Machine Learning from Data – IDC
HW5 – Support Vectors Machine

Nitai Aharoni -203626742

Theoretical questions:

1. Use Lagrange Multipliers to find the maximum and minimum values of the function subject to the given constraints:
   a. $f(x,y) = 2e^{xy}$; constraint: $2x^2 + y^2 = 32$

- נמצא נקודות מינימום ומקסימום עבור f תחת האילוץ הנתון
- נסמן – $g(x,y) = 2x^2 + y^2 - 32$
- נסמן את מערכת לגראנז׳ להיות -

$$L(x,y,\lambda) = f(x,y) - \lambda g(x,y) = 2e^{xy} - \lambda(2x^2 + y^2 - 32)$$

- נשתמש בכופלי לגראנז׳ ונחשב את $-\nabla L = \left(\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \frac{\partial L}{\partial \lambda}\right)$
  - $\frac{\partial L}{\partial x} = 2e^{xy} * y - 4\lambda x$
  - $\frac{\partial L}{\partial y} = 2e^{xy} * x - 2\lambda y$
  - $\frac{\partial L}{\partial \lambda} = -2x^2 - y^2 + 32$

- $\begin{cases} 2e^{xy} * y - 4\lambda x = 0 \\ 2e^{xy} * x - 2\lambda y = 0 \\ 2x^2 + y^2 - 32 = 0 \end{cases}$
- $\begin{cases} 2e^{xy} * y^2 - 4\lambda xy = 0 \\ 4e^{xy} * x^2 - 4\lambda xy = 0 \\ 2x^2 + y^2 - 32 = 0 \end{cases}$
- $\begin{cases} 2e^{xy} * y^2 - 4e^{xy} * x^2 = 0 \\ 4e^{xy} * x^2 + 4\lambda xy = 0 \\ 2x^2 + y^2 - 32 = 0 \end{cases}$
- $\begin{cases} y^2 - 2x^2 = 0 \\ 4e^{xy} * x^2 + 4\lambda xy = 0 \\ 2x^2 + y^2 - 32 = 0 \end{cases}$
- $\begin{cases} y^2 = 2x^2 \\ 4e^{xy} * x^2 + 4\lambda xy = 0 \\ y^2 + y^2 - 32 = 0 \end{cases}$

- $\begin{cases} y^2 = 2x^2 \\ 4e^{xy} * x^2 + 4\lambda xy = 0 \\ y^2 = 16 \end{cases}$
- $\begin{cases} x = \pm\sqrt{8} \\ y = \pm 4 \end{cases}$
- לכן נקודות הקיצון הן - $(\sqrt{8}, 4), (\sqrt{8}, 4), (-\sqrt{8}, -4), (-\sqrt{8}, 4)$
  - $f(x, y) = 2e^{xy} = 2e^{\sqrt{8}*4} = 163874.41$ - $(\sqrt{8}, 4)$ מקסימום
  - $f(x, y) = 2e^{-\sqrt{8}*4} = 0.00002$ - $(-\sqrt{8}, 4)$ מינימום
  - $f(x, y) = 2e^{\sqrt{8}*-4} = 0.00002$ - $(\sqrt{8}, -4)$ מינימום
  - $f(x, y) = 2e^{-\sqrt{8}*-4} = 163874.41$ - $(-\sqrt{8}, -4)$ מקסימום

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

- נמצא נקודות מינימום ומקסימום עבור f תחת האילוץ הנתון
- נסמן – $g(x, y) = y - \cos x$
- נסמן את מערכת לגראנז׳ להיות -

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y) = x^2 + y^2 - \lambda(y - \cos x)$$

