הרצאה (12)

אז שוב, עוסקים בבעיית ה- **k-line Center**. נתונות n נק' במישור (P) כלומר, למצוא k פסים בעובי מינימלי, שמכסים את P.
זו בעיה קשה..
נניח שקבענו את העובי (1) ונרצה לכסות את P עם כמה שפחות פסים בעובי 1.

אז אם אפשר עם k, אנחנו נמצא עם $O(k\log k)$. כלומר אם k הוא המינימלי, נוכל לכסות עם $O(k\log k)$ פסים.

גם אם לא ממש נותנים לנו את k, אפשר לנחש k ואם לא הולך, להכפיל אותו פי 2 ולעשות מין חיפוש בינארי.

הרעיון הוא להסתכל על פסים קנוניים (S): כל אחד נוגע ב-2 נק', ויש $O(n^2)$ כאלו. משקל פס כזה מתחיל מ-1. הנה שוב הלוגיקה יותר מסודר.
נגדיר $|S|$.

repeat

בוחרים מדגם אקראי R של $ck\log k$ פסים, כל פס נבחר בהסתב' פרופ' למשקלו.

אם, הם מכסים את P, עוצרים.

אחרת, נקח נק' $q\in P$ לא מכוסה ונכפיל את המשקל של כל הפסים שמכילים את q בתנאי שמשקלם הכולל של כל הפסים שמכילים את q $\le \frac{W(\text{הפסים})}{2k}$

until $8k\log n$ איטרציות מכפילות

output false (ואם לא הצלחנו)

אנליזה:

בכל הכפלה, המשקל הכולל גדל פי $\le 1+\frac{1}{2k}$.

$$W'(S) \le W(S)\left(1+\frac{1}{2k}\right)$$

המשקל בסיום:

$$\ldots \le \underset{|S|}{W_0(S)}\left(1+\frac{1}{2k}\right)^{8k\log n} \le |S|e^{4\log n} \sim n^6$$

(עם כולם)
נניח שיש כיסוי עם k פסים $\sigma_1,\sigma_2,...,\sigma_k$.
באיטרציה מכפילה, הנק' p מכוסה לפחות ע"י 1 מה-σ_i האלו.
לכן, המשקל של לפחות אחד מהם מוכפל.
היות ועושים $8k\log n$ איטרציות שבכל אחת, לפחות אחד מהם מכפיל את משקלו.
⇐ לפחות אחד מהם מוכפל $8\log n$ פעמים, ולכן משקלו לפחות $2^{8\log n}$ שזה n^8.

האם צריך תיקון לקבועים

$$n^2 e^{4\log n} = n^2\cdot 2^{4\log e\log n} = n^{4\log e+2} \overset{?}{<} n^8$$

$e=2^{\log e}$

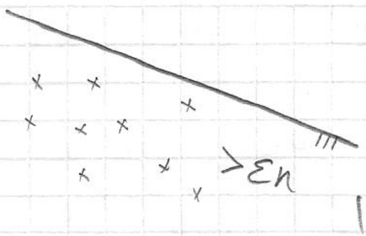
אז בעצם, צריך לטעון שרוב האיטרציות הן באמת מכפילות. כלומר, או שמצאנו, או שהעובדה שנכשלנו והכיסוי נכשל, זה בגלל שבהסתב' גבוהה אין כיסוי.
בשביל זה צריך קצת תיאוריה. לשם כך, צריך את הרעיון והנושא של ε-nets. בצורה יותר כללית:

(*) נתונה קב' A של אובייקטים (נניח נק') ונתון אוסף Σ (בגודל m) של תתי קבוצות מיוחדות של A, ונתון פרמטר $0<\varepsilon<1$

"קבוצות מיוחדות"
למשל, דוגמה קלאסית, נתונות n נק' במישור, והתת-קבוצות הן σ מה שניתן לחתוך מהקב' ע"י מישור מפריד, או למשל ע"י מלבן וכו'.

המטרה היא לבחור תת קבוצה (באופן אקראי) $N\subseteq A$ קטנה עם התכונה: כל קב' $\sigma\in\Sigma$, שגדלה לפחות εn, מכילה נקודה של N.

