Assignment Report for ELL409

Assignment 2

Shubh Goel
(2020EE10672)
Prateek Mishra
(2020EE10527)
Bharat Kumar
(2020EE10587)

Department of Electrical Engineering

December 12, 2023

Contents

1 Kernel Functions and Feature Maps

$x, z \in \mathcal{R}^d$. Let,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_d \end{bmatrix}$$

Proving that the given kernel function, $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$ is a valid kernel function and finding the corresponding feature map, ϕ.

$$\begin{aligned} K(\mathbf{x}, \mathbf{z}) &= (\mathbf{x}^T\mathbf{z})^2 \\ &= \left(\sum_{i=1}^{d} x_i z_i\right)\left(\sum_{j=1}^{d} x_j z_j\right) \\ &= \sum_{i=1}^{d}\sum_{j=1}^{d} x_i z_i x_j z_j \\ &= \sum_{i=1}^{d}\sum_{j=1}^{d} x_i x_j z_i z_j \\ &= \sum_{1 \leq i,j \leq d} (x_i x_j)(z_i z_j) \end{aligned}$$

Hence, the given kernel function is a valid kernel function that corresponds to the following feature map:

$$\phi(\mathbf{x}) = \begin{bmatrix} x_1 x_1 \\ x_1 x_2 \\ \vdots \\ x_1 x_d \\ \vdots \\ x_d x_{d-1} \\ x_d x_d \end{bmatrix} \tag{1}$$

To find the computational efficiency of the given kernel functions and their corresponding feature maps, we are going to use the Big-O notation.

1.1 $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$

From equation 1 we can see that computation of the feature map $\phi(\mathbf{x})$ requires $O(d^2)$, whereas for the computation of kernel function requires $O(d)$.

1.2 $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + c)^2$

$$\begin{aligned} K(\mathbf{x}, \mathbf{z}) &= (\mathbf{x}^T\mathbf{z} + c)^2 \\ &= (\mathbf{x}^T\mathbf{z})^2 + 2\mathbf{x}^T\mathbf{z} + c^2 \\ &= \sum_{1 \le i,j \le d} (x_i x_j)(z_i z_j) + \sum_{i=1}^{d} (\sqrt{2c}x_i)(\sqrt{2c}z_i) + c^2 \end{aligned}$$

Hence, the given kernel function is a valid kernel function that corresponds to the following feature map:

$$\phi(\mathbf{x}) = \begin{bmatrix} x_1 x_1 \\ x_1 x_2 \\ \vdots \\ x_1 x_d \\ \vdots \\ x_d x_{d-1} \\ x_d x_d \\ \sqrt{2c}x_1 \\ \sqrt{2c}x_2 \\ \vdots \\ \sqrt{2c}x_d \\ c \end{bmatrix} \tag{2}$$

Computational Efficiencies:

- For kernel function $= O(d)$.
- For $\phi(\mathbf{x}) = O(d^2 + d + 1) \approx O(d^2)$

1.3 $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^k$

$$\begin{aligned} K(\mathbf{x}, \mathbf{z}) &= (\mathbf{x}^T\mathbf{z})^k \\ &= \prod_{l=1}^{k} \left(\sum_{i_l=1}^{d} x_{i_l} z_{i_l} \right) \\ &= \sum_{i_1=1}^{d} \cdots \sum_{i_k=1}^{d} \left(\prod_{l=1}^{k} (x_{i_l} z_{i_l}) \right) \\ &= \sum_{1 \le i_1, i_2, \ldots i_k \le d} (x_{i_1} x_{i_2} \ldots x_{i_k})(z_{i_1} z_{i_2} \ldots z_{i_k}) \end{aligned}$$

Hence, the given kernel function is a valid kernel function that corresponds to the following feature map:

$$\phi(\mathbf{x}) = \begin{bmatrix} x_1^k \\ x_1^{k-1}x_2 \\ \vdots \\ x_1^{k-1}x_d \\ \vdots \\ x_d^{k-1}x_{d-1} \\ x_d^k \end{bmatrix} \tag{3}$$

