ELL409: Machine Learning and Intelligence

I Semester 2022-2023

Kernel Methods

Assignment 2

Muvva Srija (2020EE30605)

Perugu Krishna Chaitanya Yadav (2020EE10521)

Devarakonda Ronith Kumar(2020EE10486)

Date of Submission : 26th October, 2022

1 Kernel Functions and Feature Maps

We find corresponding feature mapping for the given kernel functions and computational analysis on kernels

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

$$K(x, z) =< \phi(x), \phi(z) >$$
$$K(x, z) = \phi(x)^T \phi(z)$$
$$\text{Given, } K(x, z) = (x^T z)^2$$

Here let us consider the dimension of vector x and z be d

$$K(x, z) = \left(\sum_{i=1}^{d} x_i z_i \right) \left(\sum_{j=1}^{d} x_j z_j \right)$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} x_i z_i x_j z_j$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} (x_i x_j)(z_i z_j)$$

By comparing the above with the inner product, then corresponding feature mapping is

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

From a computational efficiency point of view, it's important to remember that calculating the high-dimensional feature mapping takes $O(d^2)$ time, but finding $K(x, z)$ only takes $O(d)$ time, which is proportional to the number of input attributes.

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

$$K(x, z) =< \phi(x), \phi(z) >$$
$$K(x, z) = \phi(x)^T \phi(z)$$
$$\text{Given, } K(x, z) = (x^T z + c)^2$$
$$K(x, z) = (x^T z)^2 + 2cx_T z + c^2$$

Here let us consider the dimension of vector x and z be d

$$K(x, z) = \left(\sum_{i=1}^{d} x_i z_i\right) \left(\sum_{j=1}^{d} x_j z_j\right) + c^2 + 2c \sum_{i=1}^{d} x_i z_i$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} x_i z_i x_j z_j + \sum_{i=1}^{d} \sqrt{2c} x_i \sqrt{2c} z_i + c^2$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} (x_i x_j)(z_i z_j) + \sum_{i=1}^{d} (\sqrt{2c} x_i)(\sqrt{2c} z_i) + c^2$$

This is a Kernel with feature mapping $\phi(x)$ as,

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

From a computational efficiency point of view, it's important to remember that calculating the high-dimensional feature mapping takes $O((d+1)^2) \sim O(d^2)$ time, but finding $K(x,z)$ only takes $O(d)$ time, which is proportional to the number of input attributes.

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

$$K(x,z) =< \phi(x), \phi(z) >$$
$$K(x,z) = \phi(x)^T\phi(z)$$

Given, $K(x,z) = (x^T z)^k$ Here let us consider the dimension of vector x and z be d

$$K(x,z) = \left(\sum_{i=1}^{d} x_i z_i\right)^k$$

We can write it as multinomial expansion where each term will be

$$= \frac{k!}{r_1!r_2!...r_d!}(x_1z_1)^{r_1}(xz_2)^{r_2}....(x_dz_d)^{r_d} \text{ where } r_1 + r_2..... + r_d = k$$

$$= \left(\sqrt{\frac{k!}{r_1!r_2!...r_d!}}x_1^{r_1}x_2^{r_2}...x_d^{r_d}\right)\left(\sqrt{\frac{k!}{r_1!r_2!...r_d!}}z_1^{r_1}z_2^{r_2}...z_d^{r_d}\right)$$

we got each term then,

$$K(x,z) = \sum_{\substack{r_1=0,r_2=0,..r_d=0\\ r_1+r_2.....+r_d=k}}^{k} \left(\sqrt{\frac{k!}{r_1!r_2!...r_d!}}x_1^{r_1}x_2^{r_2}...x_d^{r_d}\right)\left(\sqrt{\frac{k!}{r_1!r_2!...r_d!}}z_1^{r_1}z_2^{r_2}...z_d^{r_d}\right)$$

So, by this we can say that size of $\phi(x)$ is a column matrix of order $\binom{d+k-1}{d-1}$ $\times 1$.

Each term of feature mapping is of the form,

$$\sqrt{\frac{k!}{r_1!r_2!...r_d!}}x_1^{r_1}x_2^{r_2}...x_d^{r_d}\ , \forall \mathrm{r}_1 \geq 0, r_2 \geq 0, ..., r_d \geq 0 \text{ and } r_1 + r_2..... + r_d = k$$

