NCTU Introduction to Machine Learning, Homework 4

109550018 郭昀

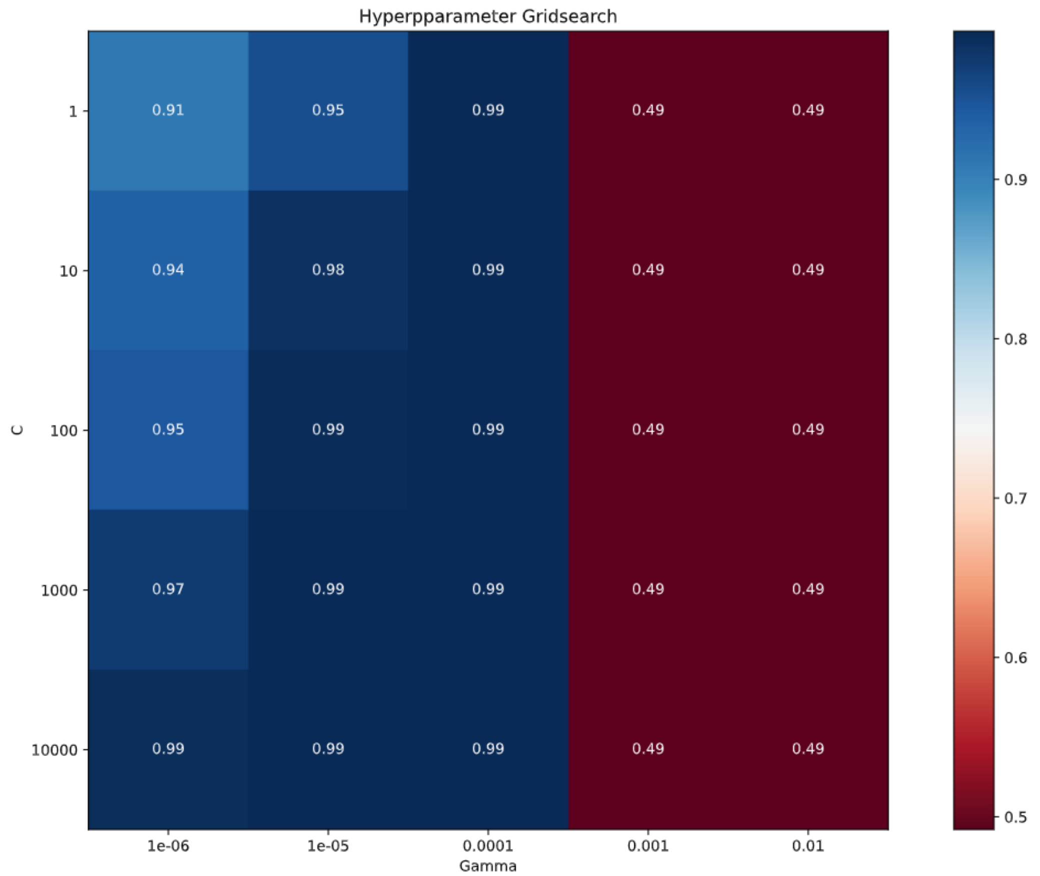
Part. 1, Coding (50%):

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)

