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: $2x^2 + y^2 = 72$

   Solution:

$$e^{xy} + \lambda(2x^2 + y^2 - 72)$$
$$f'_x = ye^{xy} + 4\lambda x = 0$$
$$f'_y = xe^{xy} + 2\lambda y = 0$$
$$f'_\lambda = 2x^2 + y^2 - 72 = 0$$
$$\frac{4\lambda x}{y} = \frac{2\lambda y}{x}$$
$$2x^2 = y^2$$
$$y^2 + y^2 - 72 = 0$$
$$y^2 = 36 \rightarrow y = \pm 6$$
$$x^2 = 18 \rightarrow x = \pm\sqrt{18}$$
$$max = e^{6\sqrt{18}} \qquad min = e^{-6\sqrt{18}}$$

   b. $f(x,y) = x^2 + y^2$; constraint: $y - \cos 2x = 0$

   Solution:

$$x^2 + y^2 + \lambda(y - \cos 2x)$$
$$f'_x = 2x + 2\lambda \sin 2x = 0$$
$$f'_y = 2y + \lambda = 0$$
$$f'_\lambda = y - \cos 2x = 0$$
$$y = \cos 2x \rightarrow \lambda = -2\cos 2x$$
$$2x - 4\cos 2x \sin 2x = 0$$
$$x = 2\sin 2x \cos 2x$$
$$x = \sin 4x$$

Option 1: $x = 0, y = 1, \lambda = -2 \rightarrow f = 1$ *max*

Option 2: $x = \pm 0.62, y = 0.32, \lambda = -0.64 \rightarrow f = 0.49$ *min*

2.

a. Consider two kernels K_1 and K_2, with the mappings φ_1 and φ_2 respectively. Show that $K = 7K_1 + 3K_2$ is also a kernel and find its corresponding φ.

Solution:

x, y are two vectors in the lower n-dimension.

Lets d_1, d_2 be the higher dimensions respectively.

By definition:

$$K_1(x,y) = \varphi_1(x) \cdot \varphi_1(y) = \sum_{i=1}^{d_1} \varphi_1(x)_i \varphi_1(y)_i$$

$$K_2(x,y) = \varphi_2(x) \cdot \varphi_2(y) = \sum_{i=1}^{d_2} \varphi_2(x)_i \varphi_2(y)_i$$

$$K(x,y) = 7K_1(x,y) + 3K_2(x,y) = 7\sum_{i=1}^{d_1} \varphi_1(x)_i \varphi_1(y)_i + 3\sum_{i=1}^{d_2} \varphi_2(x)_i \varphi_2(y)_i$$

$$= 7\varphi_1(x)_1\varphi_1(y)_1 + \cdots + 7\varphi_1(x)_{d_1}\varphi_1(y)_{d_1} + 3\varphi_2(x)_1\varphi_2(y)_1 + \cdots$$
$$+ 3\varphi_2(x)_{d_2}\varphi_2(y)_{d_2} = \varphi(x) \cdot \varphi(y)$$

Where $\varphi(x) = \left(\sqrt{7}\varphi_1(x)_1, \ldots, \sqrt{7}\varphi_1(x)_{d_1}, \sqrt{3}\varphi_2(x)_1, \ldots, \sqrt{3}\varphi_2(x)_{d_2}\right)$ Q.E.D

b. Consider a kernel K_1 and its corresponding mapping φ_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 φ_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 φ, the mapping corresponding to the kernel K, is mapping?

If YES, find the linear classifier weight vector.

If NO, prove why not.

Solution:

We saw in a that

$$\varphi(x) = \left(\sqrt{7}\varphi_1(x)_1, \ldots, \sqrt{7}\ \varphi_1(x)_{d_1}, \sqrt{3}\varphi_2(x)_{d_1+1}, \ldots, \sqrt{3}\varphi_2(x)_{d_1+d_2}\right)$$
$$= \left[\sqrt{7}\varphi_1(x), \sqrt{3}\varphi_2(x)\right]$$

We also know the linear separator w which separates instances of the two classes in R^m. Given the mapping φ above, notice that R^m is a subspace in the mapping space corresponding to φ. We will now define the linear classifier to be:

$$w' = (w_1, \dots, w_m, 0, \dots, 0)$$

This linear classier gives the same result as w, in terms of the sign, when we calculate the dot product with each instance – therefore it will separate the data just as w did.

c. What is the dimension of the mapping function φ 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?

