Foundations of Machine Learning

Subject Code: CS5590

Assignment-2

Submitted by:

Antala Aviraj
Roll No: CS24MTECH14011

Department of Computer Science and Engineering
IIT-Hyderabad

Date: September 23, 2024

September 23, 2024

Answer 1)

Hard-Margin SVM with Original Constraints

For a hard-margin case,the margin boundaries are given by:

$$w \cdot x + b = +1 \quad \text{(Positive class)}$$

$$w \cdot x + b = -1 \quad \text{(Negative class)}$$

Margin is the distance between these two hyperplanes.
So margin m can be calculated as:

$$m = \frac{b+1}{\|w\|} - \frac{b-1}{\|w\|}$$

$$m = \frac{2}{\|w\|}$$

Now our objective is to maximize the margin, which is equivalent to minimizing $\frac{1}{2}\|w\|$

$$y_i(w \cdot x_i + b) \geq 1 \quad \forall i$$

where $y_i \in \{-1, +1\}$

Now let's put γ as some arbitary constant:

Put $+1$ and -1 with arbitrary constants $+\gamma$ and $-\gamma$, where $\gamma > 0$. then, So margin m can be calculated as:

$$m = \frac{b+\gamma}{\|w\|} - \frac{b-\gamma}{\|w\|}$$

$$m = \frac{2\gamma}{\|w\|}$$

Our objective is still to maximize the margin, which is equivalent to minimizing $\frac{1}{2}\|w\|$
So from above both equation we can observe that our objective is same.

Divide the equation in both side by γ:

$$\frac{y_i(w \cdot x_i + b)}{\gamma} \geq \frac{\gamma}{\gamma} \quad \forall i$$

Rewrite this as:

$$y_i \left(\frac{w}{\gamma} \cdot x_i + \frac{b}{\gamma} \right) \geq 1 \quad \forall i$$

Now take $w' = \frac{w}{\gamma}$ and $b' = \frac{b}{\gamma}$. The new problem becomes:

$$y_i(w' \cdot x_i + b') \geq 1 \quad \forall i$$

Which is exactly same as the original SVM formulation, with new parameters w' and b'.
So we can say that solution for the maximum margin hyperplane is to minimize $\frac{1}{\|w\|}$

Now, if we had arbitrarily chosen γ instead of 1, the margin would be:

$$\frac{2\gamma}{\|w\|}$$

So this is just a change in scaling, and the optimization problem is same in both case.

Answer 2)

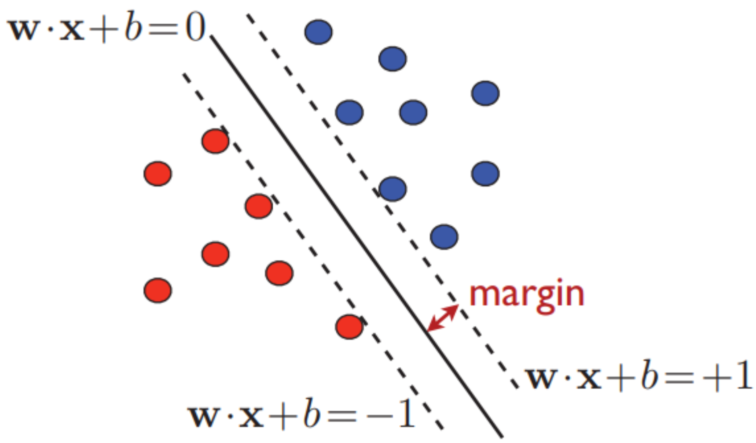

Figure 1: SVM Hyperplane and Support vector

Now in question already defined that half margin of the maximum margin is:

$$\rho = \frac{1}{\|w\|}$$

In the context of support vector machines, we know that the target values for the support vectors are either $+1$ or -1.
We can represent these target values using y_i, where y_i takes the value $+1$ for positive class and -1 for the negative class.

For a support vector, we have the equation:

$$w \cdot x_i + b = y_i$$

So, the bias term b can be:

$$b = y_i - w \cdot x_i \tag{1}$$

From the dual formulation of the SVM, the weight vector w can be expressed as:

$$w = \sum_{j=1}^{N} \alpha_j y_j x_j \tag{2}$$

α_j are the Lagrange multipliers, y_j are the target values, and x_j are the support vectors. Putting this into the bias term equation (1), we get:

$$b = y_i - \sum_{j=1}^{N} \alpha_j y_j (x_j \cdot x_i) \tag{3}$$

Multiplying both sides by $\alpha_i y_i$ and taking the sum leads to:

$$\sum_{i=1}^{N} \alpha_i y_i b = \sum_{i=1}^{N} \alpha_i y_i^2 - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{4}$$

Target value y_i can either $+1$ or -1,for support vector we know that y_i^2 is equal to 1.

Therefore, we can replace y_i^2 with 1 in our equations(4).