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.

```
(1, 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.

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.

From valid kernel to positive semi-definite:

$\because$ For any nonzero vector Z and Gram matrix of a valid kernel, we can compute that

$$\Rightarrow \boldsymbol{z}^T \boldsymbol{K} \boldsymbol{z} = \sum_{i=1}^{d} \sum_{j=1}^{d} z_i \, \boldsymbol{K}_{ij} \, z_j = \sum_{i=1}^{d} \sum_{j=1}^{d} z_i \, \boldsymbol{\phi}(\boldsymbol{x}_i)^T \boldsymbol{\phi}(\boldsymbol{x}_j) \, z_j$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} z_i \left(\sum_{k=1}^{d} \phi_k(\boldsymbol{x}_i) \, \phi_k(\boldsymbol{x}_j) \right) z_j = \sum_{k=1}^{n} \sum_{i=1}^{d} \sum_{j=1}^{d} z_i \, \phi_k(\boldsymbol{x}_i) \, \phi_k(\boldsymbol{x}_j) \, z_j$$

$$= \sum_{k=1}^{d} \left(\sum_{i=1}^{d} z_i \, \phi_k(\boldsymbol{x}_i) \right)^2 \geq 0$$

$\therefore$ The gram matrix of a valid kernel $\boldsymbol{K}$ is positive semi-definite

From positive semi-definite to valid kernel:

$\because$ $\boldsymbol{K}$ is symmetric.

$\therefore$ We have $\boldsymbol{K} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$, where $\boldsymbol{V}$ is an orthonormal matrix v_t , and the diagonal matrix $\boldsymbol{\Lambda}$ contains the eigenvalues λ_t of $\boldsymbol{K}$

$\Rightarrow$ if $\boldsymbol{K}$ is positive semidefinite, all eigenvalues are non-negative

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

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

$\Rightarrow$ So that the kernel matrix should be positive semidefinite is the necessary and sufficient condition for $k(x, x')$ to be a 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.

It can be derived from the Maclaurin series and the formula from slide Ch6 p.15:

$k(x, x') = q(k_1(x, x'))$, where $q(\bullet)$ is a polynomial with nonnegative coefficients. (6.15)

$$\because exp(k_1(x, x')) = \sum_{n=0}^{\infty} \frac{k_1(x, x')^n}{n!} = 1 + k_1(x, x') + \frac{1}{2} * k_1(x, x')^2 + \frac{1}{6} * k_1(x, x')^3 + \ldots$$

$\therefore$ It can be inferred that $k(x, x') = exp(k_1(x, x'))$ is also a valid kernel function

(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 $q(x) = x + 1$, and by formula (6.15)

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

$\therefore k(x . x')$ is a valid kernel

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

Suppose the Gram matrix of $k_1(x, x')$ is $K = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$

We can compute that $(0.8 - \lambda)^2 = 0.2^2 \rightarrow \lambda = 1$ or 6 , all the eigenvalues of K are positive

$\Rightarrow$ valid, matrix K satisfy the property of positive semi-definite

$\because k(x, x') = k_1(x, x') - 1$

$\therefore$ Gram matrix of $k(x, x') = \begin{bmatrix} 0.8 - 1 & 0.2 - 1 \\ 0.2 - 1 & 0.8 - 1 \end{bmatrix} = \begin{bmatrix} -0.2 & -0.8 \\ -0.8 & -0.2 \end{bmatrix} = K'$

$\Rightarrow$ compute eigenvalues of K' , $(-0.2 - \lambda)^2 = 0.8^2 \rightarrow \lambda = 0.6$ or -1 not all nonnegative

$\Rightarrow K'$ does not satisfy the property of positive semi-definite

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

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

We can define $f(x) = exp(\|x\|^2)$, where $f(x) \geq 0$ for all x

$\Rightarrow$ rewrite $k(x, x') = k_1(x, x')^2 + exp(\|x\|^2)exp(\|x'\|^2) = k_1(x, x')^2 + f(x)f(x')$

$\because$ For $k(x, x') = k_1(x, x')^2 + f(x)f(x')$, we can define a polynomial with nonnegative coefficients

$G(x) = x^2 + C$, where $C = f(x)f(x') \geq 0$ for all x and x'

$\therefore$ $k(x, x') = G(k_1(x, x')) = k_1(x, x')^2 + f(x)f(x')$ is also a valid kernel by formula (6.15)

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

$\because$ $exp(k_1(x, x'))$ can be replaced by a Maclaurin series

$$\Rightarrow k(x, x') = k_1(x, x')^2 + exp(k_1(x, x')) - 1$$

$$= k_1(x, x')^2 + 1 + k_1(x, x') + \frac{1}{2}k_1(x, x')^2 + \frac{1}{6}k_1(x, x')^3 + \ldots - 1$$

$$= k_1(x, x') + \frac{3}{2}k_1(x, x')^2 + \frac{1}{6}k_1(x, x')^3 + \ldots$$

$\Rightarrow$ $k(x, x') = f(k_1(x, x'))$, where all the coefficients of $f(k_1(x, x'))$ are nonnegative

$\therefore$ $k(x, x')$ is valid kernel by formula (6.15)

(10%) Consider the optimization problem

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

State the dual problem.

Rewrite the constrain $(x + 3)(x - 1) \leq 3 \rightarrow -(x + 3)(x - 1) \geq -3$

Lagrangian Function: $L(x, \lambda) = (x - 2)^2 - \lambda\left[-(x + 3)(x - 1) + 3\right]$

$$= (x - 2)^2 - \lambda\left[-x^2 - 2x + 6\right]$$

$$\Rightarrow \frac{dL}{dx} = 2(x - 2) - \lambda(-2x - 2) = 2x - 4 + 2\lambda x + 2\lambda = 0 \Rightarrow x = \frac{2 - \lambda}{1 + \lambda}$$

$$\Rightarrow \frac{dL}{d\alpha} = x^2 + 2x - 6 = 0$$

$\Rightarrow$ the Lagrangian function reaches its minimum at $x = \dfrac{2 - \lambda}{1 + \lambda}$, where $\lambda \neq -1$

$\therefore$ The dual problem:

maximize $L(\lambda) = \dfrac{-7\lambda^2 + 2\lambda}{1 + \lambda}$, subject to $\lambda \geq 0$