Machine Learning from Data – IDC
HW5 – Support VectorsMachine

Theoretical questions:

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

$L(x,y) = e^{xy} + \lambda(2x^2 + y^2 - 72) = 0$

$\nabla x \rightarrow L(x,y) = ye^{xy} + 4\lambda x = 0 \rightarrow ye^{xy} = -4\lambda x$

$\nabla y \rightarrow L(x,y) = xe^{xy} + 2\lambda y = 0 \rightarrow xe^{xy} = -2\lambda y$

נחלק את המשוואה הראשונה בשנייה ונקבל:

$$\frac{ye^{xy}}{xe^{xy}} = \frac{-4\lambda x}{-2\lambda y} \rightarrow \frac{y}{x} = \frac{2x}{y} \rightarrow y^2 = 2x^2$$

$\nabla \lambda \rightarrow L(x,y) = 2x^2 + y^2 - 72 = 0$

אז נציב במשוואה השלישית ונקבל:

$$y^2 + y^2 - 72 = 0 \rightarrow 2y^2 = 72 \rightarrow y^2 = 36 \rightarrow y = \pm 6$$

$$y = 6 \rightarrow 36 = 2x^2 \rightarrow x = \pm\sqrt{18}$$

לכן:

$$f(x,y) = e^{xy}$$
$$f(6,\sqrt{18}) = e^{6\sqrt{18}}$$
$$f(6,-\sqrt{18}) = e^{-6\sqrt{18}}$$
$$f(-6,\sqrt{18}) = e^{-6\sqrt{18}}$$
$$f(-6,-\sqrt{18}) = e^{6\sqrt{18}}$$

max = $e^{6\sqrt{18}}$ **, min** =$e^{-6\sqrt{18}}$

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

$$L(x,y) = x^2 + y^2 + \lambda(y - cos2x) = 0$$

$\nabla x \rightarrow L(x,y) = 2x + 2\lambda sin2x = 0 \rightarrow x = -\lambda sin2x$

$\nabla y \rightarrow L(x,y) = 2y + \lambda = 0 \rightarrow 2y = -\lambda$

$\nabla\lambda \rightarrow L(x,y) = y - cos2x = 0 \rightarrow y = cos2x$

נחלק את המשוואה הראשונה בשנייה ונקבל:

$\frac{x}{2y} = -\frac{\lambda sin2x}{-\lambda} \rightarrow \frac{x}{2y} = sin2x$

נציב את המשוואה שלישית בזאת ונקבל:

$$\frac{x}{2cos2x} = sin2x \rightarrow x = 2sin2xcos2x \rightarrow 2x = 4sin2xcos2 \rightarrow 2x = sin4x \rightarrow$$
$$x = 0,0.619,-0.619$$
$$x = 0 \rightarrow y = cos2x \rightarrow y = cos0 = 1$$
$$x = 0.619 \rightarrow y = \cos(2 * 0.619) = cos1.238 = 0.326$$
$$x = -0.619 \rightarrow y = \cos(2 * -0.619) = cos - 1.238 = 0.326$$
$$f(x,y) = x^2 + y^2$$
$$f(0,1) = 0 + 1 = 1$$
$$f(0.619,0.326) = (0.619)^2 + (0.326)^2 = 0.383 + 0.106 = 0.489$$
$$f(-0.619,0.326) = (0.619)^2 + (0.326)^2 = 0.383 + 0.106 = 0.489$$
$$\text{min} = 0.489, \quad \text{max} = 1$$

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.

יהיו x,y וקטורים במרחב כלשהו, לפני המיפוי ויהיו d_1, d_2 מימדים ממימד כלשהו לאחר המיפוי אז לפי ההגדרה של קרנל:

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

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

ולכן:

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

$$7(\varphi_1(x)_1 \cdot \varphi_1(y)_1 + \varphi_1(x)_2 \cdot \varphi_1(y)_2 + \cdots + \varphi_1(x)_n \cdot \varphi_1(y)_n)$$
$$+ 3(\varphi_2(x)_1 \cdot \varphi_2(y)_1 + \varphi_2(x)_2 \cdot \varphi_2(y)_2 + \cdots + \varphi_2(x)_m \cdot \varphi_2(y)_m)$$
$$= \varphi_3(x) \cdot \varphi_3(y)$$

$$\varphi_3(x) = \sqrt{7}\varphi_1(x)_1, \ldots, \sqrt{7}\varphi_1(x)_n, \sqrt{3}\varphi_2(x)_1, \ldots, \sqrt{3}\varphi_2(x)_m$$

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.

בפתיחה בסעיף a ראינו ש:

$$\varphi(x) = (\varphi_1(x)_1, \varphi_1(x)_2, .. \varphi_1(x)_{d_1}, \varphi_2(x)_{d1+1}, \varphi_2(x)_{d1+2} \ldots \varphi_2(x)_{d_2})$$

בנוסף, w הוא מפריד לינארי במימד m, עבור קרנל k1. נוסיף משקולות בגודל 0 לפי הצורך כלומר נקבל-$w' = (w_1, w_2 \ldots w_m, 0,0 \ldots 00)$, אנו יודעים שהוא מסווג את המידע נכון עבור k1.

