NCTU Introduction to Machine Learning, Homework 4

Deadline: Nov. 29, 23:59

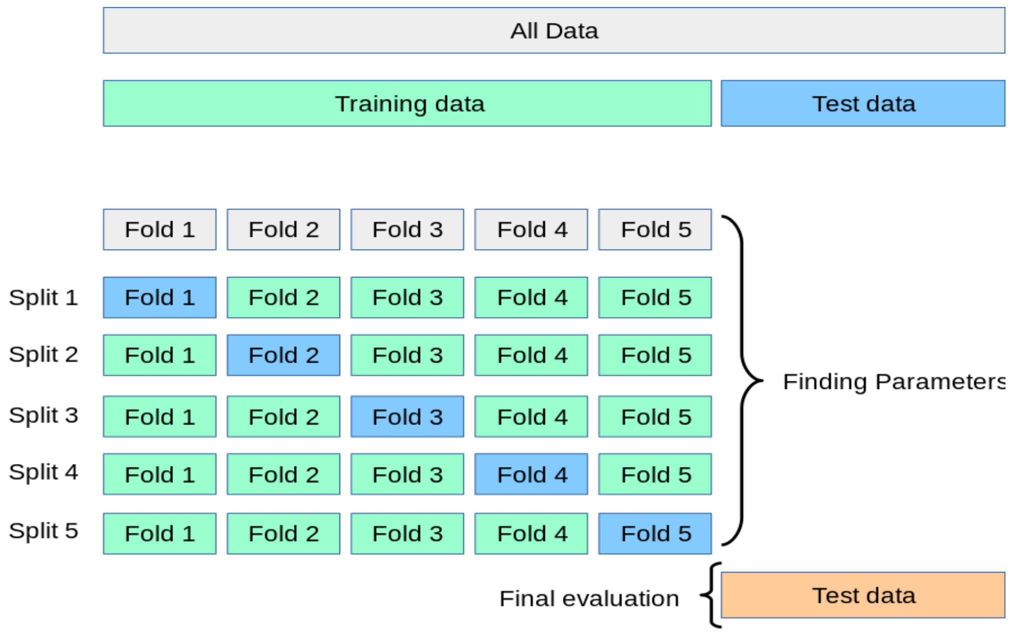
Part. 1, Coding (50%):