We know that partial derivative with respect to b is given by:

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \tag{5}$$

Putting it into equation (4), So we have the following equation:

$$\sum_{i=1}^{N} \alpha_i y_i^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{6}$$

$$\|w\|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i^T x_j) \tag{7}$$

Putting equation(7) into equation(6)

$$\sum_{i=1}^{N} \alpha_i y_i^2 = \|w\|^2 \tag{8}$$

From initial equation we have,

$$\|w\|^2 = \frac{1}{\rho^2} = \sum_{i=1}^{N} \alpha_i y_i^2$$

Hence, the proof is complete.

Answer 3)

To determine if any kernel $k(x, z)$ is valid or not , we need to verify two properties: **symmetry** and **positive semi-definiteness (PSD)**

a) $k(x, z) = k_1(x, z) + k_2(x, z)$

Symmetry:

For $k(x, z)$ to be symmetric, it must satisfy $k(x, z) = k(z, x)$.

Assume $k_1(x, z)$ and $k_2(x, z)$ are valid kernels (it is mention in the question also) and are individually symmetric:

$$k_1(x, z) = k_1(z, x) \quad \text{and} \quad k_2(x, z) = k_2(z, x)$$

So,the sum of these kernels is also symmetric:

$$k(x, z) = k_1(x, z) + k_2(x, z) = k_1(z, x) + k_2(z, x) = k(z, x)$$

Positive Semi-Definiteness (PSD):

A kernel $k(x, z)$ is PSD if for any finite set of points $x_1, x_2, \ldots, x_n$ and any real coefficients $a_1, a_2, \ldots, a_n$:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

In the question mention that k_1 and k_2 are valid kernels, they are both PSD. So:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k_1(x_i, x_j) \geq 0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k_2(x_i, x_j) \geq 0$$

Adding these inequalities gives:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (k_1(x_i, x_j) + k_2(x_i, x_j)) \geq 0$$

so form we can say that:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

So we can say that $k(x, z)$ is both symmetric and PSD, it is a valid kernel.

b) $k(x, z) = k_1(x, z)k_2(x, z)$

Symmetry:

For $k(x, z)$ to be symmetric, it must satisfy $k(x, z) = k(z, x)$.
Assume that $k_1(x, z)$ and $k_2(x, z)$ are both valid kernels, they are individually symmetric:

$$k_1(x, z) = k_1(z, x) \quad \text{and} \quad k_2(x, z) = k_2(z, x)$$

Thus, the product of these kernels is also symmetric:

$$k(x, z) = k_1(x, z)k_2(x, z) = k_1(z, x)k_2(z, x) = k(z, x)$$

Positive Semi-Definiteness (PSD):

A kernel $k(x, z)$ is PSD if for any finite set of points $x_1, x_2, \ldots, x_n$ and any real coefficients $a_1, a_2, \ldots, a_n$:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

Given that both k_1 and k_2 are valid kernels, they are both PSD. So:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k_1(x_i, x_j) \geq 0$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k_2(x_i, x_j) \geq 0$$

Now, we can express the product $k(x, z)$ as:

$$k(x, z) = k_1(x, z)k_2(x, z)$$

For prove that $k(x, z)$ is PSD, we can use the fact that if k_1 and k_2 are both PSD, then for any non-negative vectors $\mathbf{a}$:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k_1(x_i, x_j) k_2(x_i, x_j)$$

By the properties of positive semi-definiteness, we can state:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k_1(x_i, x_j) \geq 0$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k_2(x_i, x_j) \geq 0$$

Using the fact that the product of two PSD matrices is also PSD , we can conclude that:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

So we can say that $k(x, z)$ is both symmetric and PSD, it is a valid kernel.

c) $k(x, z) = h(k_1(x, z))$

Symmetry:

For $k(x, z)$ to be symmetric, it must satisfy $k(x, z) = k(z, x)$.
Assume that $k_1(x, z)$ is a valid kernel, and so it is symmetric:

$$k_1(x, z) = k_1(z, x)$$

Let h be a polynomial function with positive coefficients, we can write that as:

$$h(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

where $a_i \geq 0$ for all i. Applying this polynomial function to the kernel $k_1(x, z)$:

$$k(x, z) = h(k_1(x, z)) = a_0 + a_1 k_1(x, z) + a_2 (k_1(x, z))^2 + \cdots + a_m (k_1(x, z))^m$$

We know that $k_1(x, z)$ is symmetric:

$$k_1(x, z) = k_1(z, x)$$

So:

$$k(x, z) = h(k_1(x, z)) = h(k_1(z, x)) = k(z, x)$$

So we can say that $k(x, z)$ is symmetric.

Positive Semi-Definiteness (PSD):

Now $h(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$, where $a_i > 0$ for all i.
We want to prove that $k(x, z) = h(k_1(x, z))$ is a PSD , given that $k_1(x, z)$ is a PSD and valid kernal.

Applying this polynomial function to the kernel $k_1(x, z)$:

$$k(x, z) = h(k_1(x, z)) = a_0 + a_1 k_1(x, z) + a_2 (k_1(x, z))^2 + \cdots + a_m (k_1(x, z))^m$$

Each term $(k_1(x, z))^m$ (for $m = 1, 2, \ldots, n$) is a product of the kernel $k_1(x, z)$ with itself m times.