N נקראת ε-net.



באינטואיציה, אם נדגום נק' מ-A באופן אקראי בהסתברות p כ"א, ואם $p\sim\frac{1}{\varepsilon n}$ היינו מצפים שבתוחלת נדגום נק' אחת. כלומר אם יש לנו קב' מגודל εn, ודוגמים מתוך ה-n המקורית בהסתב' $\frac{1}{\varepsilon n}$, היינו מצפים לדגום נק' אחת מהקב'.

ולכן לא מספיק טוב
אבל זה קצת חלק מדי. כלומר בתוחלת זה יהיה 1, ~~[illegible]~~ ואנחנו
רוצים שבהסתב' גבוהה תבחר נקודה.
מה עושים? בוחרים קצת יותר נק'.

בוחרים כל נק' בהסתב' $p=\frac{c}{\varepsilon n}\log n$ (במקום $\frac{1}{\varepsilon n}$).

עבור חצי מישור אחד ספציפי שמכיל לפחות εn נק', ההסתב' שלא לבחור אף אחת מהן, היא $(1-p)^{\varepsilon n}$

$$\approx \left(1-\frac{c}{\varepsilon n}\log n\right)^{\varepsilon n} = e^{-c\log n} = \frac{1}{n^c}$$

ככל שנגדיל את c, ההסתב' שלא נבחר אף אחת הולכת וקטנה.

אבל רגע, כמה תת קב' יש? המון! אבל כמה תת הקבוצות המוגדרות ע"י חצאי מישורים יש? רק $O(n^2)$! כל אחד מוגדר ע"י 2 נק'.
לכן הסיכוי שהדגימה תכשל, אפילו בחצי מישור אחד היא לכל היותר $\frac{1}{n^{c-2}}$ (לפי ה- union bound).

כלומר אם נדגום למשל $\frac{4}{\varepsilon n}\log n$, נדגום נק' מכל חצי מישור כבד.

אם למשל בוחרים תת קבוצות עם מעגלי יחידה, (כאילו גם יש קטנים, שנוגעים בנק'...) $O(n^2)$ כי ניקח מעגל, ונזיז עד שיגע ב-2 נק'. למעשה זה מגדיר 4 תת' קב', כזו שלא מכילה את a ו-b, כזו שמכילה רק את a, כזו שמכילה רק את b, וכזו שמכילה את שניהם, כי כל פעם מזיזים את המעגל קצת בהתאם.

גם במצב לא כללי, כי יש הרבה נק' על השפה

עבור מעגלים כלליים - $O(n^3)$.
עבור מלבנים מקבילים לצירים - $O(n^4)$.

כל עוד מספר הקב' הוא פולינומיאלי, ואנחנו מסודרים, ומספיק גודל מדגם של $O(\frac{1}{\varepsilon}\log n)$.

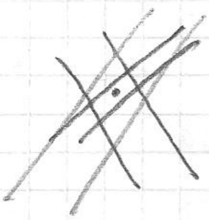
התוצאה הקשה יותר טכנית קשה ומעניינת היא שאפשר
להסתפק במדגם שגודלו $\boxed{O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})}$ (בכלל) לא תלוי ב-n !!

התנאי לכך הוא שמספר תתי הקבוצות יהיה $O(n^d)$, ושזה יהיה
נכון גם אם נחליף את A בכל קבוצה כלשהי בגודל m, $O(m^d)$.

ח״ק מציין שזה לא יתן הסתב' גבוהה אלא הסתברות קבועה.
ומיכה עונה שזה סבבה אם ε מספיק קטן או מספיק גדול
או מספיק מה שצריך.

למה זה משדרג לנו את הח״ק? כי אפשר לקחת קבוצות קטנות
"ומייצגת". אם למשל ניקח מדגם שמכיל 10% מהקב' הקטנה, אז
ע״י מעגל גשש
הוא כנראה יכיל גם בערך 10% מהקב' המקורית.
יש לזה ∞ שימושים גיאומטריים מגניבים וכו'.

