HW4

Part. 1, Coding (50%):

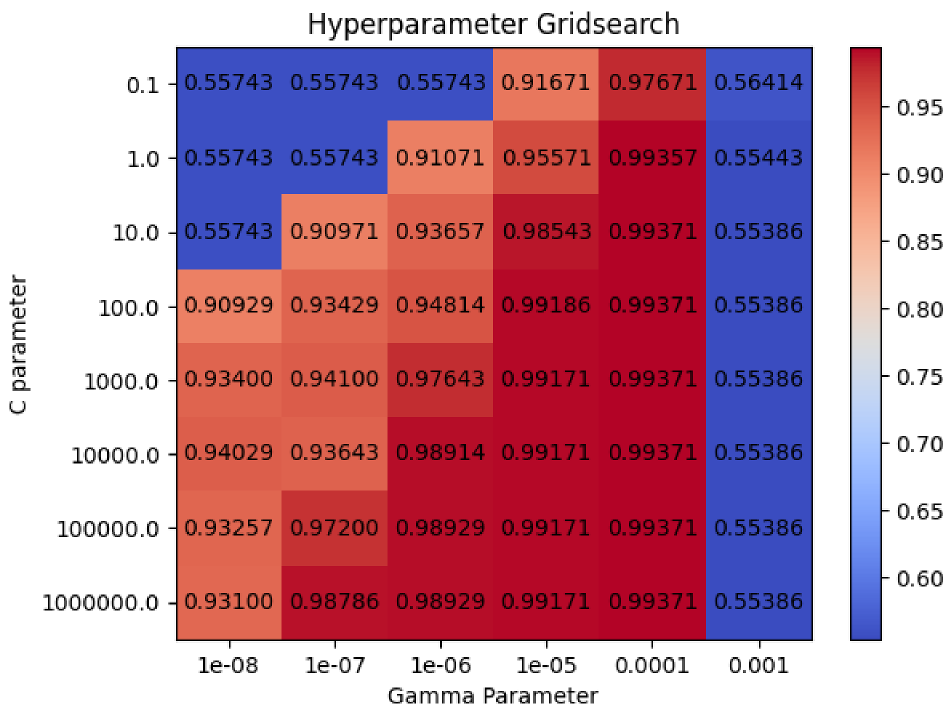
1. (10%) K-fold data partition: Implement the K-fold cross-validation function. Your function should take K as an argument and return a list of lists (*len(list) should equal to K*), which contains K elements. Each element is a list containing two parts, the first part contains the index of all training 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)

```
kfold_data = cross_validation(x_train, y_train, k=10)
assert len(kfold_data) == 10 # should contain 10 fold of data
assert len(kfold_data[0]) == 2 # each element should contain train fold and validation fold
assert kfold_data[0][1].shape[0] == 700 # The number of data in each validation fold should equal to training data divieded by K
✓ 0.4s
```