- נשתמש בכופלי לגראנז׳ ונחשב את $\nabla L = \left(\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \frac{\partial L}{\partial \lambda}\right)$-
  - $\frac{\partial L}{\partial x} = 2x - \lambda \sin x$
  - $\frac{\partial L}{\partial y} = 2y - \lambda$
  - $\frac{\partial L}{\partial z} = -y + \cos x$
- $\begin{cases} x = \frac{\lambda \sin x}{2} \\ y = \frac{\lambda}{2} \\ \cos x = y \end{cases}$
- $\begin{cases} x = y \sin x \\ y = \frac{\lambda}{2} \\ \cos x = y \end{cases}$
- $\begin{cases} x = \cos x \sin x \\ y = \frac{\lambda}{2} \\ \cos x = y \end{cases}$

- $\begin{cases} 2x = 2\cos x \sin x \\ y = \frac{\lambda}{2} \\ \cos x = y \end{cases}$
- $\begin{cases} 2x = \sin 2x \\ y = \frac{\lambda}{2} \\ \cos x = y \end{cases}$
- $\begin{cases} t = \sin t \;\rightarrow\; t = 2x = 0 \\ y = \cos x \end{cases}$
- $\begin{cases} x = 0 \\ y = 1 \end{cases}$
- לכן נקודת הקיצון היא – $(0,1)$ –
  - בדיקת מקסימום/מינימום:
  - $x = -\frac{pi}{2}$, $y = \cos x = 0$: $f\left(-\frac{pi}{2}, 0\right) = \frac{pi^2}{4} = 2.467$
  - $x = 0$, $y = 1$: $f(0,1) = 1$
  - $x = \frac{pi}{2}$, $y = \cos x = 0$: $f\left(\frac{pi}{2}, 0\right) = \frac{pi^2}{4} = 2.467$
- לכן הנקודה היא נקודת מינימום

2.

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

- יהיו: $K_1: Y \times Y \rightarrow R$, $K_2: Y \times Y \rightarrow R$ (Y מימד הinstances) פונקציות קרנל הממפות כך:
  $\varphi_1: Y \rightarrow F_1$, $\varphi_2: Y \rightarrow F_2$ בהתאמה (F מימד ה-attributes).
- ידוע ש- $K_1(x,y) = \varphi_1(x) \cdot \varphi_1(y)$, ו- $K_2(x,y) = \varphi_2(x) \cdot \varphi_2(y)$
- נסמן: $K = K_1 + K_2$ ויהיו $x, y \in Y$ אזי:

$$K(x,y) = K_1(x,y) + K_2(x,y) = \varphi_1(x) \cdot \varphi_1(y) + \varphi_2(x) \cdot \varphi_2(y) = (\varphi_1(x), \varphi_2(x)) \cdot (\varphi_1(y) \cdot \varphi_2(y))$$

- נסמן $\varphi(x) = (\varphi_1(x), \varphi_2(x))$.
- לכן $K(x,y) = (\varphi_1(x), \varphi_2(x)) \cdot (\varphi_1(y) \cdot \varphi_2(y)) = \varphi(x) \cdot \varphi(y)$
- כלומר K הינה פונקציית קרנל ופונקציית המיפוי היא: $\varphi(x) = (\varphi_1(x), \varphi_2(x))$

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 $K = K_1 + K_2$ 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.

נכון.

- נגדיר פונקציית קרנל $K_1: R^n X R^n \to R$ הממפה כך: $\varphi_1: R^n \to R^m$ כך ש- $m > n$
- נגדיר פונקציית קרנל $K_2: R^n X R^n \to R$ הממפה כך: $\varphi_2: R^n \to R^{m'}$ כך ש- $m' > n$
- נגדיר $K = K1 + K2$ פונקציית קרנל בעלת מיפוי φ
- לכן מסעיף a ידוע ש- $\varphi(x) = (\varphi_1(x), \varphi_2(x))$
- לכן φ ממפה אל-$R^{m+m'}$
- יהי $w' \in R^{m+m'}$, כך ש-w' וקטור ב-$R^{m+m'}$ ובו m הערכים הראשונים שווים לערכי w, והערכים בו מ-m עד m' שווים ל-0.
- נסמן $x \in R^n$ להיות instance מה-data
- נשים לב ש- $w' \cdot \varphi(x) = (w, 0) \cdot (\varphi_1(x), \varphi_2(x)) = w \cdot \varphi_1(x)$
- ולכן היות ונתון שניתן להפריד את משקולות w בעזרת מפריד לינארי עבור כל Instance, אזי w' הינו מפריד לינארי במרחב הגבוה $R^{m+m'}$

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?

