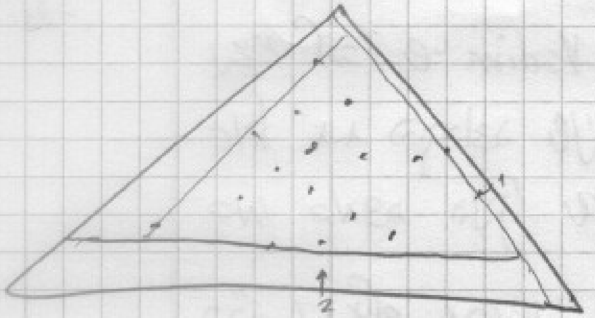
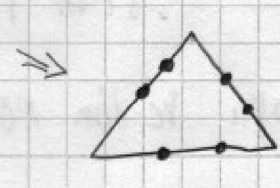
נק' ומשולשים ב- R^2:



לוקחים משולש. מצמצמים אותו ומסובבים
אותו



ולכן $O(n^6)$
אבל הוכחנו שה- VC-dim הוא 7.

אפשר גם להסתכל על זה הפוך.
אם ה-Corollary של הלמה נכון, כלומר יש n^δ טווחים, אז ניתן
לומר משהו על ה-VC-dim.

אז אם A, בגודל x הוא הכי גדול שניתן לנפץ, אז # ה-ranges
ב-A הוא 2^x והרי $2^x \le x^\delta \Leftarrow x \le \delta \log x \rightarrow x = O(\delta \log \delta)$

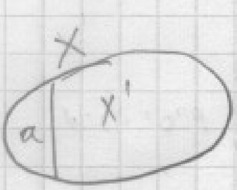
VC הראו: אם בוחרים $\frac{c\delta}{\varepsilon^2}\log\frac{1}{\varepsilon}$ נק', אנו מקבלים ε-קירוב בהסתב' גבוהה.

HW הראו: אם בוחרים $\frac{c\delta}{\varepsilon}\log\frac{1}{\varepsilon}$ נק', אנו מקבלים ε-net בהסתב' גבוהה.

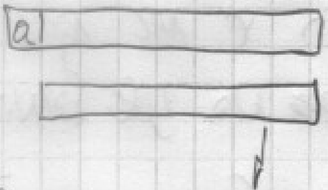
הרצאה

היום ממשיכים עם הנושא הקודם ועם ההוכחה של הלמה שמיוחה ג"כ לקו של
סאור-שלה.
אם כן, (X,R) הוא ה- range space עם $vcdim = \delta$, $|X|=n$.
אז הלמה אומרת ש- $|R| \le \binom{n}{0}+\binom{n}{1}+\dots+\binom{n}{\delta}$

עושים אינדוקציה על $n+\delta$.



$a \in X$
מסתכלים על $X \setminus \{a\}$, וה-ranges restricted שהם
לכל זה.
הטווחים $\ge F(n-1,\delta)$



range r

range r' on x'

אם זה גם אוסף של X

אז סופרים אותו פעם.

כלומר, את הטווחים ב-x' סופרים פעמיים.

$\Rightarrow \quad |R| \le F(n-1,\delta) +$ # הטווחים שסופרים פעמיים

נסתכל על הטווחים שנספרים פעמיים כטווחים על $x' = R'$

vc-dim של $(x',R') \le \delta-1$

$\equiv$ לא ניתן לנפץ סט בגודל δ.

ניתן לנפץ את ה-subset הזה $\cup \{a\}$ $\longrightarrow (a \mid \delta \; x')$

ב- (X,R)

$\Leftarrow$ סתירה לכך שה-vc-dim הוא δ.

$$\Rightarrow F(n,\delta) \le F(n-1,\delta) + F(n-1,\delta-1)$$

וזה קשור איכשהו למשולש פסקל ומקבלים את הטענה הרצויה.

בגדול, לא כל כך הבנתי מה הולך פה.

ובעת ל- ε-net ו- ε-approx.

