HW5 – Theory + SVM

1. <u>PAC Learning and VC dimension (30 pts)</u>

   Let $X = \mathbb{R}^2$. Let

   $$C = H = \left\{ h(r_1, r_2) = \left\{ (x_1, x_2) \middle| \begin{matrix} x_1^2 + x_2^2 \geq r_1 \\ x_1^2 + x_2^2 \leq r_2 \end{matrix} \right\} \right\}, \text{ for } 0 \leq r_1 \leq r_2,$$

   the set of all origin-centered rings.

   a. (8 pts) What is the $VC(H)$? Prove your answer.
   b. (14 pts) Describe a polynomial sample complexity algorithm L that learns C using H. State the time complexity and the sample complexity of your suggested algorithm. Prove all your steps.

   In class we saw a bound on the sample complexity when H is finite.

   $$m \geq \frac{1}{\varepsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right)$$

   When $|H|$ is infinite, we have a different bound:

   $$m \geq \frac{1}{\varepsilon}\left(4\log_2\frac{2}{\delta} + 8VC(H)\log_2\frac{13}{\varepsilon}\right)$$

   c. (8 pts) You want to get with 95% confidence a hypothesis with at most 5% error. Calculate the sample complexity with the bound that you found in b and the above bound for infinite $|H|$. In which one did you get a smaller m? Explain.

2. <u>VC dimension (20 pts)</u>

   Let $X = \mathbb{R}$ and $n \in \mathbb{N}$.

   Define "x-node decision tree" for any $x = 2^n - 1$ to be a full binary decision tree with x nodes (including the leaves).

   Let H_m be the hypothesis space of all "x-node decision tree" with $n \leq m$.

   a. (5 pts) What is the $VC(H_3)$? Prove your answer.
   b. (15 pts) What is the $VC(H_m)$? Prove your answer.

3. <u>Kernels and mapping functions (25 pts)</u>

   a. (20 pts) Let $K(x, y) = (x \cdot y + 1)^3$ be a function over $\mathbb{R}^2 \times \mathbb{R}^2$ (i.e., $x, y \in \mathbb{R}^2$).

      Find ψ for which $K$is a kernel. (It may help to first expand the above term on the right-hand side).

   b. (2 pts) What did we call the function ψ in class if we remove all coefficients?

c. (3 pts) How many multiplication operations do we save by using $K(x, y)$ versus $\psi(x) \cdot \psi(y)$?

4. <u>Lagrange multipliers (15 pts)</u>

Let $f(x, y) = 2x - y$. Find the minimum and the maximum points for f under the constraint $g(x, y) = \frac{x^2}{4} + y^2 = 1$.

5. See notebook exercise (10 pts)

1. <u>PAC Learning and VC dimension (30 pts)</u>

Let $X = \mathbb{R}^2$. Let

$$C = H = \left\{h(r_1, r_2) = \left\{(x_1, x_2) \middle| \begin{matrix} x_1^2 + x_2^2 \geq r_1 \\ x_1^2 + x_2^2 \leq r_2 \end{matrix}\right\}\right\}, \text{ for } 0 \leq r_1 \leq r_2,$$

the set of all origin-centered rings.

a. (8 pts) What is the $VC(H)$? Prove your answer.

b. (14 pts) Describe a polynomial sample complexity algorithm L that learns C using H. State the time complexity and the sample complexity of your suggested algorithm. Prove all your steps.

In class we saw a bound on the sample complexity when H is finite.

$$m \geq \frac{1}{\varepsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right)$$

When $|H|$ is infinite, we have a different bound:

$$m \geq \frac{1}{\varepsilon}\left(4\log_2\frac{2}{\delta} + 8VC(H)\log_2\frac{13}{\varepsilon}\right)$$

c. (8 pts) You want to get with 95% confidence a hypothesis with at most 5% error. Calculate the sample complexity with the bound that you found in b and the above bound for infinite $|H|$. In which one did you get a smaller m? Explain.