Feature mapping can also be as following.. If we allow the duplicate terms in the feature mapping , then we can re write the kernel as

$$K(x,z)=\left(\sum_{i=1}^{d} x_i z_i\right)^k$$

$$K(x,z)=\left(\sum_{i_1=1}^{d} x_{i_1} z_{i_1}\right)\left(\sum_{i_2=1}^{d} x_{i_2} z_{i_2}\right)....\left(\sum_{i_k=1}^{d} x_{i_k} z_{i_k}\right)$$

$$=\sum_{i_1=1}^{d}\sum_{i_2=1}^{d}...\sum_{i_k=1}^{d}(x_{i_1}x_{i_2}...x_{i_k})(z_{i_1}z_{i_2}...z_{i_k})$$

$$=\phi^T(x)\phi(z)$$

$$\phi(x)=\begin{bmatrix} x_1x_1...x_1 \\ x_1^{k-1}x_2 \\ \vdots \\ x_2^{k-1}x_1 \\ x_2^k \\ \vdots \\ x_d^{k-1}x_{d-1} \\ x_d^k \end{bmatrix}$$

where each term of $\phi(x)$ is $x_{i_1}x_{i_2}...x_{i_k}$ for any $i_1, i_2..., i_k \in [1,d]$ So, by this we can say that size of feature mapping is a column matrix of order $d^k \times 1$.

From a computational efficiency point of view, it's important to remember that calculating the high-dimensional feature mapping takes $O(d^k)$ time, but finding $K(x,z)$ only takes $O(d)$ time, which is proportional to the number of input attributes.

2 Kernel Algebra

To develop Kernel Algebra and formulation correspondence between kernel function and feature mapping.

2.1 $\mathbf{K(x,z) = \boldsymbol{f}K_a(x,z), \boldsymbol{f} > 0}$

$$\text{We know that, } K(x,z) =< \phi(x), \phi(z) >$$
$$K(x,z) = fK_a(x,z)$$
$$= f < \phi_a(x), \phi_a(z) >$$
$$=< \sqrt{f}\phi_a(x), \sqrt{f}\phi_a(z) >$$
$$\text{So, the feature mapping for given kernel function is ,}$$
$$\phi(x) = \sqrt{f}\phi_a(x)$$

2.2 K(x,z)=x$^T Az$, A is a positive semi definite matrix

Given A is positive semi-definite matrix which implies that there exists a matrix D such that

$$A = D^T D$$
$$\implies \text{K(x,z)=x}^T Az$$
$$= x^T D^T Dz$$
$$= (Dx)^T (Dz)$$
$$\text{we know that, } K(x,z) = \phi(x)^T \phi(z)$$
$$\text{Hence, feature mapping of } K(x,z),\ \phi(x) = Dx$$

2.3 K(x,z)=f(x)f(z)K$_a(x,z)$

$$\text{We know that, } K(x,z) =< \phi(x), \phi(z) >$$
$$K(x,z) = f(x)f(z)K_a(x,z)$$
$$K(x,z) = f(x)f(z) < \phi_a(x), \phi_a(z) >$$

$$= f(x)f(z)\left(\sum_{i=1}^{d} \phi_a^{(i)}(x)\phi_a^{(i)}(z)\right)$$

$$= \sum_{i=1}^{d}(\phi_a^{(i)}(x)f(x))(f(z)\phi_a^{(i)}(z))$$
$$=< f(x)\phi_a(x), f(z)\phi_a(z) >$$

So feature mapping, $\phi(x)$= $f(x)\phi_a(x)$

2.4 $\mathbf{K(x,z) = K_a(x,z)K_b(x,z)}$

By point-wise product,we can say that if $K_a(x, z)$, $K_b(x, z)$ are both valid kernel functions, then their product will also be valid kernel function so we can find feature mapping for $K(x, z)$. Now, let

$$K_a(x,z) = \phi_a^T(x)\phi_a(z)$$

$$\phi_a(x) = \begin{bmatrix}\phi_a^{(1)}(x) & \phi_a^{(2)}(x) & \cdots & \phi_a^{(m)}(x)\end{bmatrix}^T$$

$$K_b(x,z) = \phi_b^T(x)\phi_b(z)$$

$$\phi_b(x) = \begin{bmatrix}\phi_b^{(1)}(x) & \phi_b^{(2)}(x) & \cdots & \phi_b^{(n)}(x)\end{bmatrix}^T$$

ϕ_a is a vector of m-dimension and ϕ_b is a vector of n-dimension

$$\text{We know that, } K(x,z) =< \phi(x), \phi(z) >$$
$$K(x,z) = K_a(x,z)K_b(x,z)$$
$$K(x,z) =< \phi_a(x), \phi_a(z) > . < \phi_b(x), \phi_b(z) >$$

$$= \left(\sum_{i=1}^{m} \phi_a^{(i)}(x)\phi_a^{(i)}(z)\right)\left(\sum_{j=1}^{n} \phi_b^{(j)}(x)\phi_b^{(j)}(z)\right)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}(\phi_a^{(i)}(x)\phi_b^{(j)}(x)).(\phi_a^{(i)}(z)\phi_b^{(j)}(z))$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}(\phi^{(ij)}(x)\phi^{(ij)}(z)) = \phi(x)^T\phi(z)$$

where $\phi^{(ij)}(x) = \phi_a^{(i)}(x)\phi_b^{(j)}(x)$, $\phi(x)$ is a $m \times n$ dimensional vector

$$\phi(x) = \begin{bmatrix} \phi_a^{(1)}(x)\phi_b^{(1)}(x) \\ \phi_a^{(1)}(x)\phi_b^{(2)}(x) \\ \vdots \\ \phi_a^{(1)}(x)\phi_b^{(n)}(x) \\ \phi_a^{(2)}(x)\phi_b^{(1)}(x) \\ \phi_a^{(2)}(x)\phi_b^{(2)}(x) \\ \vdots \\ \phi_a^{(m)}(x)\phi_b^{(n-1)}(x) \\ \phi_a^{(m)}(x)\phi_b^{(n)}(x) \end{bmatrix}$$