נזכיר ש-ε-net : עבור range space (X,R), $N \subseteq X$ הוא ε-net,

ונניח ש-X סופי, אז לכל אוסף $r \in R$ עם $\varepsilon|X|$ נק' מכיל נק' של N.

Hausler - Welzl הוכיחו:

דגימה אקראית של $c\delta\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}$ של נק' מ-X $(|X|=n)$ עם $vc\text{-}dim(X,R)=\delta$ אז כן

ε-net בסבירות גבוהה.

נראה לרגע הוכחה יותר איטית, שבה במקום $\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}$ ניקח $c\delta\frac{1}{\varepsilon}\log n$.

ניקח אזור r המכיל $\ge \varepsilon m$ נק' מ-X.
המודל שלנו הוא כרגיל - דוגמים כל נק' ב-N באופן ב"ת בהסתב' $p=\frac{1}{n}[k]$

מה ההסתב' שלא נבחר אף נק' מתוך m הנק'? היא: $(1-p)^m$ = ההסתב' לכשלון

$$\sim e^{-pm} \le e^{-p\varepsilon n} = e^{-\varepsilon k} = e^{-c\delta \log n} = \frac{1}{n^{c\delta}}$$

כי ההסתב' שאזור אחד יהיה לא טוב. כמה אזורים יש לנו? נשתמש בלמה שראינו הרצ', ונשתמש ב-prob. union bound ולכן ההסתב' שאזור כלשהו נכשל היא לכל היותר:

$$O(n^{\delta})\cdot\frac{1}{n^{c\delta}}$$

מספיק שנבחר $c=2$ ואז נקבל הסתב' $\frac{1}{n^{\delta}}$ כלומר סבירות נמוכה מאוד שנכשל.

כעת נתמודד לטענה המקורית - כלומר במקום $\log n$ נעבור ל-$\log\frac{1}{\varepsilon}$.
ההוכחה שלהם כי מסובכת שמשתמשת בדגימה כפולה ועוד...
אנחנו נעשה משהו אחר, גם מסובך למדי.

ניקח מ-(X,R) ← נבנה ε-approx. $A\subseteq X$ בגודל $\frac{c\delta}{\varepsilon^2}\log\frac{1}{\varepsilon}$

ואז, בעזרת הבניה שעשינו לפני רגע, $N\subseteq A$, ε-net עבור (A,R) בגודל $\frac{c'\delta}{\varepsilon}\log|A|$
היות ו-A לא תלוי ב-n והוא יהיה קרוב ל-$\frac{1}{\varepsilon^3}$ (או משהו) הדבר לקירוב נקבל: $\frac{c''\delta}{\varepsilon}\log\frac{1}{\varepsilon}$
ואז N הוא 2ε-net ל-(X,R).

<u>הוכחה:</u>

ניקח אזור r, $|r|\ge 2\varepsilon n$ pts., $r'=r\cap A$ ⇐ $\left|\frac{|r'|}{|A|}-\underbrace{\frac{|r|}{|X|}}_{>2\varepsilon}\right|<\varepsilon$

$$\Rightarrow \frac{|r'|}{|A|}>\varepsilon$$

⇐ N מנקב את r' ⇐ N מנקב את r.

עכשיו מישהו אומר משהו על כך שניתן לקחת מלמעלה ובחיים ואחרים כדי שלא יהיה ה-2, אבל לא זה עד כדי קבוע.

הדבר היחיד שנותר להראות הוא כי קיים ε-קירוב $A\subseteq X$ בגודל שנתנו $\frac{c\delta}{\varepsilon^2}\log\frac{1}{\varepsilon}$

הרעיון הוא שנקבל את הסט שלנו ע"י גז' מהנק' שלו, ונצא אותו ע"י גז' ממנו וכן הלאה... ונראה שהשגיאה שנצבור לאורך הדרך מספיק קטנה.

אם כן, השאלה היא מה קורה כשלומדים גז' מסט.

ניתן לחשוב על זה כמו צביעה של כל נק' ב-X כמו $+1, -1$, בהסתב' $\frac{1}{2}$.