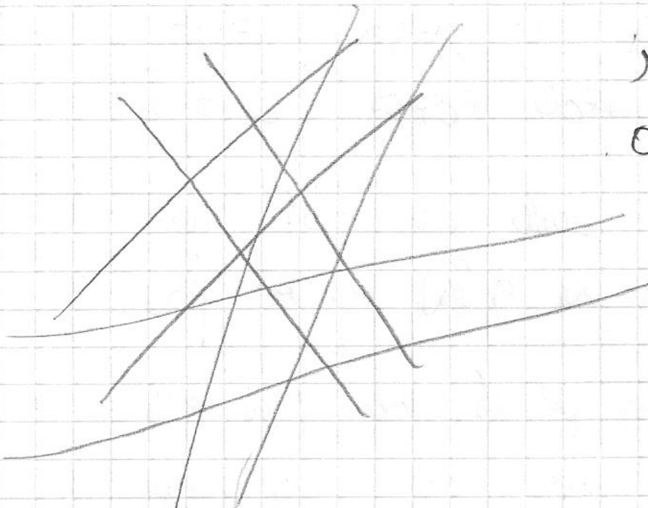
חזרה לבעיה שלנו, באלג' A = קב' הפסים עם משקלים. נעמיד פנים
שפס במשקל w הוא בעצם w פסים.



תתי קב': כל הפסים שנדקרים ע״י נק'.
דגמנו תת קב' N של פסים, $|N| = ck\log k$.
כלומר, ה-ε הוא בערך $\frac{1}{k}$.
והרי אנו יודעים שכל קבוצת פסים שמכילה נק' מסוימת ויש
בה יותר מ-$\frac{W}{k}$ פסים, תכיל פס מהמדגם (N).

במילים אחרות, אם מצאנו נק' שלא מוכלת באף פס מהמדגם,
סימן שמספר הפסים (המשוכפלים, כלומר משקל הפסים) שמכיל אותה הוא
קטן מ-$\frac{W}{k}$. היינו רצינו $\frac{W}{2k}$ אז גם קודם צריך להכפיל את 2k... אוף
ההסתב' להצלחה היא $\frac{1}{2}$, בעצם כל איטרציה "תצליח", כי כאמור
ההסתב' היא קבועה...

הדבר היחיד שצריך לדרוש בעצם, הוא שכמות הפסים שניתן לדגום
ע״י נק' היא פולינומיאלית. אבל כאן זה אפילו יותר פשוט



יש W פסים וצמת התא
קבוצת הנקודות היא $O(|W|^2)$.
כלומר $O(|W|^2)$ אפשרויות לדקור
ולשלול את קבוצה של פסים.

בעצם, מה שעשינו הוא תהליך שאפשר להכליל לבעיות דומות...
יש לנו אלג' קירוב לבעיית Hitting Set בסיטואציות גיאומטריות דומות.

יש לנו קב' נקודות P ואוסף של תת' קב' מיוחדת. מטרתנו לבחור את קב' $H \subseteq P$ מינימלית שדוקרת את כל הקבוצות.
(בניגוד לכסות את הקבוצות הנתונות)

המינוח המוכלל והמדויק הוא עם כל מיני היפר גרפים והיפר קשתות, אבל לא ניכנס לזה...
למשל אם יש מעגלים ורוצים לדקור אותם. (המטרה היא למצוא מינ' נק' שידקרו כל מעגל)
למצוא בעיה מינימלית היא NP-קשה, והמטרה היא לקרב מבחינת הגודל של הקב'. פורמלית:

סיטואציה דואלית
Hitting
↕
Cover

המטרה: לקרב מבחינת גודל H, $k=|H_{opt}|$. נמצא קב' דוקרת בגודל $ck\log k$ בדיוק באותו אופן.

ומיכה מוציא את זה כתרגיל. ממשקלים את הנק' של P. הנכונות, הפולינומיות והכל בדיוק אותו דבר. כל דבר הוא $\frac{1}{k}$-net (או $\frac{1}{2k}$-net) בהסתב' גבוהה בגלל משפט ה-ε-net, והגרסה הממושקלת תפיל משקל קטן של נק' ומביא לתוצאה הנכונה.

פקטור הקירוב הוא $O(\log opt.)=O(\log k)$.