(א) נוכיח כי $VC(H) = 2$ בכך שנוכיח ש-$VC(H) \geq 2$ ונוכיח ש-$VC(H) < 3$

(*) $VC(H) \geq 2$:

נראה שקיימת תת קבוצה $S \subseteq X$ כך ש-$|S| = 2$ וכן ש-H מנתצת אותה:

תהי: $S = \{(1,2), (2,4)\}$. ונסמן $x_1 = (1,2)$, $x_2 = (2,4)$

נשים לב כי $1^2 + 2^2 = 5$, $2^2 + 4^2 = 20$. אזי:

א. עבור ההשמות $c(x_1) = 1$ $c(x_2) = 0$ נשים לב ש: $h(2,3) = \left\{(k_1, k_2) : \begin{matrix} k_1^2 + k_2^2 \geq 4 \\ k_1^2 + k_2^2 \leq 9 \end{matrix}\right\}$

לכן $x_1 \in h(2,3)$ ואילו $x_2 \notin h(2,3)$ ולכן לכל i $h(x_i) = c(x_i)$.

ב. עבור הרשמות $c(x_1)=1$ ו $c(x_2)=1$ נשים לב ש: $h(2,5)=\left\{(k_1,k_2): \begin{matrix} k_1^2+k_2^2\ge 4 \\ k_1^2+k_2^2\le 25\end{matrix}\right\}$

ולכן $x_1\in h(2,5)$, $x_2\in h(2,5)$ ולכן לכל i $h(x_i)=c(x_i)$

ג. עבור הרשמות $c(x_1)=0$ ו $c(x_2)=1$ נשים לב ש: $h(4,5)=\left\{(k_1,k_2): \begin{matrix} k_1^2+k_2^2\ge 16 \\ k_1^2+k_2^2\le 25\end{matrix}\right\}$

ולכן $x_1\notin h(4,5)$, $x_2\in h(4,5)$ ולכן לכל i $h(x_i)=c(x_i)$

ד. עבור הרשמות $c(x_1)=0$ ו $c(x_2)=0$ נשים לב ש: $h(1,2)=\left\{(k_1,k_2): \begin{matrix} k_1^2+k_2^2\ge 1 \\ k_1^2+k_2^2\le 4\end{matrix}\right\}$

ולכן $x_1\notin h(1,2)$, $x_2\notin h(1,2)$ ולכן לכל i $h(x_i)=c(x_i)$

אכן הראינו שקיימת תת-קבוצה $S\subseteq X$, $|S|=2$ כך שH מנתצת את S.

(*) $VC(H)<3$:

תהי $S=\{k_1,k_2,k_3\}$ כך ש $S\subseteq X$ ו $|S|=3$ נראה שH לא מנתצת אותה:

א. נניח שk_1 וk_2 הן בעלות מרחק זהה מראשית הצירים (בלי הגבלת הכלליות) ויהי מרחק זה r.

אזי, עבור ההשמה $c(k_1)=1$, $c(k_2)=0$, $c(k_3)=1$ נתייחס לכל ההיפותזות שעבורן $h(k_1)=c(k_1)=1$ ונסמנם ב H'.

נשים לב כי לכל $\epsilon_1,\epsilon_2\in R_+$: $H'=\left\{h(r-\epsilon_1,r+\epsilon_2)=\left\{(x_1,x_2): \begin{matrix} x_1^2+x_2^2\ge(r-\epsilon_1)^2 \\ x_1^2+x_2^2\le(r+\epsilon_2)^2\end{matrix}\right\}\right\}$

תהי $h'\in H'$. אזי מהיות $\|k_1\|=\|k_2\|=r$, לכן $h'(k_1)=h'(k_2)=1$

ולכן $1=h'(k_2)\ne c(k_2)=0$

ב. נניח שהנקודות בעלות מרחק שונה מראשית הצירים ותהי k_1 הנקודה במרחק הקצר ביותר, d_1 וk_3 הנקודה במרחק הארוך ביותר, d_3 (בלי הגבלת הכלליות)

אזי, עבור ההשמה $c(k_1)=1$, $c(k_2)=0$, $c(k_3)=1$

ועבור $k_1=(k_{11},k_{12})$, $k_2=(k_{21},k_{22})$, $k_3=(k_{31},k_{32})$ נניח שקיימת $h(r_1,r_2)\in H$ כך ש:

$k_1,k_3\in h$, $k_2\notin h$ אך מהיות k_3 רחוקה יותר מהצירים מk_2, מתקיים כי:

$\underset{k_1,k_3\in h}{\underset{\downarrow}{r_1}}\le k_{11}^2+k_{12}^2\le k_{21}^2+k_{22}^2\le k_{31}^2+k_{32}^2\le \underset{k_1,k_3\in h}{\underset{\downarrow}{r_2}}$ ולכן גם $k_2\in h$ בסתירה לכך ש $h(k_2)=0$

(ט) נמצא $h \in H$ כך ש h עקבית עם איזשהו $c \in C$ לכל sample מ D

האלגוריתם:

א. נעבור על כל הsamples ב D ונתייחס רק ל $d \in D$, כך ש $d \in c$

ב. מבין כל הsamples השייכים לc נבחר את d בעל המרחק הקצר ביותר מהראשית ונגדיר את r_1 להיות המרחק הנ"ל.