ניקח את ~~...~~ אחד ה-subsets, לשמור את כל אלו שהן $+1$.

אז אולי ניקח לא בדיוק גז' מהנק', אבל משהו קרוב לזה.

נסתכל על זה כמו range space (X,R)

ונסתכל על ה-Discrepancy לכל $r\in R$. שזה בערך ~~...~~ מה השוני בין הכמ' של ה-$+1$ לאלו שהם -1:

$$c: X\to\{+1,-1\} \Rightarrow c(r)=\sum_{x\in r} c(x)$$

וה-$\overbrace{\text{Discrepancy}}^{\text{סטייה מהדגימה}}$ היא

$$\min_{c}\left(\max_{r\in R}|c(r)|\right)$$

ה-Discr. היא תכונה של ה-range space

ישנו <u>משפט</u> שאומר:

קיימת צביעה עם Discrepancy של $\sqrt{2n\ln(2|R|)}$

~~אם ניקח~~ $\cancel{\min_c\left(\max_{r\in R}|c(r)|\right)}$

עם $|R|$ גדול זה יכול להיות די נורא, אבל עם $d=$VC-dim $\approx \sqrt{2n\ln|R|}=\sqrt{2n\ln n^{d}}$

מי שהוא אומר שכל העניין הזה מופיע בספר של פאק ואגרוואל, למשל.

המשפט הוא מעבר של חסם צ'רנוף.

כזכור, הוא אומר שעבור מ"מ $X_1,\dots,X_n$ שהם ±1 בהסתב' $\frac{1}{2}$, ומסתכלים על הסכום שלהם: $X=\sum X_i$ - זהו מ"מ עם תוחלת אפס, ו-std של $\frac{\sqrt{n}}{2}$.

חסם צ'רנוף אומר ש-

$$\mathrm{Prob}(|X|>\Delta)<2e^{-\frac{\Delta^2}{2n}}$$

וזה בדיוק מה שיש לנו. ל-$r\in R$:

$$\mathrm{Prob}(|c(r)|>t)<2e^{-\frac{t^2}{2|r|}}\le 2e^{-\frac{t^2}{2n}}$$

n של X, לא של צ'רנוף

נסמן $t=\sqrt{2n\ln(2|R|)}$ אז $= 2e^{-\frac{2n\ln(2|R|)}{2n}} = 2e^{-\ln(2|R|)}$

$$= 2\cdot\frac{1}{2|R|} = \frac{1}{|R|}$$

⇐ עפ"י union bound prob., ההסתב' שאת מס' הא' יותר מדי לא מאוזן קטנה מ-1 ⇐ קיימת צביעה עם ה-discrepancy הרצוי.

אם כן, ה-X שנו מחולק ל-X^+ ו-X^-

ניקח $r\in R$:

$r^+=r\cap X^+$

$r^-=r\cap X^-$

$$|c(r)| = \big||r^+|-|r^-|\big| \le t$$

שזה כמו:

$$\big|2|r^+|-|r|\big| \le t$$

נחלק ב-$|X|$:

$$\left|\frac{2|r^+|}{|X|}-\frac{|r|}{|X|}\right| \le \frac{t}{|X|} = \sqrt{\frac{2\delta\ln n}{n}}$$

נרמה קצת ונרשום את $\frac{2|r^+|}{|X|}$ בתור $\frac{|r^+|}{|X|/2}$ ואז כתוב בתור $\frac{|r^+|}{|X^+|}$

אבל מיכה אומר שהרמאות הזאת מסבר כי גם ככה השגיאה בגלל הרמאות קטנה ביותר, הולכת לאפס בגבול.

$$\Rightarrow \left|\frac{|r^+|}{|X^+|}-\frac{|r|}{|X|}\right| \le \sqrt{\frac{2\delta\ln n}{n}}$$

אם כן, הרי הרשום למעלה נכון לכל r, וזה קטן הרבה יותר מ-ε אצלנו:

$X_0=X$

$X_1=X^+$

$X_2=(X^+)^+$

$\vdots$

$r_i=r\cap X_i$

$n_i=|X_i|$

$$\Rightarrow \forall r: \left|\frac{|r_i|}{|X_i|}-\frac{|r_{i-1}|}{|X_{i-1}|}\right| < \sqrt{\frac{2\delta\ln n_{i-1}}{n_{i-1}}}$$

וזה סוג של אסקופ', אם סוכמים את כל השגיאות...

