Homework 6

Lingyu Zhou

April 19, 2024

Problem 1

(a)

Initialize $\alpha = \mathbf{0} \in \mathbb{R}^n$;
while *TRUE* **do**
    m = 0 ;
    for $(\mathbf{x}_i, y_i) \in D$ **do**
        if $y_i \left(\sum_j \alpha_j y_j x_j\right)^\top x_i \leq 0$ **then**
            $\alpha_i \leftarrow \alpha_i + 1$;
            $m \leftarrow m + 1$;
        end
        if $m = 0$ **then**
            break
        end
    end
end

(b)

Initialize $\vec{\alpha} = \mathbf{0} \in \mathbb{R}^n$;
while *TRUE* **do**
    m = 0 ;
    for $(\mathbf{x}_i, y_i) \in D$ **do**
        if $y_i \sum_j \alpha_j y_j k(x_j, x_i) \leq 0$ **then**
            $\alpha_i \leftarrow \alpha_i + 1$;
            $m \leftarrow m + 1$;
        end
        if $m = 0$ **then**
            break
        end
    end
end

(c)

Initialize $\vec{\alpha} = \mathbf{0} \in \mathbb{R}^n$;
Initialize $\boldsymbol{K}$ of size $n \times n$;
for $i \in \{1, \ldots, n\}$ **do**
    for $j \in \{1, \ldots, n\}$ **do**
        $\boldsymbol{K}_{ij} = k(x_i, x_j)$
    end
end
while *TRUE* **do**
    m = 0 ;
    for $(\mathbf{x}_i, y_i) \in D$ **do**
        if $y_i \sum_j \alpha_j y_j \boldsymbol{K}_{ji} \leq 0$ **then**
            $\alpha_i \leftarrow \alpha_i + 1$;
            $m \leftarrow m + 1$;
        end
        if $m = 0$ **then**
            break
        end
    end
end

(d)

$$h(x^*) = \operatorname{sgn}\left(\sum_i \alpha_i y_i x_i x^*\right) = \operatorname{sgn}\left(\sum_i \alpha_i y_i k(x_1, x^*)\right)$$

Problem 2

Since k_1,k_2 are valid kernels, the kernel matrices related to it $K^{(1)}, K^{(2)}$ satisfies:
1. $K^{(1)}_{ij} = K^{(1)}_{ji}; K^{(2)}_{ij} = K^{(2)}_{ji}$
2. $K^{(1)} \succeq 0 \Rightarrow v^\top K^{(1)} v \geq 0; K^{(2)} \succeq 0 \Rightarrow v^\top K^{(2)} v \geq 0, \forall v \geq \mathbf{0}$.

(a)

Symmetric: $K_{ij} = k(x_i, x_j) = ck_1(x_i, x_j) = K^{(1)}_{ij} = K^{(1)}_{ji} = ck_1(x_j, x_i) = k(x_j, x_i) = K_{ji}$
Positive semidefinite: Since $k(x_i, x_j) = ck_1(x_i, x_j), \forall i, j$, it follows that $K = cK^{(1)}$, therefore for same v mentioned above, it follows that $v^\top K v = v^\top cK^{(1)} v = c \underbrace{v^\top K^{(1)} v}_{\geq 0} \geq 0, \forall c \geq 0 \Rightarrow K \succeq 0$

(b)

Symmetric:$K_{ij} = k_1(x_i, x_j) + k_2(x_i, x_j) = K^{(1)}_{ij} + K^{(2)}_{ij} = K^{(1)}_{ji} + K^{(2)}_{ji} = k_1(x_j, x_i) + k_2(x_j, x_i) = k(x_j, x_i) = K_{ji}$
Positive semidefinite: Since $k(x_i, x_j) = k_1(x_i, x_j) + k_2(x_i, x_j), \forall i, j$, it follows $K_{ij} = K^{(1)}_{ij} + K^{(2)}_{ij} \Rightarrow K = K^{(1)} + K^{(2)}$. Therefore $v^\top K v = v^\top \left(K^{(1)} + K^{(2)}\right) v = \left(v^\top K^{(1)} + v^\top K^{(2)}\right) v = \underbrace{v^\top K^{(1)} v}_{\geq 0} + \underbrace{v^\top K^{(2)} v}_{\geq 0} \geq 0, \forall v \geq \mathbf{0}$

(c)

Since k_1 is a kernel, it follows that $k_1(x_i, x_j) = \phi_1(x_i)^\top \phi_1(x_i)$, for some ϕ as a valid transformation. Also since $f(\cdot)$ is a scalar-valued function, it follows that $\sqrt{f(\cdot)}$ is also one. Also note that if $\phi(\cdot)$ is a transformation from $\mathbb{R}^d$ to $\mathbb{R}^m$, so does $c\phi(\cdot), \forall c \in \mathbb{R}$ as scalar does not affect the dimension of transformation. Therefore since $\sqrt{f(\cdot)}$ outputs a scalar, by given. $\sqrt{f(\cdot)}\phi_1(\cdot)$ is also a transformation from $\mathbb{R}^d$ to $\mathbb{R}^m$. Let $\phi(x) = \sqrt{f(x)}\phi_1(x)$, therefore $\phi(x_i)^\top \phi(x_j) := k(x_i, x_j)$ is a **valid kernel**, which by what we conclude above, simplifies to:

$$
\begin{aligned}
&= \left(\sqrt{f(x_i)}\phi_1(x_i)\right)^\top \left(\sqrt{f(x_j)}\phi_1(x_j)\right) \\
&= \sqrt{f(x_i)}\phi_1(x_i)^\top \sqrt{f(x_j)}\phi_1(x_j) \\
&= \sqrt{f(x_i)}\phi_1(x_i)^\top \phi_1(x_j)\sqrt{f(x_j)} \\
&= \sqrt{f(x_i)}k_1(x_i, x_j)\sqrt{f(x_j)}
\end{aligned}
$$

Hence we showed a product of 2 scalar-valued functions and another valid kernel, whose name is $k(x_1, x_2)$ is a **valid kernel**.

Problem 3

(a)

$$
\begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) \\ k(x_2, x_1) & k(x_2, x_2) \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}
$$

(b)

The τ^2 is like the penalty term work as a regularization term to prevent overfitting. By modifying τ, one can control the scale of the Kernel to modify the scale of penalty. $+\tau^2 I$ increases the diagonal entries by τ^2 many and help the kernel less suspectible to numeric instability especially when the kernel matrix is nearly singular.
Also, since K is already positive semi-definite, $K + \tau^2 I$ will make the new kernel matrix positive definite as $v^\top K' v = v^\top (K + \tau^2 I)v = v^\top K v + v^\top \tau^2 I v \geq \tau^2 > 0$. Hence making the kernel matrix symmetric positive-definite can guarantee its invertiblilty.

(c)

$$
\alpha = K^{-1} y = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}
$$

$h(x_1^*) = \sum_i k(x_1^*, x_i)\alpha_i = k(x_1^*, x_1)\alpha_1 + k(x_1^*, x_2)\alpha_2 = k(0, -1) \times 0 + k(0, 1) \times \frac{1}{2} = \frac{1}{2}$
$h(x_2^*) = \sum_i k(x_2^*, x_i)\alpha_i = k(x_2^*, x_1)\alpha_1 + k(x_2^*, x_2)\alpha_2 = k(2, -1) \times 0 + k(2, 1) \times \frac{1}{2} = \frac{9}{2}$