- Since $k_1(x, z)$ is a kernel, $(k_1(x, z))^m$ is also a kernel due to the closure property of kernels under multiplication.(we already prove that in question 2)
- If $k_1(x, z)$ is a kernel, then $(k_1(x, z))^m$ is a kernel.
- Each term $a_m (k_1(x, z))^m$ is a positive combination of kernels, as $a_m > 0$(because in question it is mention that with positive co-efficients.

So, we con say that each term $a_m (k_1(x, z))^m$ for $m = 1, \ldots, n$ is a kernel.
Now, $h(k_1(x, z))$ is a finite sum of kernels multiplied by positive coefficients:

$$h(k_1(x, z)) = a_n (k_1(x, z))^n + a_{n-1} (k_1(x, z))^{n-1} + \ldots + a_1 k_1(x, z) + a_0.$$

According to question 1, since each term is a kernel, the sum $h(k_1(x, z))$ is also a kernel.

So we can say that $k(x, z)$ is both symmetric and PSD, it is a valid kernel.

d) $k(x, z) = \exp(k_1(x, z))$

Symmetry:

For $k(x, z)$ to be symmetric, it must satisfy $k(x, z) = k(z, x)$.
Assume that $k_1(x, z)$ is a valid kernel, and so it is symmetric:

$$k_1(x, z) = k_1(z, x)$$

So the exponential function $\exp(x)$ is symmetric:

$$\exp(k_1(x, z)) = \exp(k_1(z, x))$$

So, the $k(x, z) = \exp(k_1(x, z))$ is also symmetric:

$$k(x, z) = \exp(k_1(x, z)) = \exp(k_1(z, x)) = k(z, x)$$

Positive Semi-Definiteness (PSD):

A kernel is PSD if for any set of points $\{x_1, x_2, ..., x_n\}$, the Gram matrix K with entries $K_{ij} = k(x_i, x_j)$ satisfies $z^T K z \geq 0$ for all $z \in R^n$.

We assume that $k_1(x, z)$ is a valid kernel, so its Gram matrix K_1 is PSD. For any vector $z \in R^n$:

So we can say that,

$$z^T K_1 z \geq 0$$

Applying the exponential function to $k_1(x, z)$, we get $k(x, z) = \exp(k_1(x, z))$.

To prove that $k(x, z) = \exp(k_1(x, z))$ defines a positive semi-definite (PSD) kernel, given that $k_1(x, z)$ is a valid kernel(given in the question).

We know that the exponential function can be expressed as a power series,so we can write like this:

$$\exp(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!}.$$

Applying this to our kernel function:

$$k(x, z) = \exp(k_1(x, z)) = \sum_{i=0}^{\infty} \frac{(k_1(x, z))^i}{i!}.$$

For a finite set of points $\{x_1, x_2, \ldots, x_n\}$, let K_1 be the kernel matrix relate to $k_1(x, z)$.

The matrix K with entries $K_{ij} = \exp(k_1(x_i, x_j))$ can be expressed as:

$$K = \sum_{i=0}^{\infty} \frac{(K_1)^i}{i!}.$$

In question it is given that $k_1(x,z)$ is a valid kernel, the matrix K_1 is PSD. A known result in matrix theory states that if M is a PSD matrix, then the matrix exponential series:

$$\sum_{i=0}^{\infty} \frac{M^i}{i!}$$

is also PSD. This implies that the matrix K is PSD.

So we can say that $k(x,z) = \exp(k_1(x,z))$ is valid kernal.

e) $k(x,z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$

Symmetry:

To check whether $k(x,z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$ is symmetric, we need to verify that if $k(x,z) = k(z,x)$.

So we know that:

$$\|x-z\|^2 = \|z-x\|^2$$

So let's look this:

$$k(x,z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\|z-x\|^2}{2\sigma^2}\right) = k(z,x)$$

So we can say that $k(x,z)$ is symmetric.

Positive Semi-Definiteness (PSD):

Let's start with the Gaussian kernel:

$$k(x,z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$$

Expanding the squared term:

$$\|x-z\|^2 = \|x\|^2 - 2x^T z + \|z\|^2$$

So now we have:

$$k(x,z) = \exp\left(-\frac{\|x\|^2 - 2x^T z + \|z\|^2}{2\sigma^2}\right)$$

We also write this as:

$$k(x,z) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|z\|^2}{2\sigma^2}\right) \exp\left(\frac{x^T z}{\sigma^2}\right)$$

Now assume that $g(x)$, $g(z)$ is valid kernal function:

$$g(x) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)$$

$$g(z) = \exp\left(-\frac{\|z\|^2}{2\sigma^2}\right)$$

$$k_1(x, z) = \frac{x^T z}{\sigma^2}$$

finally we have:

$$k(x, z) = g(x)g(z)\exp\left(k_1(x, z)\right)$$

And we already know that from question b) that the product of two valid kernels is also a valid kernel.

So we can say that $g(x)g(z)$ and $\exp\left(k_1(x, z)\right)$ are both valid kernels, their product $g(x)g(z)\exp\left(k_1(x, z)\right)$ is also a valid kernel.