עוצרים בשלב מסוים ומסתכלים על השגיאה:

ממש בלגן..

$$\left|\frac{|r_i|}{|X_i|} - \frac{|r|}{|X|}\right| < \sum_{j=0}^{i-1}\sqrt{\frac{2\delta\ln n_j}{n_j}} \sim \sqrt{2\delta\ln n}\sum\frac{1}{\sqrt{n_j}} \sim$$

$n_j = \frac{n}{2^j}$

$$\sim\sqrt{2\delta\ln n}\sum\frac{1}{\sqrt{n/2^j}} \sim \sqrt{\frac{2\delta\ln n}{n}}\cdot\sum\sqrt{2^j} \sim \sqrt{\frac{2\delta\ln n}{n}}\cdot 2^{i/2} = \varepsilon$$

בסופו של דבר רצינו ε-קירוב.

מה שרצינו לומר זה ש-

$$\sum\sqrt{\frac{\ln n_j}{n_j}} \sim \frac{\ln n_i}{\sqrt{n_i}}$$

בקיצור - ה-log במונה גדל יותר לאט מהמכנה. וזה פרופורציונלי גאומטרי בערך לערך המחובר הסופי

$$\Rightarrow \text{sum} \sim \sqrt{2\delta}\cdot\sqrt{\frac{\ln n_i}{n_i}} = \varepsilon$$

$$\frac{n_i}{\log n_i} \sim \frac{\delta}{\varepsilon^2}$$

$$\rightarrow n_i \sim \frac{\delta}{\varepsilon^2}\log n_i = O\left(\frac{\delta}{\varepsilon^2}\log\frac{1}{\varepsilon}\right)$$

<u>האינטואיציה:</u>

בקיצור, אמנם הראינו שניתן לקבל ε-net עם $\frac{\delta}{\varepsilon}\log\frac{1}{\varepsilon}$, ועשינו זאת ע"י חלוקה ב-2 שוב ושוב עד שהיינו מרוצים מגודל הקבוצה ~~[illegible]~~ והראינו שהטעות היתה קטנה ולכן מלכתחילה יכולנו לקחת משהו בגודל זה.

הוכח ל- $\frac{\delta}{\varepsilon}\log\frac{1}{\varepsilon}$ הנ"ל. ל-ε-קירוב $O\left(\frac{\delta}{\varepsilon^2}\log\frac{1}{\varepsilon}\right)$ לא הנ"ל.

<u>כעת נעבור קצת לגיאומטריה.</u>

נעבור על כמה יישומים של כל ענין ה-sampling.

(פעם שעברה)

ראינו דוגמאות לטווחים עם כל מיני מימדי VC.

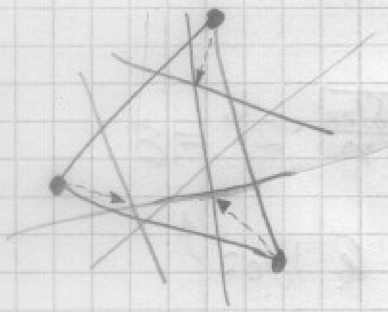
(משולשים, קטעים)

נגיד ש-X הוא כל של עצמים גיאומטריים פשוטים נק', כדורים, גלילים, קווים
במימד מוגדר, בעלי תיאור פרמטרי קבוע,

איזור פשוט ~~[illegible]~~ כלשהו

נגיד ש-R הוא אוסף כל תתי הקבוצה של X שיש להם אינטרקציה עם

עבור X ו-R כאלו $\Leftarrow$ ה-$VC\text{-}dim(X,R)$ הינו סופי.



