exp(k_1(x,x'))$ is a valid kernel by $k(x,x') = k_1(x,x') + k_2(x,x')$ - 6.17

④ $k(x,x') = k_2(x,x') - 1$ is not a valid kernel according to 3.b.

(10%) Consider the optimization problem

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

State the dual problem.

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

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

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

$[(1+\lambda)x^2+(-4+2\lambda)x+(4-6\lambda)]' = 2(1+\lambda)x+(-4+2\lambda)=0$

$\Rightarrow (1+\lambda)x = 2-\lambda$

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

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

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

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

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