That is feature mapping of $K(x,z)$ is a $m \times n$ dimensional vector, which has a coordinate $\phi^{(ij)}(.)$ where each pair $(i,j)1 \leq i \leq m, 1 \leq j \leq n$ where $\phi^{(ij)}(x) = \phi_a^{(i)}(x)\phi_b^{(j)}(x)$ Thus feature mapping of $K(x,z)$ is ϕ.

3 Valid Kernel Functions

Checking validity of kernel functions and get introduced to Gaussian kernel function.

3.1 $\mathbf{K(x,z) = exp\left(-\frac{||x-z||_2^2}{2\sigma^2}\right)}$

$$K(x,z) = exp\left(-\frac{||x-z||_2^2}{2\sigma^2}\right) = exp\left(-\frac{||x||_2^2+||z||_2^2-2x^Tz}{2\sigma^2}\right)$$
$$= exp\left(-\frac{||x||_2^2}{2\sigma^2}\right) exp\left(-\frac{||z||_2^2}{2\sigma^2}\right) exp\left(\frac{2x^Tz}{2\sigma^2}\right)$$

And we know that $k_b(x,z) = x^Tz$ is a kernel with feature mapping as $[x_i]_{n\times1}$(a column matrix with entries x_i. So we can write
$exp\left(\frac{2k_b(x,z)}{2\sigma^2}\right) = 1 + \frac{2k_b(x,z)}{2\sigma^2} + \frac{(\frac{2k_b(x,z)}{2\sigma^2})^2}{2!} + ...$
From the results of **1.3** we can observe that above equation is the summation of valid kernels, which is also a valid kernel. let it be $ck_a(x,z)$

$$exp\left(\frac{2x^Tz}{2\sigma^2}\right) = \text{c}k_a(x,z) \text{ , where c is a constant.}$$
$$\text{and let } exp\left(-\frac{||x||_2^2}{2\sigma^2}\right) = f(x)$$

then $exp\left(-\frac{||z||_2^2}{2\sigma^2}\right)$will be$f(z)$ Now the kernel $K(x,z)$ can be written as

$$K(x,z) = f(x)f(z)ck_a(x,z)$$

from **2.1** we know that $ck_a(x,z)$ is a kernel (let it be $K_a(x,z)$ and from **2.3** $f(x)f(z)K_a(x,z)$ is a kernel(if $K_a(x,z)$ is a kernel). Hence $K(x,z)$ is a kernel is a valid Kernel.
Hence we proved that Gaussian Kernel is a valid kernel.

3.2 $\mathbf{K(x,z) = \left(1 + \left(\frac{x}{||x||_2}\right)^T \left(\frac{z}{||z||_2}\right)\right)^3}$

Given Kernel can be written as,

$$K(x,z) = \left(1 + \sum_{i=1}^{n} \frac{x_i z_i}{||x||||z||}\right)^3$$

$$= {}^3C_0 + {}^3C_1 \left(\sum_{i=1}^{n} \frac{x_i z_i}{||x||||z||}\right) + {}^3C_2 \left(\sum_{i=1}^{n} \frac{x_i z_i}{||x||||z||}\right)^2 + {}^3C_3 \left(\sum_{i=1}^{n} \frac{x_i z_i}{||x||||z||}\right)^3$$

let,

$$K_1(x,z) = {}^3C_0$$

$$K_2(x,z) = {}^3C_1 \left(\sum_{i=1}^{n} \frac{x_i z_i}{||x||||z||}\right)$$

$$K_3(x,z) = {}^3C_2 \left(\sum_{i=1}^{n} \frac{x_i z_i}{||x||||z||}\right)_2$$

$$K_4(x,z) = {}^3C_3 \left(\sum_{i=1}^{n} \frac{x_i z_i}{||x||||z||}\right)^3$$

From **1.3** We know that $K(x,z) = (x^T z)^k$ is a valid kernel.
Hence, $K_1(x,z), K_2(x,z), K_3(x,z), K_4(x,z)$ are valid so by summation rule of kernels (i.e.,

if $K^*(x,z) = K_a(x,z) + K_b(x,z)$ and if $K_a(x,z), K_b(x,z)$ are valid kernels then $K^*(x,z)$ is a valid kernel.) so the given kernel $K(x,z)$ is a valid kernel