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

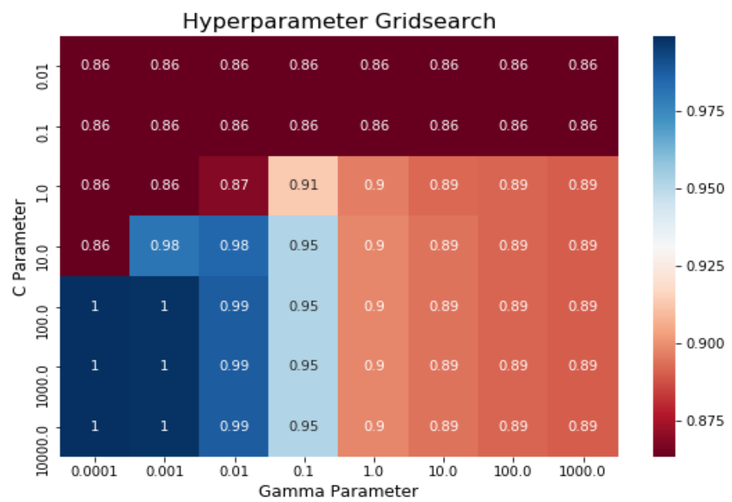
```
print(best_parameters)
```

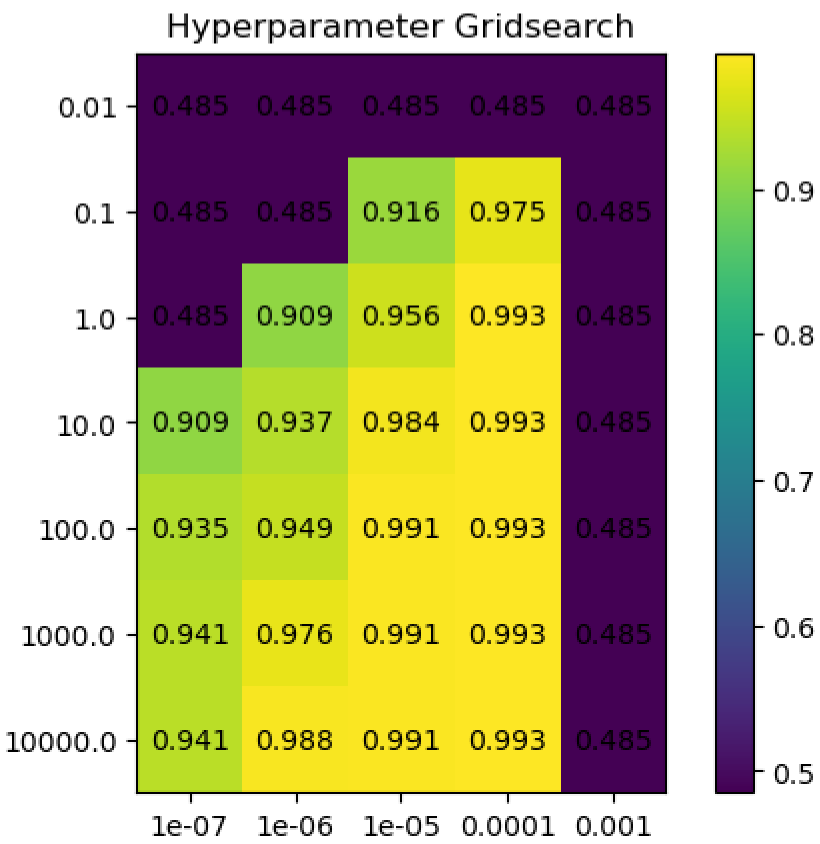
```
(1.0, 0.0001)
```

   C=1.0, gamma = 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.

   Note: This image is for reference, not the answer
   Note: matplotlib is allowed to use



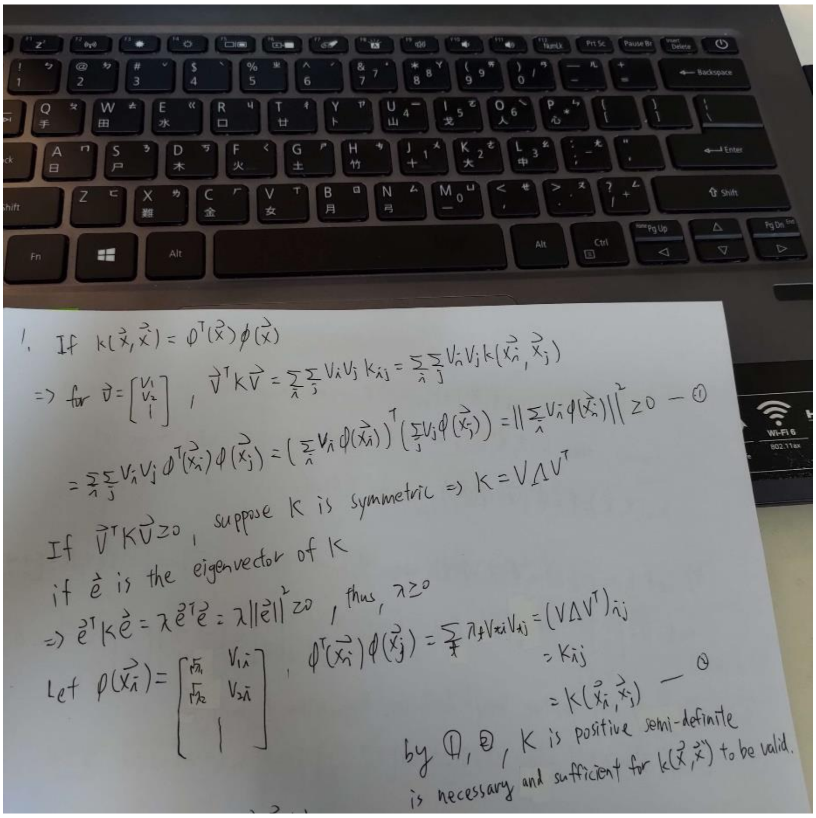


4. (10%) Train your SVM model by the best hyperparameters you found from question 2 on the whole training data and evaluate the performance on the test set.

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

Part. 2, Questions (50%):

(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. If $k(\vec{x},\vec{x}) = \phi^T(\vec{x})\phi(\vec{x})$

$\Rightarrow$ for $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \end{bmatrix}$, $\vec{v}^T K \vec{v} = \sum_i\sum_j v_i v_j K_{ij} = \sum_i\sum_j v_i v_j k(\vec{x}_i,\vec{x}_j)$

$= \sum_i\sum_j v_i v_j \phi^T(\vec{x}_i)\phi(\vec{x}_j) = \left(\sum_i v_i\phi(\vec{x}_i)\right)^T\left(\sum_j v_j\phi(\vec{x}_j)\right) = \left\|\sum_i v_i\phi(\vec{x}_i)\right\|^2 \ge 0$ — ①

If $\vec{v}^TK\vec{v}\ge 0$, suppose K is symmetric $\Rightarrow K = V\Lambda V^T$

if $\vec{e}$ is the eigenvector of K

$\Rightarrow \vec{e}^TK\vec{e} = \lambda\vec{e}^T\vec{e} = \lambda\|\vec{e}\|^2 \ge 0$, thus, $\lambda \ge 0$

