

Introduction to Machine Learning

מחברים: יובל, סלביץ

רוב העציון שפותרו העלם אחידה (ראו ככה)

[illegible]

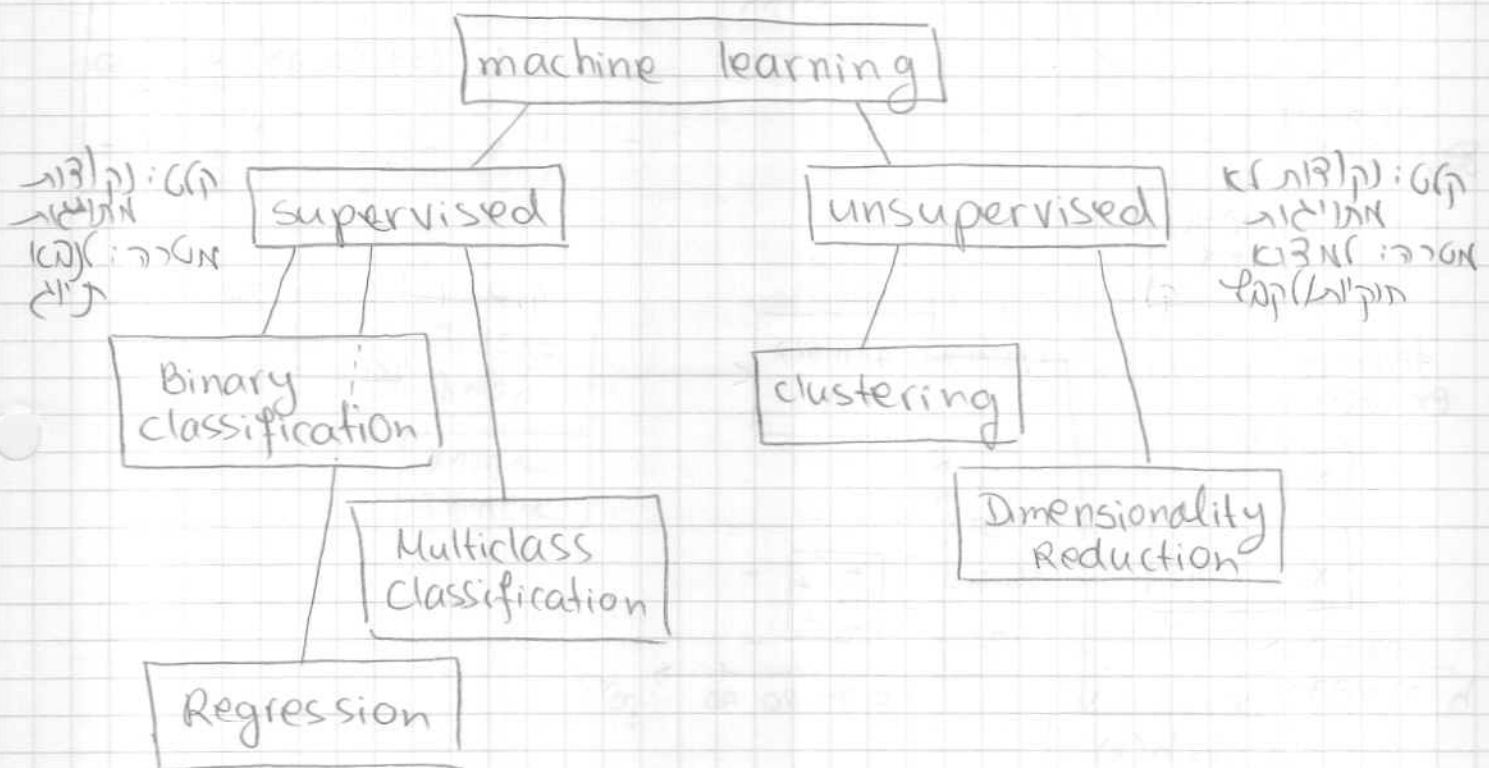
המטרה היא ליבוי כמה שיותר מדויק א קואנטל תדירות (ירי א)
היטות רבה ידועה התשובה).

כאן תחנכים את המכונה - Machine Learning?