Computational Efficiencies:

- For kernel function $= O(d)$.
- For $\phi(\mathbf{x}) = O(d^k + d^{k-1} + \cdots + 1) \approx O(d^k)$

In a high-dimensional feature space, we would prefer kernel mapping over feature mapping. We can conclude this by observing the computational efficiencies of the kernel function and their corresponding feature maps. For instance, in the case of example 1.3, even though we are working in a $O(d^k)$ feature space, computing the kernel function still only takes $O(d)$ time.

2 Kernel Algebra

For the subsequent parts, let's define the feature map of $k_a(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}_a(\mathbf{x})$ and $k_b(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}_b(\mathbf{x})$. The corresponding feature maps for the following Kernel functions are:

2.1 $K(\mathbf{x}, \mathbf{z}) = fk_a(\mathbf{x}, \mathbf{z})$

We know, $k_a(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}_a(\mathbf{x})^T \boldsymbol{\phi}_a(\mathbf{z})$. Thus,

$$\begin{aligned} K(\mathbf{x}, \mathbf{z}) &= fk_a(\mathbf{x}, \mathbf{z}) \\ &= f\boldsymbol{\phi}_a(\mathbf{x})^T \boldsymbol{\phi}_a(\mathbf{z}) \\ &= (\sqrt{f}\boldsymbol{\phi}_a(\mathbf{x}))^T (\sqrt{f}\boldsymbol{\phi}_a(\mathbf{z})) \qquad [\because, f > 0] \end{aligned}$$

Hence, feature map of $K(\mathbf{x}, \mathbf{z}) = fk_a(\mathbf{x}, \mathbf{z})$:

$$\boxed{\boldsymbol{\phi}(\mathbf{x}) = \sqrt{f}\boldsymbol{\phi}_a(\mathbf{x})}$$

2.2 $K(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T \mathbf{A} \mathbf{z}$

Since, $\mathbf{A}$ is a positive semi-definite matrix, there exists a matrix $\mathbf{B}$, such that $\mathbf{A} = \mathbf{B}^T\mathbf{B}$.

$$\begin{aligned} K(\mathbf{x}, \mathbf{z}) &= \mathbf{x}^T \mathbf{A} \mathbf{z} \\ &= \mathbf{x}^T \mathbf{B}^T \mathbf{B} \mathbf{z} \qquad \text{where } \mathbf{A} = \mathbf{B}^T\mathbf{B} \\ &= (\mathbf{B}\mathbf{x})^T (\mathbf{B}\mathbf{z}) \end{aligned}$$

Hence, feature map of $K(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T \mathbf{A} \mathbf{z}$:

$$\boxed{\boldsymbol{\phi}(\mathbf{x}) = \mathbf{B}\mathbf{x}}$$

2.3 $K(\mathbf{x}, \mathbf{z}) = f(\mathbf{x}) \cdot f(\mathbf{z}) \cdot k_a(\mathbf{x}, \mathbf{z})$

We know, $k_a(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}_a(\mathbf{x})^T \boldsymbol{\phi}_a(\mathbf{z})$. Thus,

$$
\begin{aligned}
K(\mathbf{x}, \mathbf{z}) &= f(\mathbf{x}) \cdot f(\mathbf{z}) \cdot k_a(\mathbf{x}, \mathbf{z}) \\
&= f(\mathbf{x}) \cdot \boldsymbol{\phi}_a(\mathbf{x})^T \cdot f(\mathbf{z}) \cdot \boldsymbol{\phi}_a(\mathbf{z}) \\
&= (f(\mathbf{x}) \cdot \boldsymbol{\phi}_a(\mathbf{x}))^T (f(\mathbf{z}) \cdot \boldsymbol{\phi}_a(\mathbf{z}))
\end{aligned}
$$

