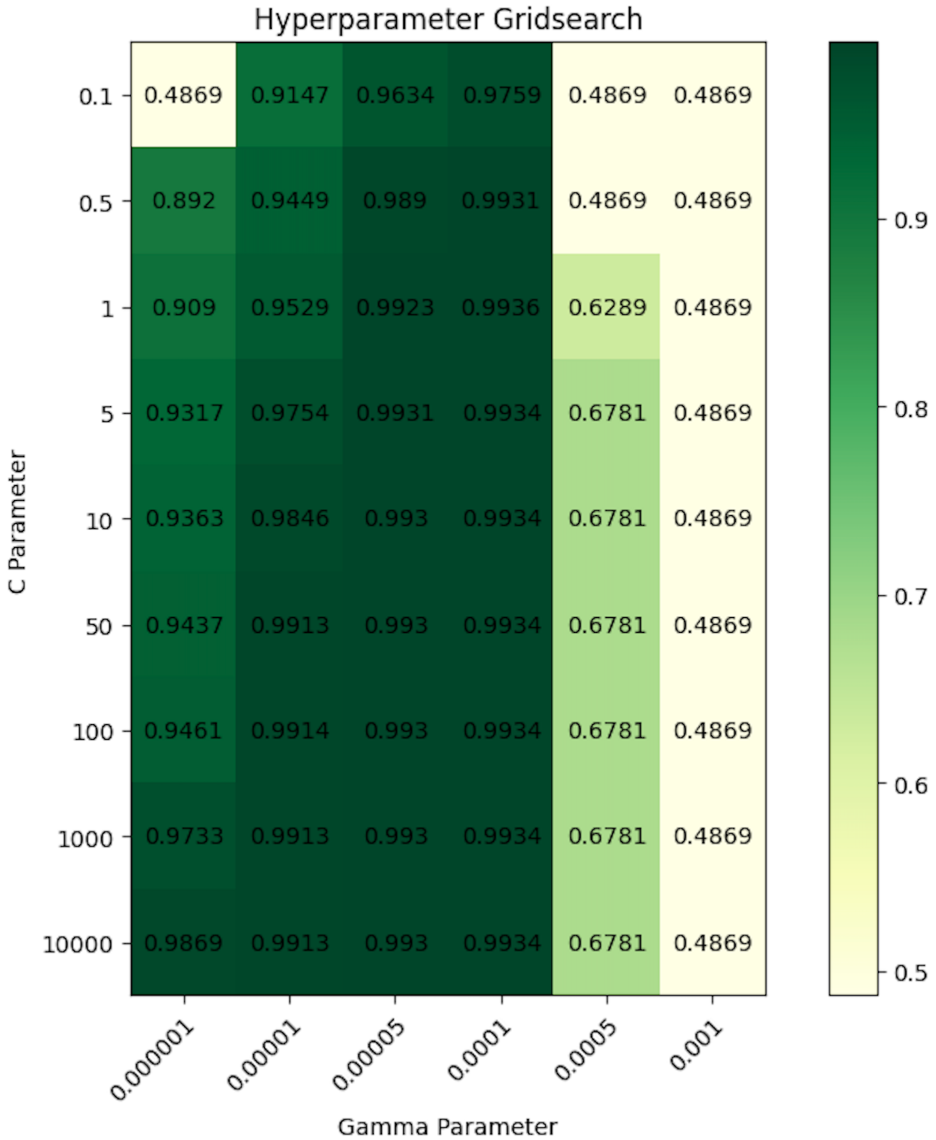
NCTU Introduction to Machine Learning, Homework 4

109550171 陳存佩

Part. 1, Coding (50%):

1. (10%) K-fold data partition
2. (20%) Grid Search & Cross-validation
3. (10%) Plot the grid search results of your SVM. The x and y represent "gamma" and "C" hyperparameters, respectively.



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%) 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 postive semidefinite matrix: all of its eigenvalues are non-negative

K為資料兩兩用內積相乘，故為 symmetric matrix (即 $i \cdot j = j \cdot i$)

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

K可分解為 $V \Lambda V^T$, V: orthonormal matrix v_t; Λ: diagomal matrix contains eigenvalues of K

存在一個 mapping function ϕ，投影 x_i 到 n 維空間 $\phi: x_i \mapsto (\sqrt{\lambda_t}\, v_{ti})_{t=1}^{n} \in \mathbb{R}^n$

故存在 function ϕ，使 $\phi(x_i)^T \phi(x_j) = \sum_{t=1}^{n} \lambda_t v_{ti} v_{tj} = (V \Lambda V^T)_{ij} = K_{ij} = k(x_i, x_j) \rightarrow$ valid

反向: if K is a valid kernel, it is positive semidefinite,

$K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$,

for any vector z: $z^T K z = \sum_i \sum_j z_i K_{ij} z_j = \sum_i \sum_j z_i \phi(x_i)^T \phi(x_j) z_j$

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

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

2. (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. $\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ taylor expansion

$k(x, x') = \exp(k_1(x, x')) = f(k_1(x, x'))$ f is a polynomial with positive coefficients

because 6.15 from textbook

$\rightarrow k(x, x') = \exp(k_1(x, x'))$ 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. $k(x, x') = k_1(x, x') + 1$

3. a. given a polynomial function with postive coefficients: $f(r) = r + 1$

let $r = k_1(x, x')$, then $k(x, x') = f(k_1(x, x')) = k_1(x, x') + 1$

because 6.15 from textbook:

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

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

3.b. 舉反例: let $K = \begin{bmatrix} 1.5 & -1 \\ -1 & 1.5 \end{bmatrix}$

$|K - \lambda I| = \begin{vmatrix} 1.5-\lambda & -1 \\ -1 & 1.5-\lambda \end{vmatrix} = \lambda^2 - 3\lambda + \frac{5}{4}$, $\lambda = \frac{1}{2}$ or $\frac{5}{2}$

$K - 1 = \begin{bmatrix} 0.5 & -2 \\ -2 & 0.5 \end{bmatrix}$

$|K - 1 - \lambda I| = \begin{vmatrix} 0.5-\lambda & -2 \\ -2 & 0.5-\lambda \end{vmatrix} = \lambda^2 - \lambda - \frac{15}{4}$, $\lambda = \frac{5}{2}$ or $\frac{-3}{2} < 0$

not valid

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

3.c. ① from 6.18, $k_1(x, x')^2$ is a valid kernel

② $exp(\|x^2\|) \cdot exp(\|x'\|^2) = (1 + x + \frac{(x^2)^2}{2!} + \frac{(x^2)^3}{3!} + \dots)(1 + x + \frac{(x'^2)^2}{2!} + \frac{(x'^2)^3}{3!} + \dots) = g(x)g(x') \geq 0$

let $C = g(x)g(x') \geq 0$ where $g(x) = 1 + x + \frac{x^4}{2!} + \frac{x^6}{3!} + \dots$

given a polynomial function with non-negative coefficients: $f(r) = r^2 + c$

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

because 6.15 from textbook:

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

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

3.d. $exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ taylor expansion

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

from 6.18, $k_1(x, x')^2$ is a valid kernel

given a polynomial function with postive coefficients

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

because 6.15 & 6.17 from textbook

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

4. (10%) Consider the optimization problem

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

State the dual problem.

4. $(x+3)(x-1) \leq 3 \longrightarrow -(x+3)(x-1) \geq 3$

lagrangian function: $L(x,a) = (x-2)^2 - a[(-x-3)(x-1)+3]$

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

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

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

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