

# NCTU Pattern Recognition, Homework 4

**Deadline: May 25, 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](https://github.com/NCTU-VRDL/CS_AT0828/tree/main/HW4)

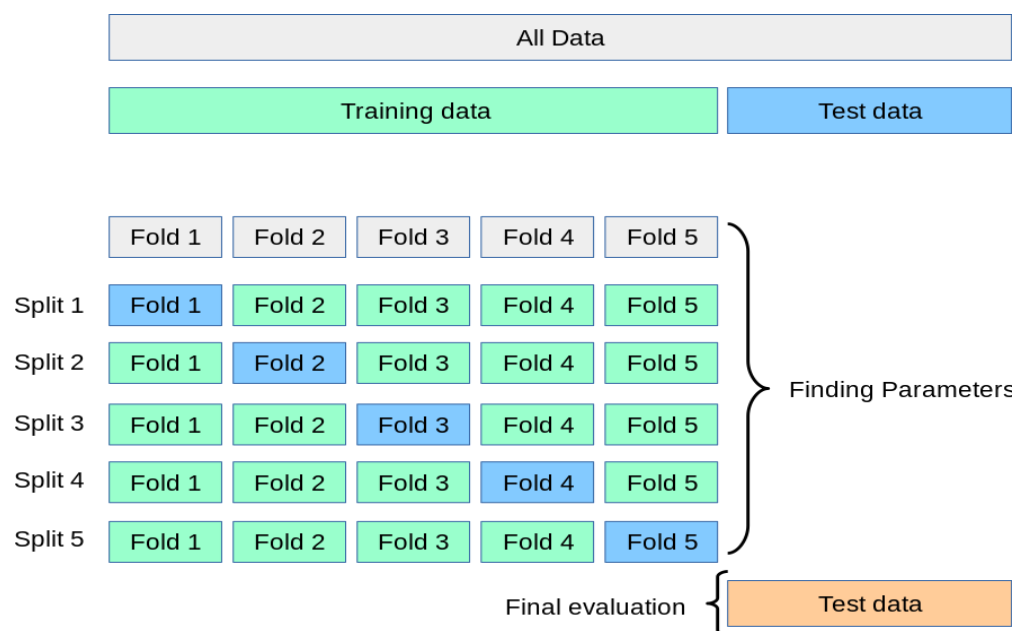
Please note that only NumPy can be used to implement cross-validation and grid search. You will get no points by simply calling [sklearn.model\\_selection.GridSearchCV](#).

- (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



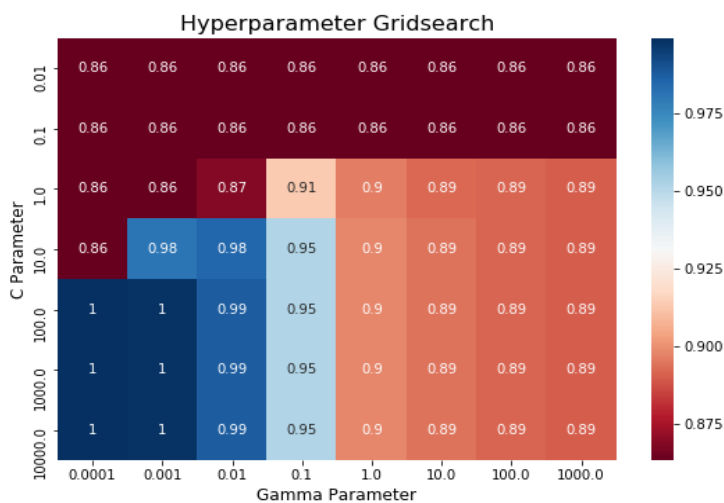
2. (20%) Grid Search & Cross-validation: using [sklearn.svm.SVC](#) to train a classifier on the provided train set and conduct the grid search of “C” and “gamma,” “kernel”=’rbf’ to find the best hyperparameters by cross-validation. Print the best hyperparameters you found.

Note: We suggest using K=5

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



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%):**

1. (10%) Given a valid kernel  $k_1(x, x')$ , prove that the following proposed functions are or are not valid kernels.
  - a.  $k(x, x') = (k_1(x, x'))^2 + (k_1(x, x') + 1)^2$
  - b.  $k(x, x') = (k_1(x, x'))^2 + \exp(\|x\|^2) * \exp(\|x'\|^2)$
2. (10%) Show that the kernel matrix  $\mathbf{K} = [k(\mathbf{x}_n, \mathbf{x}_m)]_{nm}$  should be positive semidefinite is the necessary and sufficient condition for  $k(\mathbf{x}, \mathbf{x}')$  to be a valid kernel.
3. (10%) Consider the dual formulation of the least-squares linear regression problem given on page 6 in the ppt of Kernel Methods. Show that the solution for the components  $\mathbf{a}_n$  of the vector  $\mathbf{a}$  can be expressed as a linear combination of the elements of the vector  $\boldsymbol{\phi}(\mathbf{x}_n)$ . Denoting these coefficients by the vector  $\mathbf{w}$ , show that the dual of the dual formulation is given by the original representation in terms of the parameter vector  $\mathbf{w}$ .
4. (10%) Prove that the Gaussian kernel defined by (eq 1) is valid and show the function  $\boldsymbol{\phi}(\mathbf{x})$ , where  $\mathbf{x} \in \mathbb{R}^1$ .  
(eq1)  $k(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2) = \boldsymbol{\phi}(x)^T \boldsymbol{\phi}(x')$
5. (10%) Consider the optimization problem
$$\begin{aligned} & \text{minimize } (x - 2)^2 \\ & \text{subject to } (x+3)(x-1) \leq 2 \end{aligned}$$
State the dual problem.