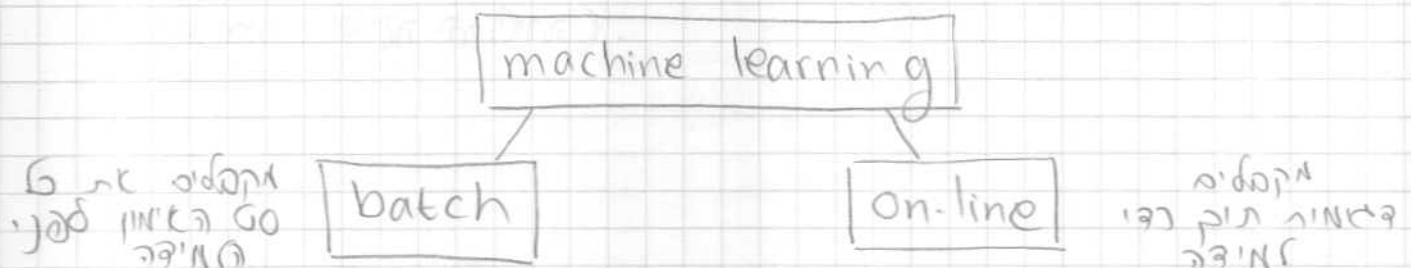
- עיבוד תמונות
- ביאנפואטיקה
- Text-mining
- קורס
- Reverse Engineering
- Junk mail filtering

מה כלל אינרנטי חייב להם טובים? - לא תמיד (יתר)
 מה פיתוח צרכנים, ובעיות חול מהם, אפסטיים צורק
 האינרנטי אפסטיים (כמו למשל המורה של דואר לבד)

סוגים של אינרנטיים אומדים:



עוד חוקר של אינרנטיים אומדים:



2

★ X - instance space

: סימן

★ $x \in X$ - input instance

$x = (0, 0, 1, 0) - !$ $X = \{0, 1\}^4$ (ערך)
 $x = \text{"I am a string"}$ $X = \Sigma^*$

★ Y - output / label space

$Y = \{\pm 1\}$, $Y = \{0, 1\}$, $Y = \mathbb{R}$ (ערך)

★ (x, y) - example

★ training set $Z = \{(x_1, y_1), \dots, (x_m, y_m)\}$

★ $S = \{x_1, \dots, x_m\}$ - סט של נקודות

i.i.d (independent identically distributed)

$$\mathcal{D}^m = \mathcal{D} \times \dots \times \mathcal{D}$$

★ $C = H = \{h \mid h: X \rightarrow Y\}$ -

Concept / Hypothesis class

★ h - concept / hypothesis classifier

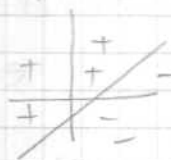
: דוגמאות

- Decision tree

- Conjunctions

- Separating hyperplanes

$\hookrightarrow$



★ c_t - target concept

הקונספט היעד

Realizable case - זה $C_t \in C$ ואז $Z = \{(x_i, C_t(x_i))\}$

Unrealizable case - זה $C_t \notin C$ ואז Z (צדק)

אם $X \times Y$ אז $\mathcal{D}$

נניח שקרה אם נאמר לא לקחת מספר פרמטרים אפילו יותר
זהו שני אנשים עם את תמונתו צופק אם את חוזה ואז
היום. במקרה זה אף איש לא היה יכול לנבא

הסבר: יש הרבה אחיזי שמספר את הצדדים שלנו, אלא שאנחנו לא
יודעים אותו ואלו אנחנו חזים עתה. הפעם היא שאנחנו עובדים
עם מתקן היפוטזה בלבד. וזהו להיות שקונסטרוקציה לא נמצא
היא במקרה זה היא Unrealizable. ואילו אם קונסטרוקציה
היא כן נמצא המתקן היפוטזה שאנחנו עובדים איתה,
אז אנחנו במקרה Realizable.

אם, אם מתקן היפוטזה היא $\mathcal{H}$ separation hyperplanes
ואילו התחום הנכון הוא דיוקא נמצא אז זה Unrealizable.

הצדדים $\mathcal{H}$ היפוטזה h ו- $\mathcal{D}$

$$\text{err}(h) = P_{\mathcal{D}}[C_t(x) \neq h(x)] = \int_X \text{ind}(C_t(x) \neq h(x)) \mathcal{D}(x) dx$$

$$\text{ind}(x) = \begin{cases} 0 & x = \text{false} \\ 1 & x = \text{true} \end{cases}$$

Validation

נניח שאנחנו חוצים ומכירים h היפוטזה h .