Hence, feature map of $K(\mathbf{x}, \mathbf{z}) = f(\mathbf{x}) \cdot f(\mathbf{z}) \cdot k_a(\mathbf{x}, \mathbf{z})$:

$$\boxed{\boldsymbol{\phi}(\mathbf{x}) = f(\mathbf{x})\boldsymbol{\phi}_a(\mathbf{x})}$$

2.4 $K(\mathbf{x}, \mathbf{z}) = k_a(\mathbf{x}, \mathbf{z}) \cdot k_b(\mathbf{x}, \mathbf{z})$

We know that $k_a(\mathbf{x}, \mathbf{z}) = \left(\sum_{i=1}^{d_1} \boldsymbol{\phi}_{ai}(\mathbf{x}) \boldsymbol{\phi}_{ai}(\mathbf{z}) \right)$ and $k_b(\mathbf{x}, \mathbf{z}) = \left(\sum_{j=1}^{d_2} \boldsymbol{\phi}_{bj}(\mathbf{x}) \boldsymbol{\phi}_{bj}(\mathbf{z}) \right)$.

$$
\begin{aligned}
K(\mathbf{x}, \mathbf{z}) &= k_a(\mathbf{x}, \mathbf{z}) \cdot k_b(\mathbf{x}, \mathbf{z}) \\
&= \left(\sum_{i=1}^{d_1} \phi_{ai}(\mathbf{x}) \phi_{ai}(\mathbf{z}) \right) \left(\sum_{j=1}^{d_2} \phi_{bj}(\mathbf{x}) \phi_{bj}(\mathbf{z}) \right) \\
&= \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \left(\phi_{ai}(\mathbf{x}) \phi_{bj}(\mathbf{x}) \right) \left(\phi_{ai}(\mathbf{z}) \phi_{bj}(\mathbf{z}) \right) \\
&= \begin{bmatrix} \phi_{a_1}(\mathbf{x}) \boldsymbol{\phi}_b^T(\mathbf{z}) & \phi_{a_2}(\mathbf{x}) \boldsymbol{\phi}_b^T(\mathbf{z}) & \ldots & \phi_{a_{d_1}}(\mathbf{x}) \boldsymbol{\phi}_b^T(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \phi_{a_1}(\mathbf{x}) \boldsymbol{\phi}_b(\mathbf{z}) \\ \phi_{a_2}(\mathbf{x}) \boldsymbol{\phi}_b(\mathbf{z}) \\ \vdots \\ \phi_{a_{d_1}}(\mathbf{x}) \boldsymbol{\phi}_b^T(\mathbf{z}) \end{bmatrix} \\
&= (\boldsymbol{\phi}_a(\mathbf{x}) \otimes \boldsymbol{\phi}_b(\mathbf{x}))^T (\boldsymbol{\phi}_a(\mathbf{x}) \otimes \boldsymbol{\phi}_b(\mathbf{x}))
\end{aligned}
$$

Where, $\otimes$ represents Kronecker's product between two matrices.

Hence, feature map of $K(\mathbf{x}, \mathbf{z}) = k_a(\mathbf{x}, \mathbf{z}) \cdot k_b(\mathbf{x}, \mathbf{z})$:

$$\boxed{\boldsymbol{\phi}(\mathbf{x}) = \boldsymbol{\phi}_a(\mathbf{x}) \otimes \boldsymbol{\phi}_b(\mathbf{x})}$$

3 Valid Kernel Functions

3.1 Prove that Gaussian Kernel is a valid kernel

Radial Basis Function or Gaussian Kernel for a bandwidth $\sigma > 0$, defined as:

$$
\begin{aligned}
K(\mathbf{x}, \mathbf{z}) &= exp\left(\frac{-\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2} \right) \\
&= exp\left(-\frac{\mathbf{x}^T\mathbf{x} + \mathbf{z}^T\mathbf{z} - 2 \cdot \mathbf{x}^T\mathbf{z}}{2\sigma^2} \right) \\
&= \left(exp\left(\frac{-\mathbf{x}^T\mathbf{x}}{2\sigma^2} \right) \cdot exp\left(\frac{-\mathbf{z}^T\mathbf{z}}{2\sigma^2} \right) \cdot exp\left(\frac{2 \cdot \mathbf{x}^T\mathbf{z}}{2\sigma^2} \right) \right) \qquad \text{[Eq. 1]}
\end{aligned}
$$