- נתונה פונקציית קרנל $K: R^n X R^n \to R$ כך ש-$K(x, y) = (\alpha x \cdot y + \beta)^d$, $(\alpha, \beta \neq 0)$, המימד הנמוך הוא n
- אזי לפי נוסחאות המולטינום:

$$K(x, y) = (\alpha x \cdot y + \beta)^d = (\alpha x^T * y + \beta)^d = \left(\alpha \sum_{i=1}^{n} x_i y_i + \beta\right)^d = \left(\alpha \sum_{i=1}^{n} x_i y_i + \left(\sqrt{\beta}\right)^2\right)^d =$$

$$= \sum_{k_1 + \cdots + k_{n+1} = d} \binom{d}{k_1, \dots, k_{n+1}} \prod_{j=1}^{n} (\alpha x_j y_j)^{k_j} * \left(\beta^{\frac{k_{n+1}}{2}}\right)^2 =$$

$$= \sum_{k_1 + \cdots + k_{n+1} = d} \sqrt{\binom{d}{k_1, \dots, k_{n+1}}} \prod_{j=1}^{n} \left(\alpha^{\frac{k_j}{2}}\right) x_j^{k_j}) \left(\beta^{\frac{k_{n+1}}{2}}\right) \sqrt{\binom{d}{k_1, \dots, k_{n+1}}} \prod_{j=1}^{n} \left(\alpha^{\frac{k_j}{2}}\right) y_j^{k_j}) \left(\beta^{\frac{k_{n+1}}{2}}\right) =$$

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

עבור – $\varphi(x) = \left(\sqrt{\binom{d}{k_1,\dots,k_{n+1}}}\left(\prod_{j=1}^{n}\left(\alpha^{\frac{k_j}{2}}\right)x_j^{k_j}\right)\left(\beta^{\frac{k_{n+1}}{2}}\right)\right)_{k_1+\cdots+k_{n+1}=d}$

- נחשב את מספר הסכומים במשוואה שקיבלנו שבעצם מהווים את המימד שהמיפוי ממפה אליו - $= \binom{n+1+d-1}{n+1-1}$ $\binom{n+d}{n}$
- לכן המרחב של המיפוי φ המתאים ל-K הוא $\binom{n+d}{n}$

d. Given two polynomial kernels
$K_1(x,y) = (\alpha_1 x\cdot y + \beta_1)^d$ and $K_2(x,y) = (\alpha_2 x\cdot y + \beta_2)^d$ (note that d is the same in both kernels), with the corresponding mappings φ_1 and φ_2. Find a mapping φ that corresponds to the kernel $K = K_1 + K_2$ and that has the same dimension as φ_1 and φ_2.

- נתונות פונקציות קרנל: $K_1, K_2 \colon Y\,X\,Y \to R$ כך ש-$K_1(x,y) = (\alpha_1 x\cdot y + \beta_1)^d$ ו- $K_2(x,y) = (\alpha_2 x\cdot y + \beta_2)^d$, בעלות מיפויים φ_1, φ_2 בהתאמה.
- תהי פונקציית קרנל $K = K_1 + K_2$ עם המיפוי φ
- יהיו $x, y \in Y$

לכן לפי סעיף c:

$$K(x,y) = K_1(x,y) + K_2(x,y) =$$

$$= \sum_{k_1+\cdots+k_{n+1}=d}\binom{d}{k_1,\dots,k_{n+1}}\prod_{j=1}^{n}(\alpha_1 x_j y_j)^{k_j}\beta_1^{k_{n+1}}$$

$$+ \sum_{k_1+\cdots+k_{n+1}=d}\binom{d}{k_1,\dots,k_{n+1}}\prod_{j=1}^{n}(\alpha_2 x_j y_j)^{k_j}\beta_2^{k_{n+1}} = \cdots$$ אלגברה כמו בסעיף c =