ג. מבין כל הsamples השייכים לc נבחר את d בעל המרחק הארוך ביותר מהראשית ונגדיר את r_2 להיות המרחק הנ"ל

ד. לבסוף, נשים $L(D)=h(r_1,r_2)=\left\{(x_1,x_2): \begin{matrix} x_1^2+x_2^2 \geq r_1^2 \\ x_1^2+x_2^2 \leq r_2^2 \end{matrix}\right\}$

נכונות:

נשים לב שמצעדים ב׳+ג׳ קיבלנו ש:

$$r_1 = \min_{d \in D \cap c}\{\text{מרחק של } d \text{ מהראשית}\}$$
$$r_2 = \max_{d \in D \cap c}\{\text{מרחק של } d \text{ מהראשית}\}$$

ולכן, עבור כל $d \in D \cap c$ נקבל גם כי $d \in h(r_1,r_2)$

בנוסף, מהנתון $C=H$ ולכן עבור כל $d \in D$ כך ש $d \notin c$ נקבל כי $d \notin h(r_1,r_2)$

לכן $h(r_1,r_2)$ עקבית עם c.

זמן ריצה:

$O(m)$ - אנחנו עוברים על כל הsamples ב D מספר קבוע של פעמים (אפשר גם בפעם אחת) ולכן זמן הריצה שלנו הוא $O(m)$

Sample complexity: נרצה לדעת כמה דגימות צריך לכל הפחות בשביל $\delta, \epsilon > 0$

יהי: $C = \left\{(x_1,x_2): \begin{matrix} x_1^2+x_2^2 \geq r_1^2 \\ x_1^2+x_2^2 \leq r_2^2 \end{matrix}\right\}$. נגדיר: $B_1 = \left\{(x_1,x_2): \begin{matrix} x_1^2+x_2^2 \geq (r_1')^2 \\ x_1^2+x_2^2 \leq r_2^2 \end{matrix}\right\}$, $B_2 = \left\{(x_1,x_2): \begin{matrix} x_1^2+x_2^2 \geq r_1^2 \\ x_1^2+x_2^2 \leq (r_2')^2 \end{matrix}\right\}$

כך ש $r_1 \leq r_1' \leq r_2' \leq r_2$ וכן $\Pi(B_i)=\frac{\epsilon}{2}$ לכל $i \in \{1,2\}$

נשים לב שכדי שההיפותזה שלנו תהיה ϵ-bad צריך להתקיים שה Data שלנו לא מבקר או ב B_1 או ב B_2 כלומר:

$$\Pi^m(D \in X^m : Err(L(D),c) > \epsilon) \leq$$
$$\leq (\Pi(X-B_1))^m + (\Pi(X-B_2))^m \leq \left(1-\frac{\epsilon}{2}\right)^m + \left(1-\frac{\epsilon}{2}\right)^m \leq 2 \cdot e^{-\frac{m\epsilon}{2}} \leq \delta \longrightarrow \ln\left(e^{-\frac{m\epsilon}{2}}\right) \leq \ln\left(\frac{\delta}{2}\right)$$
$$\longrightarrow \frac{-m \cdot \epsilon}{2} \leq \ln\left(\frac{\delta}{2}\right) \longrightarrow m \geq -\frac{2}{\epsilon}\ln\left(\frac{\delta}{2}\right) \longrightarrow m \geq \frac{2}{\epsilon}\ln\left(\frac{2}{\delta}\right)$$

כלומר, נצטרך לפחות $\frac{2}{\epsilon}\ln\left(\frac{2}{\delta}\right)$ כדי שבוודאות של $1-\delta$ ה Err יהיה לכל היותר ϵ

(c) לפי הנתונים:

$$\epsilon = 0.05 \quad , \quad 1-\delta = 0.95 \Rightarrow \delta = 0.05$$

ולכן לפי סעיף ב נצטרך לפחות: $m \geq \frac{2}{0.05} \cdot \ln\left(\frac{2}{0.05}\right) = 147.5$

כלומר, לפחות 148 דגימות.