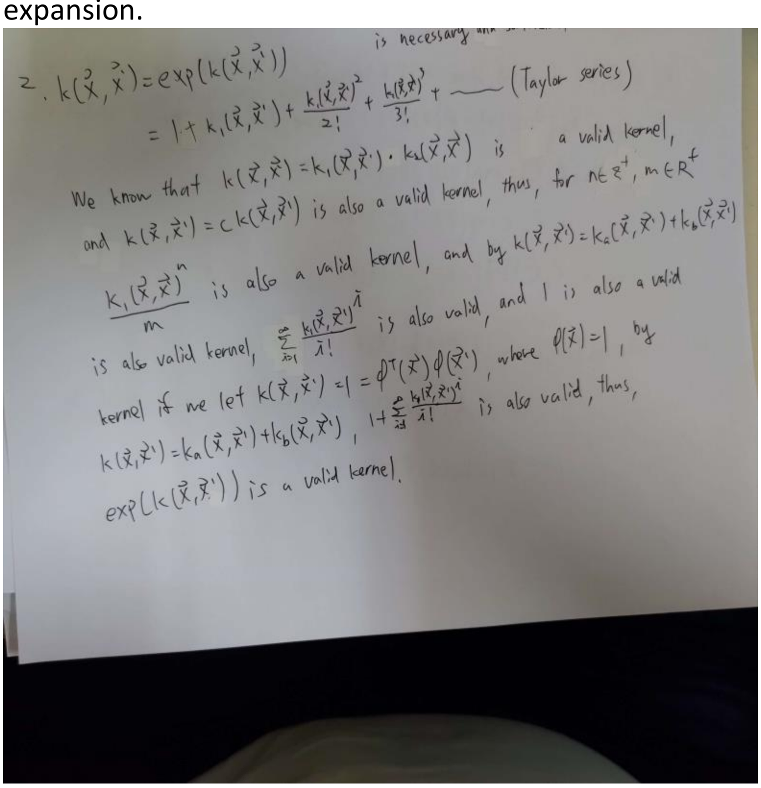
Let $\phi(\vec{x}_i) = \begin{bmatrix} \sqrt{\lambda_1} & V_{1i} \\ \sqrt{\lambda_2} & V_{2i} \\ & \vdots \end{bmatrix}$, $\phi^T(\vec{x}_i)\phi(\vec{x}_j) = \sum_t \lambda_t V_{ti}V_{tj} = (V\Lambda V^T)_{ij}$

$= K_{ij}$

$= k(\vec{x}_i,\vec{x}_j)$ — ②

by ①, ②, K is positive semi-definite is necessary and sufficient for $k(\vec{x},\vec{x}')$ to be valid.

(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. $k(\vec{x},\vec{x}') = \exp(k(\vec{x},\vec{x}'))$

$= 1 + k_1(\vec{x},\vec{x}') + \frac{k_1(\vec{x},\vec{x}')^2}{2!} + \frac{k_1(\vec{x},\vec{x}')^3}{3!} + \cdots$ (Taylor series)

We know that $k(\vec{x},\vec{x}') = k_1(\vec{x},\vec{x}')\cdot k_2(\vec{x},\vec{x}')$ is a valid kernel, and $k(\vec{x},\vec{x}') = ck(\vec{x},\vec{x}')$ is also a valid kernel, thus, for $n \in \mathbb{Z}^+$, $m \in R^+$, $\frac{k_1(\vec{x},\vec{x}')^n}{m}$ is also a valid kernel, and by $k(\vec{x},\vec{x}') = k_a(\vec{x},\vec{x}') + k_b(\vec{x},\vec{x}')$ is also valid kernel, $\sum_{i=1}^{\infty}\frac{k_1(\vec{x},\vec{x}')^i}{i!}$ is also valid, and 1 is also a valid kernel if we let $k(\vec{x},\vec{x}') = 1 = \phi^T(\vec{x})\phi(\vec{x}')$, where $\phi(\vec{x}) = 1$, by $k(\vec{x},\vec{x}') = k_a(\vec{x},\vec{x}') + k_b(\vec{x},\vec{x}')$, $1 + \sum_{i=1}^{\infty}\frac{k_1(\vec{x},\vec{x}')^i}{i!}$ is also valid, thus, $\exp(k(\vec{x},\vec{x}'))$ 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) By $k(\vec{x}, \vec{x}') = k_a(\vec{x}, \vec{x}') + k_b(\vec{x}, \vec{x}')$, we know that if 1 is a valid kernel, then $k(\vec{x}, \vec{x}')$ is a valid kernel, and 1 is valid if we let $k_2(\vec{x}, \vec{x}') = 1 = \phi^T(\vec{x})\phi(\vec{x}')$, where $\phi(\vec{x}) = 1$