ולכן לפי הסעיף הקודם ולפי ההרחבה של w נקבל את אותן התוצאות (מיפויים) כלומר, w' הוא מפריד לינארי בדיוק כמו w.

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?

נפתח בעזרת $x = (x_1, x_2, \dots, x_n)$, $y = (y_1, y_2 \dots y_n)$ יש n כאלה כי המימד שממנו מגיעים הוא n.

$$K(x, y) = (\alpha(x_1, x_2, \dots, x_n)(y_1, y_2 \dots y_n) + \beta)^d$$

$$(\alpha x_1 y_1, \alpha x_2 y_2, \dots, \alpha x_n y_n + \beta)^d =$$

כעת יש n+1 מחוברים ולכן, לפי מה שלמדנו בכיתה לפי התאוריה מולטינומית

$$\dim(\varphi) = \binom{d + n + 1 - 1}{n + 1 - 1} = \binom{d + n}{n}$$

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

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:

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

נגדיר את המיפוי הבא:

עבור כל מספר $1 \leq i \leq N$ נגדיר שהמספר יהיה 5i פעמים 1 והשאר 0, כלומר עבור N=7 ועבור x=3 נקבל (5,5,5,0,0,0,0) ועבור y=5 נקבל (5,5,5,5,5,0,0).

N יהיה למעשה המימד של φ.

נניח ש x = min(x,y) ונסמן

$$n = \sum_{i=1}^{x} 1.1 + \sum_{i=x+1}^{y} 0.1 + \sum_{i=y+1}^{N} 0.0$$

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

$$5 \min(3,5) = 5.3 = 15$$

e. Consider the space $S = \{1,2, \ldots N\}$ for some finite N (each instance in the space is a 1-dimension vector and the possible values are $1, 2, \ldots, N$) 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:

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

על מנת להוכיח שפונקציה זו היא איננה קרנל, נראה דוגמא לכך שקיים וקטור שעבורו במרחב סופי מסוים המיפוי הוא לא בהחלט חיובי.

נגדיר: s=[1,2]

נסתכל על המטריצה:

$$M = \begin{pmatrix} 2 & 4 \\ 3 & 3 \end{pmatrix}$$

ועל הוקטור:

$$v = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

לכן:

$$< v, Mv > = < \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 & 4 \\ 3 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} >=< \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \end{pmatrix} > = 1 * -2 + (-1) * 0 = \ -2$$

ולכן איננה קרנל.

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 $\alpha, \beta > 0$:

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

$$\varphi(x) \cdot \ \varphi(y) = \left(x_1^3y_1^3, +x_2^3y_2^3 + \ \sqrt{3}x_1^2x_2\sqrt{3}y_1^2y_2 + \sqrt{3}x_2^2x_1\sqrt{3}y_2^2y_1 + \ 2\sqrt{3}x_1^2\ 2\sqrt{3}y_1^2 + \ 2\sqrt{3}x_2^2 2\sqrt{3}y_2^2 + 2\sqrt{6}x_1x_2\ 2\sqrt{6}y_1y_2 + 4\sqrt{3}x_1 4\sqrt{3}y_1 + \ 4\sqrt{3}x_2 4\sqrt{3}y_2 + 8 \cdot 8 \right) =$$

$$= y_1^3 + x_2^3y_2^3 + 3x_1^2x_2y_1^2y_2 + 3x_1x_2^2y_1y_2^2 + 12x_1^2y_1^2 + 12x_2^2y_2^2 + 24x_1x_2y_1y_2 + 48x_1y_1 + 48x_2y_2 + 64$$

$$= (x \cdot y)^3 + 12(x \cdot y)^2 + 48x \cdot y + 64 = \ (x \cdot y + 4)^3 = K1$$

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

$$\varphi(x)\cdot\ \varphi(y) = \sqrt{10}x_1^2\sqrt{10}y_1^2 + \sqrt{10}x_2^2\sqrt{10}y_2^2 + \sqrt{20}x_1x_2\sqrt{20}y_1y_2 + \sqrt{8}x_1\sqrt{8}y_1 + \sqrt{8}x_2\sqrt{8}y_2 + \sqrt{2}\sqrt{2} =$$

$$= 10x_1^2y_1^2 + 10x_2^2y_2^2 + 20x_1x_2y_1y_2 + 8x_1y_1 + 8x_2y_2 + 2$$

$$= \ 10(x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2) + 8(x_1y_1 + x_2y_2 + \frac{1}{4}) =$$

$$10(xy)^2 + 8\left(xy + \frac{1}{4}\right) = 10K_1 + 8K_2$$

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 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?