באופן מאוד כללי אפשר למצוא פתרון עם יחס קירוב $O(\log n)$, ע"י אלג' חמדן:
איך? נמצא נק' שדוקרת מקס' אוסף/מעגלים שעוד לא נדקרו. נזרוק את המעגלים ונמשיך באותו אופן.

אם כן, בעיית הכיסוי (Cover) ובעיית הדקירה (Hitting)
נפתרות באותו אופן, כי הן ד' דואליות אחת לשניה, באותו אלג' שראינו.

וכעת, למשהו אחר לגמרי.
התחלנו לדבר על אלגוריתמי קירוב, וכעת נראה כמה בעיות
קלאסיות עם פתרונות מקורבים. זה קצת יקשר למה שראינו.
נתחיל מבעיות חימום, שנפתח לשיטה כללית וממדים גבוהים בהמשך.

כמה בעיות קלאסיות עם פתרונות מקורבים

חישוב קוטר ב-3D. בעיה נחמדה בפני עצמה.
נתונה קב' P של n נק' ב-$\mathbb{R}^3$.
רוצים לחשב קוטר של P = $D(P)=\max_{x,y\in P} d(x,y)$

במישור, כבר ראינו שקל לפתור את הבעיה ע"י חישוב קמור. יש מועמדים לנק' נגדיות וכו'. ב-3 ממדים, זה כבר לא פשוט. ב-$O(n^2)$ אפשר לפתור באופן טריוויאלי.

נסתכל על הבעיה כבעיית אופטימיזציה, עם פרמטר הכרעה וכל זה.
מקבלים מרחק δ, וצריך לקבוע האם $D(P)\overset{?}{\lesseqgtr}\delta$.

לכל $p\in P$, נסתכל על הכדור $B_\delta(p)$ = מרכז ב-p עם רדיוס δ.

אם הקוטר קטן מ-δ, כלומר $D(P)\le\delta$, אז $P\subseteq B_\delta(p)$ לכל $p\in P$.
אז הקב' צריכה להיות בחיתוך של הכדורים, (וגם אם"ם):

$$P\subseteq K_\delta=\bigcap_{p\in P} B_\delta(p)$$

נניח לרגע שזה לא נכון, נניח ש-P לא מוכל ב-K_δ. אז נסמן ב-P_0 את $P\cap K_\delta$, וטוען שהקוטר מושג בין זוג נק' ב-$P\setminus P_0$. כל זה די ברור. אפשר להעלים מהנק' שכן בחיתוך. כלומר:
$D(P)>\delta$, $d(x,y)=D(P)>\delta$, $y\notin B_\delta(x), x\notin B_\delta(y) \Rightarrow x,y\notin K_\delta$

אז הנה מנגנון שמאפשר לזרוק הרבה נק'...

נמיין את נק' P ; $p_1,p_2,\dots,p_n$ לפי סדר עולה של $\delta(p)=\max_{x\in P} d(p,x)$

2 הנק' שמשיגות את הקוטר, ה-δ שלהן מקסימלי.

לא באמת מחשבים את זה, כי זה עולה $O(n^2)$. רק קונספטואלית.

אם נבחר נק' אקראית p_j, ונסתכל על $\delta_j=\delta(p_j)$ (שאותו מחשבים בזמן לינארי) ונפעיל את פונק' ההכרעה על δ_j.

אם הצלחנו, δ_j הוא הקוטר, כי הוא מכיל את כל הנק' אבל הוא גם המרחק בין p_j למישהו.

אם לא הצלחנו, נסתכל על P_0, ונטען ש-$p_1,p_2,\dots,p_j\in P_0$. הנק' הללו נמצאות בכל הכדורים ואפשר לזרוק אותן. אנחנו גם ממש זורקים אותן, ובעצם אפשר לזרוק j נק' ולהמשיך ברקורסיה.

(מקס') תוחלת זמן הריצה $T(n)$ היא:

$$T(n)\le\frac{1}{n}\sum_{j=1}^{n}T(j)+\underbrace{D(n)}_{\text{עלות פונק' ההכרעה}}$$

כלומר אם בחרנו את j נשארנו עם בעיה בגודל $j-n$, וההסתברות שנבחר זאת $\frac{1}{n}$.