Now, let us define a kernel $K_a(\mathbf{x},\mathbf{z}) = \mathbf{x}^T\mathbf{z}$ and using the following property of valid kernels:

- If $K_i(\mathbf{x},\mathbf{z})$ is a valid kernel, then $c * K_i$ is also a valid kernel.
- If $K_i(\mathbf{x},\mathbf{z})$ is a valid kernel, then $exp\,(K_i(\mathbf{x},\mathbf{z}))$ is also a valid kernel.
- If $K_i(\mathbf{x},\mathbf{z})$ is a valid kernel and f($\cdot$) is some function, then $f(\mathbf{x}) \cdot K_i(\mathbf{x},\mathbf{z}) \cdot f(\mathbf{z})$ is a valid kernel.

So, since K_a is a valid kernel, $2 * \frac{\mathbf{x}^T\mathbf{z}}{2\sigma^2}$ is also a valid kernel. Further $K_b(\mathbf{x},\mathbf{z}) = exp\left(\frac{2*\mathbf{x}^T\mathbf{z}}{2\sigma^2}\right)$ is also a valid kernel.

Now $K(\mathbf{x},\mathbf{z})$ can be written as follows:

$$K(\mathbf{x},\mathbf{z}) = f(\mathbf{x}) \cdot K_b(\mathbf{x},\mathbf{z}) \cdot f(\mathbf{z})$$

where f(x) $=exp\left(\frac{-\mathbf{x}^T\mathbf{x}}{2\sigma^2}\right)$. Using the 3rd property mentioned above, $K(\mathbf{x},\mathbf{z})$ is a valid kernel. Hence, proved.

3.2 Show that $K(\mathbf{x},\mathbf{z}) = \left(1 + \left(\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right)^T \cdot \left(\frac{\mathbf{z}}{\|\mathbf{z}\|_2}\right)\right)^3$ is a valid kernel

For this part let us define a kernel, $K_a(\mathbf{x},\mathbf{z}) = \mathbf{x}^T\mathbf{z}$.

Now we will use the following properties of a valid kernel:

- If $K_a(\mathbf{x},\mathbf{z})$ is a valid kernel, then $f(\mathbf{x}) \cdot K_a(\mathbf{x},\mathbf{z}) \cdot f(\mathbf{z})$ is also a valid kernel.
- If $K_a(\mathbf{x},\mathbf{z})$ is a valid kernel, then $a_i + K_a(\mathbf{x},\mathbf{z})$ is also a valid kernel.
- If $K_i(\mathbf{x},\mathbf{z})$ is a valid kernel, then $\prod_{i=1}^{m}(K_i(\mathbf{x},\mathbf{z}))$ is also a valid kernel.

Using 1st property for $f(\mathbf{x}) = \frac{1}{\|\mathbf{x}\|_2}$ on the defined K_a we get

$$K(\mathbf{x},\mathbf{z}) = \left(\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right)^T \cdot \left(\frac{\mathbf{z}}{\|\mathbf{z}\|_2}\right)$$

which is a valid kernel.

Now, using 2nd property we get,

$$K(\mathbf{x},\mathbf{z}) = \left(1 + \left(\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right)^T \cdot \left(\frac{\mathbf{z}}{\|\mathbf{z}\|_2}\right)\right)$$

which is a valid kernel.

Now using 3rd property we get,

$$K(\mathbf{x},\mathbf{z}) = \left(1 + \left(\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right)^T \cdot \left(\frac{\mathbf{z}}{\|\mathbf{z}\|_2}\right)\right)^3$$

as a valid kernel. Hence, proved.