b) Let $k_1(\vec{x}, \vec{x}') = 0 = \phi^T(\vec{x})\phi(\vec{x})$ where $\phi(\vec{x}) = 0$, then $k(\vec{x}, \vec{x}') = -1$, Let $K = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, then $\vec{v}^T K \vec{v} = [1\ 0]\begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = [-1\ -1]\begin{bmatrix} 1 \\ 0 \end{bmatrix} = -1 < 0$

so K is not positive semi-definite. Eigenvalue λ will be

$K\vec{v} = \lambda I \vec{v} = (K - \lambda I)\vec{v}$, $\Rightarrow \begin{bmatrix} -1-\lambda & -1 \\ -1 & -1-\lambda \end{bmatrix}\vec{v} = 0$, so $(-1-\lambda)^2 - 1 = 0$, $\lambda = 0$

c) $k_1(\vec{x}, \vec{x}')^2$ is valid since $k(\vec{x}, \vec{x}') = k_a(\vec{x}, \vec{x}')k_b(\vec{x}, \vec{x}')$ is also valid,

Assume $k_2(\vec{x}, \vec{x}') = 1 = \phi^T(\vec{x})\phi(\vec{x}')$, where $\phi(\vec{x}) = 1$, k_2 is valid

Let $k_3(\vec{x}, \vec{x}') = exp(\|\vec{x}\|^2)\, k_2(\vec{x}, \vec{x}')\, exp(\|\vec{x}'\|^2)$

$= exp(\|\vec{x}\|^2) \cdot 1 \cdot exp(\|\vec{x}'\|^2)$

by $k(\vec{x}, \vec{x}') = f(x)k_a(\vec{x}, \vec{x}')f(x')$, k_3 is valid

So, $k(\vec{x}, \vec{x}') = k_1(\vec{x}, \vec{x}')^2 + k_3(\vec{x}, \vec{x}')$

$= k_1(\vec{x}, \vec{x}')^2 + exp(\|\vec{x}\|^2)\, exp(\|\vec{x}'\|)^2$ is also valid by

$k(\vec{x}, \vec{x}') = k_a(\vec{x}, \vec{x}') + k_b(\vec{x}, \vec{x}')$

d) $k_1(\vec{x},\vec{x}')^2$ is valid since $k_1(\vec{x},\vec{x}')^2 = k_1(\vec{x},\vec{x}')\cdot k_1(\vec{x},\vec{x}')$ and k_1 is valid.

$\exp(k(\vec{x},\vec{x}')) = 1 + k_1(\vec{x},\vec{x}') + \frac{k_1(\vec{x},\vec{x}')^2}{2!} + \frac{k_1(\vec{x},\vec{x}')^3}{3!} + \cdots$

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

by $k(\vec{x},\vec{x}') = ck_a(\vec{x},\vec{x}')$, $k(\vec{x},\vec{x}') = k_a(\vec{x},\vec{x}')\cdot k_b(\vec{x},\vec{x}')$, we know

for $n \in \mathbb{Z}^+, m \in R^+$, $mk_1(\vec{x},\vec{x}')^n$ is valid, and by

$k(\vec{x},\vec{x}') = k_a(\vec{x},\vec{x}') + k_b(\vec{x},\vec{x}')$, $k_1(\vec{x},\vec{x}') + \frac{k_1(\vec{x},\vec{x}')^2}{2!} + \cdots$ is valid

which means $\exp(k_1(\vec{x},\vec{x}')) - 1$ is valid, so, by $k(\vec{x},\vec{x}') = k_a(\vec{x},\vec{x}') + k_b(\vec{x},\vec{x}')$

$k_1(\vec{x},\vec{x}')^2 + \exp(k_1(\vec{x},\vec{x}')) - 1$ is valid.

(10%) Consider the optimization problem

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

State the dual problem.

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

using Lagrange multiplier, we have

$L(x,\alpha) = x^2 - 4x - 4 + \alpha(x^2+2x-6)$

$= (1+\alpha)x^2 + (2\alpha-4)x + (-6\alpha+4)$, $\alpha \geq 0$

$\frac{\partial L}{\partial x} = 2(1+\alpha)x + (2\alpha-4) = 0$, $x = \frac{-(2\alpha-4)}{2(1+\alpha)} = \frac{2-\alpha}{1+\alpha}$

thus, $L(x,\alpha)$ can be written as $L(\alpha) = \frac{(2-\alpha)^2}{1+\alpha} + \frac{-2(\alpha-2)^2}{1+\alpha} + (4-6\alpha)$

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

So, we can state the dual problem: minimize $(x-2)^2$ subject to $(x+3)(x-1) \leq 3$

by maximize $\frac{-(\alpha-2)^2}{1+\alpha} + (4-6\alpha)$ subject to $\alpha \geq 0$