בנוסף, מסעיף א $VC(H)=2$ ולכן מהנוסחה שלמדנו בהרצאה: $m(\varepsilon,\delta) \geq \frac{1}{\varepsilon}\left(4\log_2\left(\frac{2}{\delta}\right) + 8VC(H)\log_2\left(\frac{13}{\varepsilon}\right)\right)$

$$m \geq \frac{1}{0.05}\left(4\log\frac{2}{0.05} + 8\cdot 2\cdot \log\left(\frac{13}{0.05}\right)\right) = 2992.9$$

כלומר, לפחות 2993 דגימות

הסיבה שקיבלנו חסם הדוק יותר בראשון היא כי מרחב ההיפותזות יכול להשפיע הרבה יותר מאשר התפלגות.

2. <u>VC dimension (20 pts)</u>

Let $X = \mathbb{R}$ and $n \in \mathbb{N}$.

Define "x-node decision tree" for any $x = 2^n - 1$ to be a full binary decision tree with x nodes (including the leaves).

Let H_m be the hypothesis space of all "x-node decision tree" with $n \le m$.

a. (5 pts) What is the $VC(H_3)$? Prove your answer.

b. (15 pts) What is the $VC(H_m)$? Prove your answer.

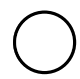
(סע) נוכיח כי $VC(H_3)=4$ בכך שנוכיח ש$VC(H_3)\ge 4$ ונוכיח ש $VC(H_3)<5$

(*) $4\le VC(H_3)$:

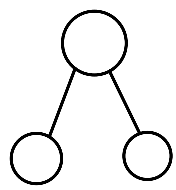
נראה שקיימת תת קבוצה $S\subseteq X$ כך ש-$|S|=4$ וכן H_3 מנתצת אותה:

ראשית, נשים לב כי ה X-desicion-tree היחידים שיש לנו הם מהצורות הבאות (מהיות $m=3$):

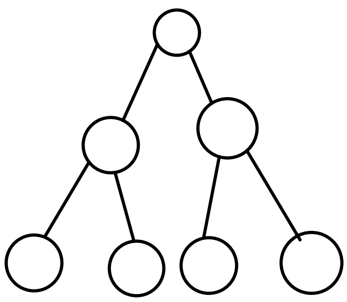
$n=1$



$n=2$:



$n=3$:



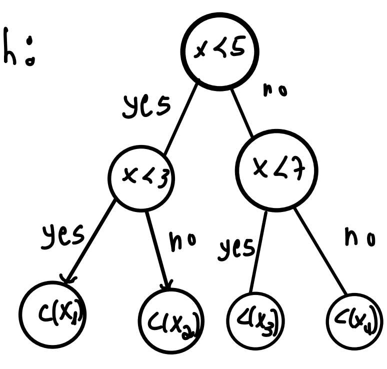
נוכיח עבור desicion tree כאשר $n=3$:

תהי $S=\{2,4,6,8\}$ ונסמן $x_i=2i$ לכל $1\le i\le 4$.

נחלק את S לשתי קבוצות S_+ ו-S_- ונסמן $c(x_i)=+$ לכל $x_i\in S_+$ ו- $c(x_i)=-1$ לכל $x_i\in S_-$

ונבנה בעץ החלטה h של H_3:

h:



נשים לב שלכל $1\le i\le 4$: $h(x_i)=c(x_i)$ ולכן h עקבית עם c

נוכיח כעת ש-$VC(H_3)<5$:

תהי $S=\{x_1,x_2,x_3,x_4,x_5\}$ כאשר $x_1,x_2,x_3,x_4,x_5\in\mathbb{R}$ בסדר עולה.

ותהי ההשמה הבאה:

לכל $i=2k$: $c(x_i)=+$

לכל $i=2k+1$: $c(x_i)=-$

נתבונן ב-decision tree כאשר $h=3$:

נשים לב כי כל מסלול מהשורש לעלה מגדיר לנו אינטרוול ולכן יש לנו 4 אינטרוולים. מאחר שיש לנו 5 נקודות, מעקרון שובך היונים בהכרח יהיו לנו 2 נקודות שונות באותו אינטרוול ולכן מתקיים כי עבור i מסויים

$h(x_i)=h(x_{i+1})=+$ או $h(x_i)=h(x_{i+1})=-$

בסתירה להשמה.