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

- נשים לב שמספר האיברים ב-φ נקבע לפי מספר האיברים בביטוי המולטינום, ב-n+1 האיברים ובחזקת-d - $\binom{n+d}{n}$
- לפי סעיף c ידוע ש- φ_1, φ_2 הם מיפויים המתאימים לקרנל פולינומי ולכן גם הם ממפים למרחב $\binom{n+d}{n}$
- לכן φ הוא מיפוי המתאים ל- K הממפה באותו אופן למרחבים כמו φ_1, φ_2

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

Find the mapping φ such that:

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

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

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

- נתון $S = \{1,2, \ldots N\}$, $f(x, y) = \min(x, y)$, $x, y \in S$
- נגדיר מיפוי $\varphi: S \to R^N$
- נסמן - $\varphi(x) = (\overbrace{1, \ldots, 1}^{x}, \overbrace{0 \ldots, 0}^{N-x})$
- לכן

$$\varphi(x) \cdot \varphi(y) = \left(\overbrace{1, \ldots, 1}^{x}, \overbrace{0 \ldots, 0}^{N-x}\right)\left(\overbrace{1, \ldots, 1}^{y}, \overbrace{0 \ldots, 0}^{N-y}\right) = \overbrace{1+\ldots+1}^{\text{min(x,y) פעמים}} + \overbrace{0+\ldots+0}^{\text{N-min(x,y) פעמים}} = \min(x, y) =$$

$$= f(x, y)$$

- כלומר $\varphi(x) \cdot \varphi(y) = f(x, y)$
- לכן $f(x,y)$ היא פונקציית קרנל עבור המיפוי φ

3. Find the Kernel function for the following mapping:

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

- נמצא פונקציית קרנל $K: R^2XR^2 \to R$ כך שעבור כל $x, y \in R^2$ מתקיים $K(x, y) = \varphi(x) \cdot \varphi(y)$
- נחשב:

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

$$= x_1^3y_1^3 + x_2^3y_2^3 + 3x_1^2x_2y_1^2y_2 + 3x_1x_2^2y_1y_2^2 + 9x_1^2y_1^2 + 9x_2^2y_2^2 + 18x_1x_2y_1y_2$$

$$+ 27x_1y_1 + 27x_2y_2 + 27 =$$

$$= (x_1y_1 + x_2y_2)^3 + (x_1y_1 + x_2y_2)^2 + 27(x_1y_1 + x_2y_2) + 27 =$$

$$= (x \cdot y^T)^3 + (x \cdot y^T)^2 + 27(x \cdot y^T) + 27 = (x \cdot y^T + 3)^3$$

b. $\varphi(x) = (\sqrt{5}x_1^2, \sqrt{5}x_2^2, \sqrt{10}x_1x_2, \sqrt{8}x_1, \sqrt{8}x_2, \sqrt{5})$

- נמצא פונקציית קרנל $K: R^2XR^2 \to R$ כך שעבור כל $x, y \in R^2$ מתקיים $K(x, y) = \varphi(x) \cdot \varphi(y)$

- נחשב:

$$K(x,y) = \varphi(\mathrm{x}) \cdot \varphi(\mathrm{y}) = 5\mathrm{x}_1^2\mathrm{y}_1^2 + 5\mathrm{x}_2^2\mathrm{y}_2^2 + 10\mathrm{x}_1\mathrm{x}_2\mathrm{y}_1\mathrm{y}_2 + 8\mathrm{x}_1\mathrm{y}_1 + 8\mathrm{x}_2\mathrm{y}_2 + 25$$
$$= 5(\mathrm{x}_1\mathrm{y}_1 + \mathrm{x}_2\mathrm{y}_2)^2 + 8(\mathrm{x}_1\mathrm{y}_1 + \mathrm{x}_2\mathrm{y}_2) + 5 = 5(\mathrm{x} \cdot y^T)^2 + 8(\mathrm{x} \cdot y^T) + 5$$