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

```
[C, gamma]
[10.0, 0.0001]
SVC(C=10.0, 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.



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%):

$$\begin{aligned}
k(\mathbf{x}, \mathbf{x}') &= ck_1(\mathbf{x}, \mathbf{x}') && (6.13)\\
k(\mathbf{x}, \mathbf{x}') &= f(\mathbf{x})k_1(\mathbf{x}, \mathbf{x}')f(\mathbf{x}') && (6.14)\\
k(\mathbf{x}, \mathbf{x}') &= q\left(k_1(\mathbf{x}, \mathbf{x}')\right) && (6.15)\\
k(\mathbf{x}, \mathbf{x}') &= \exp\left(k_1(\mathbf{x}, \mathbf{x}')\right) && (6.16)\\
k(\mathbf{x}, \mathbf{x}') &= k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}') && (6.17)\\
k(\mathbf{x}, \mathbf{x}') &= k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}') && (6.18)\\
k(\mathbf{x}, \mathbf{x}') &= k_3\left(\boldsymbol{\phi}(\mathbf{x}), \boldsymbol{\phi}(\mathbf{x}')\right) && (6.19)\\
k(\mathbf{x}, \mathbf{x}') &= \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}' && (6.20)\\
k(\mathbf{x}, \mathbf{x}') &= k_a(\mathbf{x}_a, \mathbf{x}_a') + k_b(\mathbf{x}_b, \mathbf{x}_b') && (6.21)\\
k(\mathbf{x}, \mathbf{x}') &= k_a(\mathbf{x}_a, \mathbf{x}_a')k_b(\mathbf{x}_b, \mathbf{x}_b') && (6.22)
\end{aligned}$$

1. (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.

$\Rightarrow$

K is symmetric. Thus, we have $K = V\Lambda V^T$, where V is an orthonormal matrix v_t and the diagonal matrix Λ contains the eigenvalues λ_t of K.

If K is positive semidefinite, all eigenvalues are non-negative.

Consider the feature map: $\phi : x_i \rightarrow (\sqrt{\lambda_t}v_{ti})_{t=1}^n \in \mathbb{R}^n$.

We find that $\phi(x_i)^T\phi(x_j) = \sum_{t=1}^n \lambda_t v_{ti} v_{tj} = (V\Lambda V^T)_{ij} = K_{ij} = k(x_i, x_j)$

$\Leftarrow$

If $k(x, x') = \phi(x)^T\phi(x')$ is a valid kernel.

Suppose $A = [\phi(x_1), \phi(x_2), \ldots, \phi(x_n)]^T$.

Let the kernel matrix be $K = [k(x_n, x_m)]_{nm} = AA^T$.

For any non-zero vector y, $y^T K y = y^T AA^T y = (yA^T)^T(yA^T) \geq 0$

Therefore K is positive semidefinite.

2. (10%) Given a valid kernel $k_1(x, x')$, explain that $k(x_i, x_j) = K_{ij} = (V\Lambda V^T)_{ij} = \sum_{t=1}^n \lambda_t v_{ti} v_{tj} = \phi(x_i)^T\phi(x_j)$ is also a valid kernel. Your answer may mention some terms like series or expansion.

$\exp{(x)} = \lim_{m\to\infty}(1 + \frac{x}{m})^m = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$

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

We know that is a polynomial with nonnegative coefficients as above.

According to (6.15), $\exp{(k_1(x, x'))}$ is a valid kernel.

Proven $k(x, x') = \exp{(k_1(x, x'))}$ is also a valid kernel if $k_1(x, x')$ is a valid kernel.

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

Suppose a function q be $q(x) = x + 1$ which is a polynomial with nonnegative coefficients.

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

According to (6.15), $q(k_1(x, x'))$ is a valid kernel.

Proven $k(x, x') = k_1(x, x') + 1$ is also a valid kernel if $k_1(x, x')$ is a valid kernel.

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

Suppose $K_1 = \begin{bmatrix} k_1(x_1, x_1) & k_1(x_2, x_1) \\ k_1(x_1, x_2) & k_1(x_2, x_2) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, eigenvalues $\lambda_1 = 0, \lambda_2 = 1$

Then $K = \begin{bmatrix} k(x_1, x_1) & k(x_2, x_1) \\ k(x_1, x_2) & k(x_2, x_2) \end{bmatrix} = \begin{bmatrix} k_1(x_1, x_1) - 1 & k_1(x_2, x_1) - 1 \\ k_1(x_1, x_2) - 1 & k_1(x_2, x_2) - 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & -1 \end{bmatrix}$

And eigenvalues are $\lambda_1 = \frac{-\sqrt{5}-1}{2}$, $\lambda_2 = \frac{\sqrt{5}-1}{2}$

Since $\lambda_1 < 0$, K is not positive semidefinite.

Proven $k(x, x') = k_1(x, x') - 1$ is not a valid kernel.

c. $k(x, x') = k_1(x, x')^2 + \exp\left(||x||^2\right) * \exp\left(||x'||^2\right)$

Suppose $\phi_2(x) = exp(||x||^2)$

$k_2(x, x') = \exp\left(||x||^2\right) * \exp\left(||x'||^2\right) = \phi_2(x)^T \phi_2(x')$

Therefore $k_2(x, x')$ is a valid kernel.

$k(x, x') = k_1(x, x')^2 + k_2(x, x')$

According to (6.17), (6.18), $k(x, x')$ is a valid kernel.

Proven $k(x, x') = k_1(x, x')^2 + \exp\left(||x||^2\right) * \exp\left(||x'||^2\right)$ is a valid kernel.

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

$$\begin{aligned} k_2(x, x') &= \exp\left(k_1(x, x')\right) - 1 = -1 + 1 + k_1(x, x') + \frac{k_1(x, x')^2}{2!} + \frac{k_1(x, x')^3}{3!} + \dots \\ &= k_1(x, x') + \frac{k_1(x, x')^2}{2!} + \frac{k_1(x, x')^3}{3!} + \dots \end{aligned}$$

According to (6.13), (6.17), (6.18), $k_2(x, x')$ is a valid kernel.

$k(x, x') = k_1(x, x')^2 + k_2(x, x')$

According to (6.17), (6.18), $k(x, x')$ is a valid kernel.

Proven $k(x, x') = k_1(x, x')^2 + \exp\left(k_1(x, x')\right) - 1$ is a valid kernel.

4. Consider the optimization problem *minimize* $(x - 2)^2$, *subject to* $(x + 3)(x - 1) \leq 3$, State the dual problem.

$L(x, a) = (x - 2)^2 + a((x + 3)(x - 1) - 3) = (1 + a)x^2 + (2a - 4)x + 4 - 6a$

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

$L'(a) = \frac{(2-a)^2}{1+a} + \frac{(2-a)(2a-4)}{1+a} + (4 - 6a) = \frac{-7a^2+2a}{1+a}$

Maximize $L'(a)$ subject to $a \geq 0$