באותו אופן ניתן להוכיח עבור decision tree כאשר $h=2$ או $h=1$ באמצעות עקרון שובך היונים

(b) נוכיח כי $VC(H_m)=2^{m-1}$ בכך שנוכיח ש-$VC(H_m)\ge 2^{m-1}$ ונוכיח ש-$VC(H_m)<2^m$

יהי $m\in\mathbb{N}^+$. כאשית, נשים לב כי ב-x-node decision trees מספר העלים הוא $\frac{x+1}{2}$ ולכן מהגדרה, H_m הוא מרחב ההיפותזות של כל עצי ההחלטה המלאים עם 2^{m-1} עלים לכל היותר (עבור $x=2^m-1$)

(*) $2^{m-1}\le VC(H_m)$:

נראה שקיימת תת קבוצה $S\subseteq X$ כך ש-$|S|=2^{m-1}$ וכך ש-H_m מנתצת אותה:

תהי $S=\{2,4,6,8,\dots,2^m\}$ ונסמן $x_i=2i$ לכל $1\le i\le 2^{m-1}$.

נחלק את S לשתי קבוצות S_+ ו-S_- ונסמן $c(x_i)=+$ לכל $x_i\in S_+$

$c(x_i)=-1$ לכל S_-.

טוי: עבור עץ ההחלטה h שמתנהג באופן הבא מפורש למטה:

א. נגדיר $S'=S$

ב. נשווה את העץ שנכנס 'x' לאיבר האמצעי בגודלו ב-'S (נסמן אותו ב-'x)

ג. אם 'x קטן ממנו, נלך לתת עץ השמאלי ונגדיר $S'=\{x_i\in S' : x_i<x'\}$

ד. אם 'x' גדול ממנו נלך לתת העץ הימני ונגדיר $S'=\{x_i\in S' : x'<x_i\}$

ה. נחזור לשלב א'.

כלומר, ההשוואה מתבצעת כל פעם מול החציון של אותו חלק בקבוצה אם קיבלנו שx קטן מהחציון, נלך שמאלה, אחרת נלך ימינה (בדומה למה שעשינו בסעיף א').

נשים לב שלכל $1\le i\le 2^m-1$: $h(x_i)=c(x_i)$ ולכן h עקבית עם c

(*) $\underline{VC(H_m)<2^{m-1}+1}$:

תהי $S=\{x_1,x_2,x_3,\dots,x_{2^{m-1}+1}\}$ כאשר $x_1,x_2,x_3,\dots,x_{2^{m-1}+1}\in\mathbb{R}$ בסדר עולה.

ותהי ההשמה הבאה:

לכל $i=2k$: $c(x_i)=+$

לכל $i=2k+1$: $c(x_i)=-$

ונתייחס לdesicion tree h x-node עם 2^{m-1} עלים (המקסימום עלים האפשרי בH_m):

נשים לב כי כל מסלול מהשורש לעלה מגדיר לנו אינטרוול ולכן יש לנו 2^{m-1} אינטרוולים. מאחר שיש לנו $2^{m-1}+1$ נקודות, מעקרון שובך היונים בהכרח יהיו לנו 2 נקודות שונות באותו אינטרוול ולכן מתקיים כי עבור i מסוים:

$$h(x_i)=h(x_{i+1})=+ \quad \text{או} \quad h(x_i)=h(x_{i+1})=-$$

בסתירה להשמה.

באותו אופן ניתן להוכיח עבור desicion tree עם פחות מ-2^{m-1} עלים באמצעות עקרון שובך היונים

3. <u>Kernels and mapping functions (25 pts)</u>

a. (20 pts) Let $K(x,y) = (x \cdot y + 1)^3$ be a function over $\mathbb{R}^2 \times \mathbb{R}^2$ (i.e., $x, y \in \mathbb{R}^2$).

Find ψ for which $K$is a kernel. (It may help to first expand the above term on the right-hand side).

b. (2 pts) What did we call the function ψ in class if we remove all coefficients?

c. (3 pts) How many multiplication operations do we save by using $K(x,y)$ versus $\psi(x) \cdot \psi(y)$?

