NCTU Introduction to Machine Learning, Homework 4

109550136 邱弘竣

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)
   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%):

(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 K be a positive semidefinite matrix

$K_{ij} = K(x_i, x_j) = \phi(x_i)^T\phi(x_j) = \phi(x_j)^T\phi(x_i) = K_{ji}$ ⇒ symmetric matrix

Thus, we have $K = V\Lambda V^T$

consider the feature map $\phi : x_i \mapsto (\sqrt{\lambda_t} V_{ti})_{t=1}^{n} \in R^n$

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

$\forall z$: $z^T K z = \sum_i \sum_j z_i k_{ij} z_j = \sum_i \sum_j z_i \phi(x_i)^T \phi(x_j) z_j$

↑
vector

$= \sum_i \sum_j z_i \sum_k \phi_k(x_i) \phi_k(x_j) z_j$

$= \sum_k \sum_i \sum_j z_i \phi_k(x_i) \phi_k(x_j) z_j$

$= \sum_k \left(\sum_i z_i \phi_k(x_i) \right)^2 \geq 0$

(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.

$exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$ (taylor expansion) → a polynomial with positive coefficients

$k(x, x') = exp(k_1(x, x')) = f(k_1(x, x'))$

each polynomial → a product of kernels with a positive coefficient

↓ textbook 6.15 $k(x, x') = g(k_1(x, x'))$

$k(x, x') = exp(k_1(x, x'))$ 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) $f(r) = r+1 \rightarrow$ polynomial function with positive coefficient

let $r = k_1(x,x')$

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

$\downarrow$ textbook 6.15

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

0.36
0.25
-0.16

(b) let $K = \begin{bmatrix} 0.5 & 0.4 \\ 0.4 & 0.5 \end{bmatrix}$

$|K - \lambda I| = \begin{vmatrix} 0.5-\lambda & 0.4 \\ 0.4 & 0.5-\lambda \end{vmatrix} = \lambda^2 - \lambda + 0.09 = (\lambda - 0.1)(\lambda - 0.9)$, $\lambda = 0.1, 0.9$

$K - 1 = \begin{bmatrix} -0.5 & -0.6 \\ -0.6 & -0.5 \end{bmatrix}$

$|K - 1 - \lambda I| = \begin{vmatrix} -0.5-\lambda & -0.6 \\ -0.6 & -0.5-\lambda \end{vmatrix} = \lambda^2 + \lambda - 0.11 = (\lambda + 1.1)(\lambda - 0.1)$, $\lambda = 0.1, -1.1$

< 0 #

$$\frac{2032}{17968}$$

3.(c) $k_1(x,x')^2$ is a valid function $\longleftarrow$ textbook 6.18

$$\exp(\|x^2\|)\exp(\|x'\|^2) = \left(1+x+\frac{(x^2)^2}{2!}+\frac{(x^2)^3}{3!}+\cdots\right)\left(1+x+\frac{(x'^2)^2}{2!}+\frac{(x'^2)^3}{3!}+\cdots\right)$$

$$= g(x)\,g(x') \geq 0$$

give $f(r) = r^2 + c$ with non-negative coefficient

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

$$\Big\downarrow \text{textbook 6.15}$$

$k(x,x') = k_1(x,x')^2 + \exp(\|x\|^2)\exp(\|x'\|^2)$ is a valid kernel

3.(d). $\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$ taylor expansion

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

$\Big\downarrow$ 6.18. $k_1(x,x')^2$ is a valid kernel

given $f(k_1(x,x')) = k_1(x,x') + \frac{k_1(x,x')^2}{2!} + \frac{k_1(x,x')^3}{3!} + \cdots$

$$\Big\downarrow 6.15,\ 6.17$$

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

(10%) Consider the optimization problem

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

State the dual problem.

4.

$(x+3)(x-1) \le 3 \; . \rightarrow -(x+3)(x-1) \ge 3$

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

$\frac{dL}{dx} = 2x-4+2xa+2a=0 \quad , \quad x = \frac{2-a}{1+a}$

$\frac{dL}{da} = x^2+2x-6=0$

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

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

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

the dual problem = maximize $L(a) = \frac{-7a^2+2a}{1+a}$, subject to $a \ge 0$