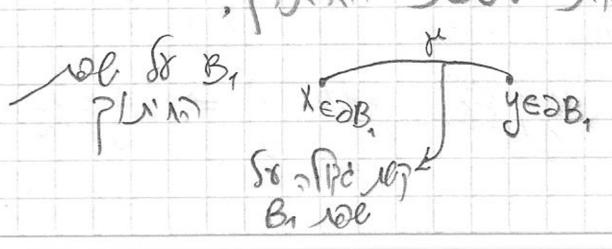
נראה ש-$D(n)=O(n\log n)$ וזה גורר שגם $T(n)=O(n\log n)$.

בהנחה סמויה, שהפלט, כלומר החיתוך הוא בסיבוכיות $O(n\log n)$.

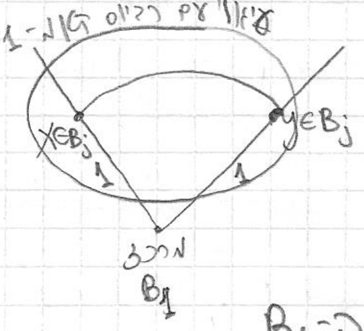
אז מה יש לנו ב-$D(n)$. צריך לחשב את החיתוך של n כדורי יחידה ב-3D, ואז לבדוק מי בפנים ומי בחוץ. החלק השני יחסית קל.
החלק הראשון קצת יותר קשה, ולשם כך צריך להיזכר בחישוב קמור בתלת ממד עם אלג' רנדומי אינקרמנטלי.

סיבוכיות החיתוך היא לינארית, וזה קצת לא מובן מאליו, כי אם הכדורים לא היו כדורי יחידה אלא כלליים היא הייתה ריבועית, והחיתוך של כדורי יחידה גם כן יכול להיות ריבועי! במקרה שלנו יש לנו מזל, והחיתוך הוא קב' קמורה, המתקבל ע"י קב' קמורות. הטענה היא שכל כדור תורם $\ge$ פאה אחת לשפת החיתוך.

נוכיח:



הדרך הקצרה ביותר לחבר 2 נק' על כדור היא ע"י שיהיו על מעגל גדול, כמו קו המשווה שמרכזו במרכז הכדור. כך נראה שכל הנק' ב-B_1 נמצאות על אותה שפה וכך כל הפאה מוכרת לעצמה.



ההוכחה על בררוק?

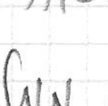
החיתוך של המישור הנפרד עם B_j הוא עיגול ברדיוס < 1

אז מה שרואים זה שכל הקשת מוכל ב-B_j.
לכן כל הקשת בעיגול, וכל הקשת ב-B_j, ולכן כל כדור תורם פאה אחת.
לפי אוילר, ומספר מישורים וכו' נקבל שסיבוכיות X היא $O(n)$.

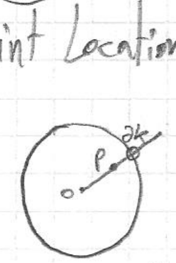
אלג' רנדומי אינקרמנטלי דומה לחישוב קמור בתלת ממד. שם היה לנו מין Conflict list - החזקנו רשימה של כל הנק' שעוד לא הכנסנו ושכאשר נכניס תהרוג את אותו feature. כשהיו פרפרים, ברשימה היו כל הנק' שרואות את הפרפר מבחוץ.
כל נק' שנכנסת, מעדכנת את כל הרשימות...

מה שנעשה כאן זה נחזיק את שפת X, אחרי הוספת כל כדור. כל קשת של X תצביע על כל הכדורים שעוד לא נוספו שלא מכילים את הקשת בשלמותה.
כל פעם שנוסיף כדור, הוא "ילך" לכל הקשתות שהוא פוגע בהן, ויחשב את כל הקשתות שמקוצרות (או נעלמות) ע"י פגיעת הכדור.

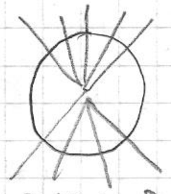
אם סדר ההכנסה הוא רנדומי, אז תוחלת אורך הרשימות מוגבלת. הגבול בהן הוא $O(n\log n)$.



