

# Machine Learning from Data – IDC

## HW5 – Support Vectors Machine

Theoretical questions:

1. Use Lagrange Multipliers to find the maximum and minimum values of the function subject to the given constraints:
  - a.  $f(x, y) = e^{xy}$ ; constraint:  $2x^2 + y^2 = 72$
  - b.  $f(x, y) = x^2 + y^2$ ; constraint:  $y - \cos 2x = 0$

2.

- a. Consider two kernels  $K_1$  and  $K_2$ , with the mappings  $\varphi_1$  and  $\varphi_2$  respectively. Show that  $K = 7K_1 + 3K_2$  is also a kernel and find its corresponding  $\varphi$ .
- b. Consider a kernel  $K_1$  and its corresponding mapping  $\varphi_1$  that maps from the lower space  $R^n$  to a higher space  $R^m$  ( $m > n$ ). We know that the data in the higher space  $R^m$ , is separable by a linear classifier with the weights vector  $w$ .

Given a different kernel  $K_2$  and its corresponding mapping  $\varphi_2$ , we create a kernel  $K = 7K_1 + 3K_2$  as in section a above. Can you find a linear classifier in the higher space to which  $\varphi$ , the mapping corresponding to the kernel  $K$ , is mapping?

If YES, find the linear classifier weight vector.

If NO, prove why not.

- c. What is the dimension of the mapping function  $\varphi$  that corresponds to a polynomial kernel  $K(x, y) = (\alpha x \cdot y + \beta)^d$ , ( $\alpha, \beta \neq 0$ ), where the lower dimension is  $n$ ?
- d. Consider the space  $S = \{1, 2, \dots, N\}$  for some finite  $N$  (each instance in the space is a 1-dimension vector and the possible values are  $1, 2, \dots, M$ ) and the function  $f(x, y) = \min(x, y)$ .

Find the mapping  $\varphi$  such that:

$$\varphi(x) \cdot \varphi(y) = 5 \min(x, y)$$

For example, if the instances are  $x = 3, y = 5$ , for some  $N \geq 5$ , then:

$$\varphi(x) \cdot \varphi(y) = \varphi(3) \cdot \varphi(5) = 5 \min(3, 5) = 15$$

- e. Consider the space  $S = \{1, 2, \dots, N\}$  for some finite  $N$  (each instance in the space is a 1-dimension vector and the possible values are  $1, 2, \dots, M$ ) and the function  $f(x, y) = \max(x, y)$ .

Prove that the function  $\max(x, y)$  is not a kernel, i.e., there is no mapping  $\varphi$  such that:

$$\varphi(x) \cdot \varphi(y) = \max(x, y)$$

3. Find the Kernel function for the following mapping. Provide a representation of the form  $\alpha K_1 + \beta K_2$ , where both  $K_1$  and  $K_2$  are polynomial kernels and  $\alpha, \beta \geq 0$ :
  - a.  $\varphi(x) = (x_1^3, x_2^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, 2\sqrt{3}x_1^2, 2\sqrt{3}x_2^2, 2\sqrt{6}x_1x_2, 4\sqrt{3}x_1, 4\sqrt{3}x_2, 8)$
  - b.  $\varphi(x) = (\sqrt{10}x_1^2, \sqrt{10}x_2^2, \sqrt{20}x_1x_2, \sqrt{8}x_1, \sqrt{8}x_2, \sqrt{2})$
4. The purpose of this exercise is to demonstrate the usefulness of using kernels. Write a Python script that performs the following:
  - a. Draw 20,000 vectors with 20 dimensions.
  - b. Use the kernel  $(x \cdot y + 1)^2$  to calculate the Gram matrix for these data. That is: each cell  $i, j$  in the matrix is the result of applying the kernel function on the vectors  $i$  and  $j$  ( $i=j$  is also a valid input) :  $K(x_i, x_j)$
  - c. Find the associated mapping function  $\varphi$ . Into which dimension?
  - d. Use  $\varphi$  to map the vectors to the higher dimension.
  - e. Calculate the matrix where each cell  $i, j$  is the result of the dot product of the mapping images of the vectors  $i$  and  $j$ :  $\varphi(x_i) \cdot \varphi(x_j)$
  - f. Compare the matrices from sections b and e (use `np.isclose`) – they should be the same.  
\* NOTE: both matrices should be of the same size 20,000x20,000.
  - g. Compare the time it took your machine to calculate the two matrices. What do you observe?