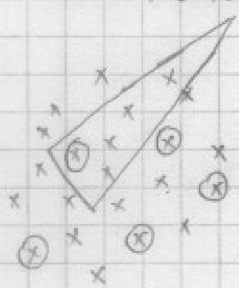
דוגמה משמאל - קדקדי המשולש נמצאים במפגש של המערך. אנו יכולים להזיז את קדקדי המשולש בתוך כל פאה (שבפאה) כך שעדיין אותם הישרים חותכים את המשולש.

לכן ניתן להעביר את הקדקדים של המשולש לנק' חיתוך

לכן כמות ה-ranges הינה לכל היותר $(n^2)^3 = \binom{n^2}{3}$. (הנוצרים ע"י משולשים.)

למרות שיש ∞ משולשים אפשריים.

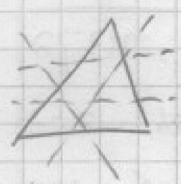
נניח שיש לנו קב' של נק' במישור, ה-ranges הם משולשים ונרצה



ועוד דוגמה:

L = קב' של n ישרים ב-R^2

R = אוסף של קבוצות של L שחותכים משולש.

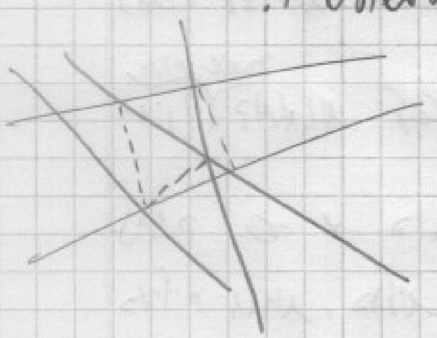


דגימה מקרית של $O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$ ישרים של L

היא ε-net בהסתברות גבוהה.

זה אומר שאם משולש Δ אינו נחתך ע"י שום ישר של N, אזי הוא נחתך ע"י $< \varepsilon n$ ישרים של L.

נבנה את המערך $A(N)$, ונחלק כל פאה למשולשים.



מחשבים את מבנה המשולש, שיכלול [illegible] הפינים פ"יים.

אבל משאירים ישרים שחותכים את הפנים של משולש. אם חותך קדקד, לא משאירים.

גם כאן ה-vc-dim עם הסיפור של ϵn ...
שזה בעצם cross - לחצות את הפנים.
intersect

קיבלנו אוסף של משולשים, כך שכל משולש נחצה ע"י לכל היותר ϵn ישרים.

נבחר $r=\frac{1}{\epsilon}$ וכעת $|N|=O(r\log r)$
ואז כל משולש הינו crossed (נחצה) ע"י $\le \frac{n}{r}$ ישרים של L.

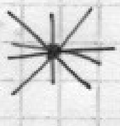
מס' המשולשים בפאה הוא לכל היותר #הצלעות פחות 2.

לכן: $\# \text{המשולשים} \le \sum_{f\ \text{faces}} (e(f)-2)$

$$2E-2F=2V+4$$

$$\Rightarrow O(V)=O(t^2)$$

מה הוא פרמטר.
אזי קיימת חלוקה (partition) של המישור ל- $O(r^2\log^2 r)$ משולשים, כך שכל אחד מהם נחצה ע"י לכל היותר $\frac{n}{r}$ ישרים של L.

ישנם כמה מכשולים. הגבולות (boundaries) הם בעייתיים.
לדוגמא  אם נתקרב לקדקוד עם משולש הישרים יחתכו אותו.

עוד משהו - החסם הזה איננו הכי טוב.

קודם כל נקרא לילד בשמו- חלוקה שכזאת נקראת $\frac{1}{r}$-cutting.

<u>הטענה</u>: לכל $\frac{1}{r}$-cutting יש בהכרח $\Omega(r^2)$ משולשים.
זה בהנחת מצב כללי. (למרות שגם אם לא זה בסדר).
בהנחת מצב כללי, לא צריך לדאוג לגבולות.