Solution:

The dimension of the mapping φ is: $\binom{n+d}{n}$ as shown in class.

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, …, N) and the function $f(x, y) = 5\min(x, y)$.

Find the mapping φ 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:

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

Solution:

Given x, y and N the dimension of mapping φ will be N and it will be constructed as follows:

For any number x (=instance) the mapping will be x times $\sqrt{5}$ and then $N - x$ times 0. For example:

If $x = 3, y = 5$ and $N = 6$ the mappings will be:

$$\varphi(x) = \varphi(3) = \sqrt{5}(1,1,1,0,0,0)$$

$$\varphi(y) = \varphi(5) = \sqrt{5}(1,1,1,1,1,0)$$

$$\varphi(x) \cdot \varphi(y) = \varphi(3) \cdot \varphi(5) = 5 + 5 + 5 + 0 + 0 + 0 = 15 = 5\min(3,5)$$

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, …, N) 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 φ such that:

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

Solution:

Kernel function should create a psd (positive semi definite) matrix from all the $\binom{n+1}{2}$ possible choices of 2 instances from n with replacement.

Let's assume we have only 2 instances in the data $\{1, 2\}$.

The function $\max(x, y)$ will create the following matrix:

$$\begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$$

The determinant of this matrix is 2-4=-2

A psd matrix has a determinant equal or bigger than 0.

Therefor, $\max(x, y)$ is not a kernel.

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. $\varphi(x) = (x_1^3, x_2^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, 2\sqrt{3}x_1^2, 2\sqrt{3}x_2^2, 2\sqrt{6}x_1x_2, 4\sqrt{3}x_1, 4\sqrt{3}x_2, 8)$

<u>*Solution*</u>:

$$\varphi(x) \cdot \varphi(y) = x_1^3y_1^3 + x_2^3y_2^3 + 3x_1^2x_2y_1^2y_2 + 3x_1x_2^2y_1y_2^2 + 12x_1^2y_1^2 + 12x_2^2y_2^2$$
$$+ 24x_1x_2y_1y_2 + 48x_1y_1 + 48x_2y_2 + 64$$
$$= (x \cdot y)^3 + 12(x \cdot y)^2 + 48x \cdot y + 64 = (x \cdot y + 4)^3 = K(x, y)$$

b. $\varphi(x) = (\sqrt{10}x_1^2, \sqrt{10}x_2^2, \sqrt{20}x_1x_2, \sqrt{8}x_1, \sqrt{8}x_2, \sqrt{2})$

<u>*Solution*</u>:

$$\varphi(x) \cdot \varphi(y) = 10x_1^2y_1^2 + 10x_2^2y_2^2 + 20x_1x_2y_1y_2 + 8x_1y_1 + 8x_2y_2 + 2$$
$$= 10(x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2) + 8x_1y_1 + 8x_2y_2 + 2$$
$$= 10(x_1y_1 + x_2y_2)^2 + 8x_1y_1 + 8x_2y_2 + 2$$
$$= (x_1y_1 + x_2y_2)^2 + 2x_1y_1 + 2x_2y_2 + 1 + 9(x_1y_1 + x_2y_2)^2 + 6x_1y_1 + 6x_2y_2 + 1$$
$$= (x \cdot y)^2 + 2x \cdot y + 1 + 9(x \cdot y)^2 + 6x \cdot y + 1$$
$$(x \cdot y + 1)^2 + (3x \cdot y + 1)^2 = K_1(x, y) + K_2(x, y) = K(x, y)$$

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 φ. Into which dimension?

d. Use φ 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?