

Pattern Recognition

Assignment #3

By Professor Ying Shen

2031566/Yang Han

November 24, 2020

1. Kernel computation cost

1). Consider we have a two-dimensional input space such that the input vector is $x = (x_1, x_2)^T$. Define the feature mapping $\phi(x) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)^T$. What is the corresponding kernel function, i.e. $K(x, z)$? Do not leave $\phi(x)$ in your final answer.

Solution:

$$K(x, z) = \phi(x)^T \cdot \phi(z) = (x_1^2, \sqrt{2}x_1x_2, x_2^2) \cdot (z_1^2, \sqrt{2}z_1z_2, z_2^2)^T = x_1^2z_1^2 + 2x_1x_2z_1z_2 + x_2^2z_2^2$$

2). Suppose we want to compute the value of the kernel function $K(x, z)$ from the previous question, on two vectors $x, z \in \mathbb{R}^2$. How many additions and multiplications are needed if you

(a) map the input vector to the feature space and then perform the dot product on the mapped features?

Solution:

To compute this projection, we need 4 multiplications, there exists two points, so need 8 multiplications, and then the dot product itself requires 3 multiplications and 2 additions. So we totally need 11 multiplications and 2 additions.

(b) compute through the kernel function you derived in question 1?

Solution:

$$K(x, z) = x_1^2z_1^2 + 2x_1x_2z_1z_2 + x_2^2z_2^2 = (x_1z_1 + x_2z_2)^2$$

From the above equation, we can easily find that we need 3 multiplications and 1 addition.

2. Kernel functions, Consider the following kernel function:

$$K(x, x') = \begin{cases} 1 & \text{if } x = x' \\ 0 & \text{otherwise} \end{cases}$$

1). Prove this is a legal kernel. That is, describe an implicit mapping: $\Phi : X \rightarrow \mathbb{R}^m$ such that $K(x, x') = \Phi(x) \cdot \Phi(x')$. (You may assume the instance space X is finite.)

Solution:

We get the kernel matrix $K = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$, and K is a positive semidefinite matrix, so this

is a legal kernel.

2). In this kernel space, any labeling of points in X will be linearly separable. Justify this claim.

Solution:

$K(x_i, x_j) = \Phi(x_i)^T \cdot \Phi(x_j) = \begin{cases} 1 & \text{if } x_i = x_j \\ 0 & \text{otherwise} \end{cases}$ Then $\Phi(x_i) = e_i$, e_i is the column vector which the i th element is 1 as $[0, 0, 0, \dots, 1, \dots, 0]^T$

So the decision boundary is $f(x) = W^T \Phi(x) = W^T e_i = W_i = y_i$, all data points will be linearly separable.

3). Since all labelings are linearly separable, this kernel seems perfect for learning any target function. Why is this actually a bad idea?

Solution:

Because W will overfit the data.