(a). ראשית, נשים לב כי:

$$K(x,y) = (x \cdot y + 1)^3 = (x_1 \cdot y_1 + x_2 \cdot y_2 + 1)^3 =$$

$$= (x_1^2 \cdot y_1^2 + x_1 \cdot x_2 \cdot y_1 \cdot y_2 + x_1 y_1 + x_1 \cdot x_2 \cdot y_1 \cdot y_2 + x_2^2 \cdot y_2^2 + x_2 \cdot y_2 + x_1 \cdot y_1 + x_2 \cdot y_2 + 1) \cdot (x_1 \cdot y_1 + x_2 \cdot y_2 + 1) =$$

$$= (x_1^2 \cdot y_1^2 + x_2^2 \cdot y_2^2 + 2 x_1 x_2 y_1 y_2 + 2 x_1 \cdot y_1 + 2 x_2 \cdot y_2 + 1)(x_1 \cdot y_1 + x_2 \cdot y_2 + 1) =$$

$$= x_1^3 \cdot y_1^3 + x_1 x_2^2 \cdot y_1 y_2^2 + 2 x_1^2 x_2 y_1^2 y_2 + 2 \cdot x_1^2 \cdot y_1^2 + 2 \cdot x_1 \cdot x_2 \cdot y_1 \cdot y_2 + x_1 y_1 + x_1^2 \cdot x_2 \cdot y_1^2 \cdot y_2 + x_2^3 \cdot y_2^3 + 2 x_1 \cdot x_2^2 \cdot y_1 \cdot y_2^2 + 2 x_1 \cdot x_2 \cdot y_1 \cdot y_2 +$$

$$+ 2 \cdot x_2^2 \cdot y_2^2 + x_2 \cdot y_2 + x_1^2 \cdot y_1^2 + x_2^2 \cdot y_2^2 + 2 x_1 x_2 y_1 y_2 + 2 x_1 \cdot y_1 + 2 x_2 \cdot y_2 + 1 =$$

$$= x_1^3 y_1^3 + x_2^3 \cdot y_2^3 + 3 x_1 \cdot x_2^2 \cdot y_1 \cdot y_2^2 + 3 x_1^2 \cdot x_2 \cdot y_1^2 \cdot y_2 + 6 x_1 \cdot x_2 \cdot y_1 \cdot y_2 + 3 x_1^2 \cdot y_1^2 + 3 x_2^2 \cdot y_2^2 + 3 x_1 y_1 + 3 x_2 y_2 + 1$$

נשים לב כי עבור פונקציית המיפוי $\psi : R^2 \longrightarrow R^{10}$:

$$\psi(x_1, x_2) = (x_1^3, x_2^3, \sqrt{3} x_1 \cdot x_2^2, \sqrt{3} \cdot x_1^2 \cdot x_2, \sqrt{6}\, x_1 \cdot x_2, \sqrt{3} \cdot x_1^2, \sqrt{3}\, x_2^2, \sqrt{3} \cdot x_1, \sqrt{3} \cdot x_2, 1) \in R^{10}$$

ונקבל ש $\forall x, y \in R^2$:

$$\psi(x) = (x_1^3, x_2^3, \sqrt{3} x_1 \cdot x_2^2, \sqrt{3} \cdot x_1^2 \cdot x_2, \sqrt{6}\, x_1 \cdot x_2, \sqrt{3} \cdot x_1^2, \sqrt{3}\, x_2^2, \sqrt{3} \cdot x_1, \sqrt{3} \cdot x_2, 1)$$

$$\psi(y) = (y_1^3, y_2^3, \sqrt{3} y_1 \cdot y_2^2, \sqrt{3} \cdot y_1^2 \cdot y_2, \sqrt{6}\, y_1 \cdot y_2, \sqrt{3} \cdot y_1^2, \sqrt{3}\, y_2^2, \sqrt{3} \cdot y_1, \sqrt{3} \cdot y_2, 1)$$