אח"כ יש את אותו החלק השני היותר פשוט שמבצע ע"י point location.



מה שעושים זה תופסים נק' o בתוך החיתוך, ורוצים לבדוק האם p בתוך החיתוך או לא. אז שולחים קרן מ-o דרך p ורואים איפה שפה הקרן חותכת, ובודקים לאיזה כדור שייכת נק' היציאה.



אפשר להכין גם את המפה המושרית שהיא
על פני כדור ועליו להכין מבנה Point Location...

ב-4 ממדים זה כבר לא יעבוד כי כל המבנה יהיה בעל סיבוכיות
ריבועית, וצריך משהו אחר.

מבחן
9/2/14

אלגוריתמי קירוב

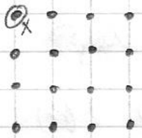
לפעמים נתקלים בבעיה קשה, שמ-קשה או סתם בעיה כמו קוטר שלוקחת
$O(n^2)$ ורוצים יותר טוב.

הרבה פעמים מה שנעשה זה נבנה גריד, והנק' המקוריות יזוזו לנק' הגריד.
נשתמש בבעיית הקוטר כבעיה מבחן.

בד"כ, מה שמחפשים הוא PTAS = poly. Time Approx. Scheme.
ושאומרים פולינומיאלי מתכוונים poly.$(n, \frac{1}{\varepsilon})$. ומחפשים פתרון עד כדי קירוב $(1\pm\varepsilon)$.
שוב, במקרה של הקוטר כבר יש לנו, ורק רוצים יותר טוב.
מכאן, לפעמים נחפש LTAS = Linear.... כלומר $O(n\cdot f(\frac{1}{\varepsilon}))$.
ולפעמים מדברים על strongLTAS שזה $O(n+f(\frac{1}{\varepsilon}))$.

שוב, בבעיית הקוטר (ב-R^d) ופתרון (קצת משעמם לפי מיכה) הוא Grid.

בונים שריג.



נניח שביעילות ובמהירות יש לנו (פונק') floor וניתן
להחליף כל נק' בקודקוד של תא
במקום כל תא והנק' שלו, בוחרים נק'
אחת שמייצגת את כולם

כמו כן, לפעמים גם נוכל להיעזר בעובדה שאנחנו על גריד. למשל
אם מחפשים קוטר ויש לנו

אז כל הנק' באמצע לא מעניינות...

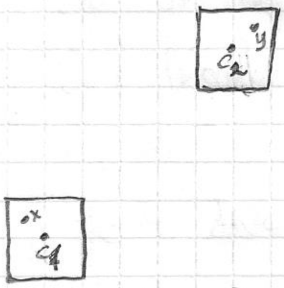
אם עושים לא היינו צריכים אם קירוב של ε אלא נגיד היינו מעוניינים בימס קירוב 2, זה טריוויאלי (קצת כמו האלג' של גונזלס).
נבחר נק' $p_0 \in P$ כלשהי, ונחשב $\delta(p_0) = \delta_0$. זה לפחות $\frac{1}{2}$ הקוטר ולכל היותר הקוטר.

$$D(p) = d(x,y) \le d(p_0,x) + d(p_0,y) \le 2\delta(p_0)$$

עכשיו נבנה סריג. גודל הסריג יהיה $\varepsilon\delta_0$.

בכל קואורדינטה המרחק בין הקואור' המינימלית למקסימלית של נק' P הוא לכל היותר $2\delta_0$. (פשוט הטלה על ציר x של הדיון הקודם, או לגמרי דיוק של הקוטר, שבהטלה רק קטן).
לכן, יהיו $\frac{2}{\varepsilon}$ תאים בכל ציר. נסמן $E = \frac{1}{\varepsilon}$ ולכן יש לנו $\left(\frac{2}{\varepsilon}\right)^d = O(E^d)$ תאים.