נסתכל על כל הקדקודים של $A(L)$: ישנם $\binom{n}{2}$ כאלו.
משולש של ה-cutting נחצה ע"י $\frac{n}{r}$ ישרים לכל היותר.

⇐ לכל היותר $\binom{n/r}{2}$ קודקודים בתוך כל משולש.

⇐ # המשולשים $\geq \frac{\binom{n}{2}}{\binom{n/r}{2}} \sim r^2$

עם עוד קצת עבודה, אפשר להראות שקיימת $\frac{1}{r}$-cutting עם $\Theta(r^2)$ משולשים.

סיכום: מתחילים עם דגימה של רק r ישרים.

נקבל משולשים שנחתכים ע"י ישרים רבים מדי.

ע"י שימוש בגודל הרשת של ε-net, כל משולש נחתך

ע"י $\leq \varepsilon n = \frac{cn}{r}\log r$ בהסתב' גבוהה.

$\left[\varepsilon \sim \frac{1}{r}\log\frac{1}{\varepsilon} \sim \frac{1}{r}\log r\right.$

יש לנו $O(r^2)$ משולשים, אבל לא בדיוק $\frac{1}{r}$-cutting.

השלב השני - תיקון משולשים לא מוצלחים.

בממוצע, משולשים לא נחתכו ע"י יותר מדי ישרים. בממוצע רק ע"י $\frac{n}{r}$. (נגמול יקיים..)

העודף (Excess) של משולש Δ, t_Δ יהיה: $t_\Delta = \frac{\text{\# הישרים החותכים את } \Delta}{n/r}$

Δ נחתך ע"י $\frac{t_\Delta n}{r}$ ישרים.

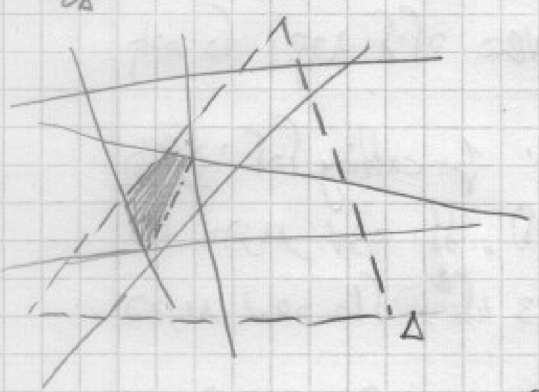
אם $t_\Delta < 1$, מפצלים את Δ. איך?

נסתכל על L_Δ - הישרים החותכים את Δ.

$|L_\Delta| \sim t_\Delta \frac{n}{r}$

נבנה $\frac{1}{t_\Delta}$-cutting של $A(L_\Delta)$, וכעת כל משולש נחתך ע"י $\leq \frac{|L_\Delta|}{t_\Delta} = \frac{n}{r}$ ישרים של L_Δ

נקבל $O(t_\Delta^2 \log^2 t_\Delta)$



אבל עכשיו יש גם יותר משולשים! כמה?

$\#\text{משולשים הכולל} = O(r^2) + \sum_{t_\Delta > 1} t_\Delta^2 \log^2 t_\Delta$

לשם כך, באה למה, שלא נוכיח, הנקראת Exponential decay Lemma

המשולשים המקוריים של ה-cutting הראשון עם $t_\Delta \ge t$ היא $O(r^2\cdot 2^{-t})$ בהסתב' גבוהה.

זה מספק את הסכום המבוקש של כמות המשולשים החדשים לשהו תב"ב:

$$O(r^2)+\sum_{t_\Delta>1} t_\Delta^2\log^2 t_\Delta \to \sum_{t=1}^{\infty} t^2\log^2 t \cdot \#\text{המשולשים עם } t_\Delta\ge t \sim r^2\cdot\sum 2^{-t}\cdot t^2\log^2 t = O(r^2)$$

הסיפור שעשינו כאן ב-2 שלבים, נכון לא רק למשולשים, והוא נכון
גם במימדים יותר גבוהים