ובנוסף:

$$\psi(x) \cdot \psi(y) = (x_1^3, x_2^3, \sqrt{3} x_1 \cdot x_2^2, \sqrt{3} \cdot x_1^2 \cdot x_2, \sqrt{6}\, x_1 \cdot x_2, \sqrt{3} \cdot x_1^2, \sqrt{3}\, x_2^2, \sqrt{3} \cdot x_1, \sqrt{3} \cdot x_2, 1)(y_1^3, y_2^3, \sqrt{3} y_1 \cdot y_2^2, \sqrt{3} \cdot y_1^2 \cdot y_2, \sqrt{6}\, y_1 \cdot y_2, \sqrt{3} \cdot y_1^2, \sqrt{3}\, y_2^2, \sqrt{3} \cdot y_1, \sqrt{3} \cdot y_2, 1)$$

$$= x_1^3 y_1^3 + x_2^3 \cdot y_2^3 + 3 x_1 x_2^2 \cdot y_1 \cdot y_2^2 + 3 x_1^2 \cdot x_2 \cdot y_1^2 \cdot y_2 + 6 x_1 x_2 \cdot y_1 \cdot y_2 + 3 x_1^2 y_1^2 + 3 x_2^2 \cdot y_2^2 + 3 x_1 y_1 + 3 x_2 y_2 + 1 = K(x,y)$$

ולכן מהיות כך שψ מקיימת את כך שלכל $x, y \in R^2$: $\psi(x) \cdot \psi(y) = K(x,y)$

לכן ψ היא פונקציית המיפוי שעבורה K הוא Kernel.

(ב). בהרצאה 6 עמוד 42, קראנו לפונקציה מסוג זה בלי המקדמים: full rational varieties

ונשים לב שהפונקציה אכן דומה לפונקציה מעמוד 49 (עמוד לפני) ללא המקדמים:

$$(x_1,x_2)\longmapsto(1,x_1,x_2,x_1x_2,x_1^2,x_2^2,x_1^2x_2,x_1x_2^2,x_1^3,x_2^3)$$

אחת התכונות של פונקציה מסוג זה היא שהמימד שאליו הפונקציה ממפה הוא $\binom{n+r}{r}$ כאשר r זה הדרגה של כל איבר ו n זה המימד של הוקטור לפני המיפוי.

נשים לב כי בפועל זה: $r=3$, $n=2$

ולכן: $\binom{r+n}{r}=\binom{3+2}{3}=\binom{5}{3}=10$

(c) נשים לב שמספר הפעולות שאנחנו מבצעים ב $K(x,y)=(x\cdot y+1)^3$ הוא:

א. 2 פעולות כפל ופעולת חיבור במכפלה הפנימית $x\cdot y$ - 3 פעולות סה"כ

ב. פעולת חיבור בהוספת 1: $x\cdot y+1$ - סה"כ פעולה אחת

ג. מהפעולות א+ב קיבלנו סקלר ולכן העלאה בחזקת 3 שקולה ל 2 פעולות הכפלה - סה"כ 2 פעולות

לכן, מספר הפעולות ב $K(x,y)$ הוא 6 פעולות

לעומת זאת, מספר הפעולות שאנחנו מבצעים ב $\psi(x)\cdot\psi(y)$ הוא:

א. החישוב של $\psi(x)$ לוקח לנו 18 פעולות הכפלה (הסבר למטה *)

ב. החישוב של $\psi(y)$ לוקח לנו 18 פעולות הכפלה (הסבר למטה *)

ג. המכפלה הפנימית בין $\psi(x)$ ל $\psi(y)$ לוקחת לנו 10 פעולות הכפלה (בין קואורדינטות) ו 9 פעולות חיבור (חיבור המכפלות) - סה"כ 19 פעולות

לכן מספר הפעולות ב $\psi(x)\cdot\psi(y)$ הוא 55 פעולות

$$\left(\overset{2}{\uparrow}\!\!\!\!\!\!\;x_1^3,\ \overset{2}{\uparrow}\!\!\!\!\!\!\;x_2^3,\ \overset{3}{\uparrow}\sqrt{3}x_1\cdot x_2^2,\ \overset{3}{\uparrow}\sqrt{3}\cdot x_1^2\cdot x_2,\ \overset{2}{\uparrow}\sqrt{6}\,x_1\cdot x_2,\ \overset{2}{\uparrow}\sqrt{3}\cdot x_1^2,\ \overset{2}{\uparrow}\sqrt{3}\,x_2^2,\ \overset{1}{\uparrow}\sqrt{3}\cdot x_1,\ \overset{1}{\uparrow}\sqrt{3}\cdot x_2,\ \overset{0}{\uparrow}1\right)\ *$$

לכן חסכנו $55-6=49$ (כלומר 49 פעולות) בשימוש ב $K(x,y)$ במקום ב $\psi(x)\cdot\psi(y)$

4. Lagrange multipliers (15 pts)

Let $f(x,y) = 2x - y$. Find the minimum and the maximum points for f under the constraint $g(x,y) = \frac{x^2}{4} + y^2 = 1$.