נפתור את הבעיה ב- brute force, בזמן ריבועי עבור מרכזי התאים. זה יעלה $O(E^{2d})$, וקיבלנו קוטר.



$$d(x,y) \le d(c_1,c_2) + d(x,c_1) + d(y,c_2)$$
$$\le d(c_1,c_2) + \frac{2\varepsilon\delta_0\sqrt{d}}{2}$$

$x,y \in P$

פיתגורס, ב-d ממדים או משהו

$$\Rightarrow \underbrace{d(x,y) \le D(Q) + \varepsilon\sqrt{d}\,\delta_0}$$

$D(p) \le \delta_0 \le 2D(p)$

$$\underbrace{D(p)(1+\cdots\varepsilon)}_{\text{באופן דומה}} \ge D(Q) \ge D(p)(1-\varepsilon\sqrt{d})$$

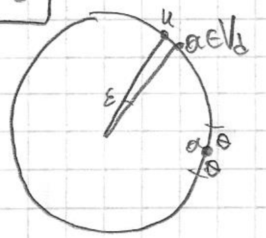
(נגיד אם מדוייקים על $D(Q)$ אז על 2 נק' מ-P שיוצרות את הקוטר...)
אז קיבלנו אלג' שרץ ב- $O(n+E^{2d})$
לשלוח נק' לתאים, ולחשב δ_0.

אפשר לשפר. הנה עוד שיטה, שנוגעת לגרעינים...

2) נסתכל על שפת כדור היחידה ב-d ממדים (S^{d-1}), ונרצה אותה בנק' V_d עם התכונה שלכל $u \in S^{d-1}$ קיים $a \in V_d$ כך שהזווית בין a ל-u $\approx$ המרחק ביניהם על הכדור קטן מ-θ.

$\boxed{\cos\theta = \frac{1}{1+\varepsilon}}$.

שגיאה בערך $\sim 1-\varepsilon$



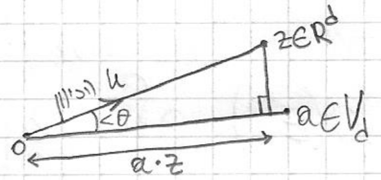
סביבת ε של נק' a ~ כדור ברדיוס ε ב-$d-1$ ממדים.

רוצים לרצף את הספירה, מרצפים אותה במין כיפות... יש לנו פה על מישורים ועל כדורים... בלגן.
הנפח הזה $\sim \theta^{d-1}$.
*להעזר ע"י כדורי a...

$$\cos\theta = \frac{1}{1+\varepsilon} \sim 1-\varepsilon \sim 1-\frac{\theta^2}{2} \rightarrow \theta = \sqrt{2\varepsilon}$$

לכן:

$$O\left(\varepsilon^{\frac{d-1}{2}}\right) = |V_d| = O\left(\frac{1}{\theta^{d-1}}\right)$$



הנה דוגמה לנק' z והכיוון שנוצר והכיוון עבור a שלוקחים במקומו.
נ"ל (מכפלה סקלרית) את הנק' המקורית.

$$a \cdot z \geq \|z\|\cos\theta = \frac{\|z\|}{1+\varepsilon} \quad \text{(מכפלה סקלרית)}$$

רוצים לחשב בערך את הקוטר (גודל הקוטר) בין זוג נק' שמשיגות את הקוטר. במקום זה יהיו לנו כיוונים אחרים.
כאמור, מחפשים:

$$\max_{x,y \in P} \|x-y\|$$

מה האי שיוויון מהמכפלה הסקלרית גורם לנו?

$$\forall a \in V_d \rightarrow \max(x-y)\cdot a = \max_{x \in P} x \cdot a - \min_{y \in P} y \cdot a$$

ומקבלים 2 תת' בעיות פלגי גלויות שבהן סה"כ צריך לחשב מקס', ומוצאים את ההפרש הכי גדול, בקלות, עם קירוב עד כדי $(1+\varepsilon)$.
זה נותן אלג' ב- $O(n\varepsilon^{\frac{d-1}{2}})$. בערך מטילים את הקוטר על כיוונים...
אם קודם נחשב את הגריד ואז נחשב על הגריד, נקבל משהו שמשתמש LTAS
$O(n+\varepsilon^{\frac{3d}{2}})$.