נמצא מקסימום של f באמצעות Lagrange multipliers:

$$L(x,y,\lambda) = 2x - y + \lambda\left(\frac{x^2}{4} + y^2 - 1\right)$$

א. $\frac{\partial}{\partial x} L(x,y,\lambda) = 2 + \frac{\lambda \cdot x}{2} = 0$

ב. $\frac{\partial}{\partial y} L(x,y,\lambda) = -1 + 2\lambda \cdot y = 0$

ג. $\frac{\partial}{\partial \lambda} L(x,y,\lambda) = \frac{x^2}{4} + y^2 - 1 = 0$

נשים לב של א: $2 + \frac{\lambda x}{2} = 0 \longrightarrow \frac{\lambda x}{2} = -2 \longrightarrow \lambda = \frac{-4}{x}$

ושל ב: $-1 + 2\lambda \cdot y = 0 \longrightarrow 2\lambda \cdot y = 1 \longrightarrow \lambda = \frac{1}{2y}$

ולכן קיבלנו ש: $\frac{-4}{x} = \frac{1}{2y} \longrightarrow x = -8y$

נציב בג ונקבל: $\frac{(-8y)^2}{4} + y^2 - 1 = 0 \longrightarrow \frac{64y^2}{4} + y^2 - 1 = 0 \longrightarrow 16y^2 + y^2 = 1 \longrightarrow y^2 = \frac{1}{17} \longrightarrow y = \pm\frac{1}{\sqrt{17}}$

ולכן $x = \mp\frac{8}{\sqrt{17}}$

נציב בא ונקבל:

עבור $x = -\frac{8}{\sqrt{17}}$: $2 - \frac{8}{2\sqrt{17}} \cdot \lambda = 0 \longrightarrow 2 = \frac{4}{\sqrt{17}}\lambda \longrightarrow \lambda = \frac{\sqrt{17}}{2}$

עבור $x = \frac{8}{\sqrt{17}}$: $2 + \frac{8}{2\sqrt{17}}\lambda = 0 \longrightarrow \frac{4}{\sqrt{17}}\lambda = -2 \longrightarrow \lambda = -\frac{\sqrt{17}}{2}$

ולכן קיבלנו 2 פתרונות:

1. $\lambda = \frac{\sqrt{17}}{2}$, $x = -\frac{8}{\sqrt{17}}$, $y = \frac{1}{\sqrt{17}}$

2. $\lambda = -\frac{\sqrt{17}}{2}$, $x = \frac{8}{\sqrt{17}}$, $y = -\frac{1}{\sqrt{17}}$

נבדוק איזה אחד מהם הוא המקסימום ואיזה הוא המינימום ע"י הצבה בf:

(*) נציב את 1 בf: $f\left(-\frac{8}{\sqrt{17}}, \frac{1}{\sqrt{17}}\right) = 2 \cdot \left(-\frac{8}{\sqrt{17}}\right) - \frac{1}{\sqrt{17}} = -\frac{17}{\sqrt{17}} = -\sqrt{17}$

(*) נציב את 2 בf: $f\left(\frac{8}{\sqrt{17}}, -\frac{1}{\sqrt{17}}\right) = 2 \cdot \left(\frac{8}{\sqrt{17}}\right) - \left(-\frac{1}{\sqrt{17}}\right) = \frac{17}{\sqrt{17}} = \sqrt{17}$

ולכן, עבור $\lambda = \frac{\sqrt{17}}{2}$, $x = -\frac{8}{\sqrt{17}}$, $y = \frac{1}{\sqrt{17}}$ נקבל את נקודת המינימום של f תחת האילוץ g

ועבור $\lambda = -\frac{\sqrt{17}}{2}$, $x = \frac{8}{\sqrt{17}}$, $y = -\frac{1}{\sqrt{17}}$ נקבל את נקודת המקסימום של f תחת האילוץ g.