המטרה: לשער / לחסום את $\text{err}(h)$

נניח מרחב נתונים $\{(x_1, y_1), \dots, (x_m, y_m)\}$ i.i.d. $\mathcal{D}^m$

נסמן את הטעות האמפירית

$$\hat{\text{err}}(h) = \frac{1}{m} \sum_{i=1}^m \text{ind}(y_i \neq h(x_i))$$

מה הקשר בין הטעות האמפירית לטעות האמיתית?

4) 30.10.04
ת. מ' אולמברג

אנחנו נחזיק במסלול אחריותה של משפט הופצ'נג ואז נסבור (פצ'נג)
יותר מעניינים

הצורה: עבור $0 \leq p, q \leq 1$ (גורם אחד)
 $D_{KL}(p||q) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q}$
 וצורה $p=q=0$ $0 \ln \frac{0}{0} = 0$

זה נקרא אחדות Kullback-Leibler divergence (ח) & סבור
 אחדות זה לא מקיים את אי שוויון הנחשבות אבל הוא כן מאר שמושי
 נוכח גורם קצת שונה של משפט הופצ'נג.

משפט: (הופצ'נג-צ'ינג) : עבור $X_1, \dots, X_m$ i.i.d. $X_i \in \{0, 1\}$
 כאשר $X_i = \begin{cases} 0 & \text{בהסת' } 1-p \\ 1 & \text{בהסת' } p \end{cases}$ $0 < \varepsilon$ δ

$$\begin{aligned} 1) \quad P\left\{\frac{1}{m} \sum_{i=1}^m X_i \geq p + \varepsilon\right\} &\leq e^{-\Theta(p + \varepsilon \|p\|)m} \leq e^{-2\varepsilon^2 m} \\ 2) \quad P\left\{\frac{1}{m} \sum_{i=1}^m X_i \leq p - \varepsilon\right\} &\leq e^{-\Theta(p - \varepsilon \|p\|)m} \leq e^{-2\varepsilon^2 m} \end{aligned}$$

חזקה

הערבה: $EX_i = 0 \cdot (1-p) + 1 \cdot p = p$

גם שזה דומה (ה) זה שיש הופצ'נג אחר ונשתמש באנחנו בחזקה
 ונזכור לחסום את המרחק בין הטעות האופיינית לבין התוחלת של
 הטעות האמיתית משפט הופצ'נג-צ'ינג אפשר לנו לעשות זאת.

קורחבה: (נסו) $q = p + \varepsilon$ וכן

$$P\left\{\sum_{i=1}^m X_i \geq mq\right\} = P\left\{e^{\lambda \sum X_i} \geq e^{\lambda mq}\right\} =$$

שטח צ'ינג $\lambda > 0$!

$$= P\left\{\prod_{i=1}^m e^{\lambda X_i} \geq e^{\lambda mq}\right\} \leq \frac{E\left(\prod_{i=1}^m e^{\lambda X_i}\right)}{e^{\lambda mq}} =$$

אם אחדות

$$= \frac{\prod_{i=1}^m E e^{\lambda X_i}}{e^{\lambda mq}} = \left(\frac{E e^{\lambda X_1}}{e^{\lambda q}}\right)^m = \left(\frac{p e^{\lambda} + (1-p)}{e^{\lambda q}}\right)^m =$$

משפט שסבור

$$= (p e^{\lambda(1-q)} + (1-p) e^{-\lambda q})^m = (*)$$

זה נתון $\delta < \lambda$ ואנחנו רוצים תסם כמה שיותר בדיוק, אז אנחנו רוצים למצוא את $(*)$. אחרי גזירה והשוואה (אפס) (ר' אתר

הקורס) נקבל שהמינוס מתמצה עבור $\lambda = \ln\left(\frac{(1-p)q}{(1-q)p}\right)$. אז

$$\Leftrightarrow \frac{1-p}{1-q} > 1 \text{ וזה } \frac{q}{p} > 1 \text{ וזה } q = p + \epsilon > p$$

$$\Leftrightarrow \frac{1-p}{1-q} \cdot \frac{q}{p} > 1 \Leftrightarrow \ln\left(\frac{1-p}{1-q} \cdot \frac{q}{p}\right) > 0 \text{ , אז אנחנו מסתדר.}$$

אם נציב λ ב- $(*)$ (ר' אתר הקורס) נקבל $(*) = e^{-\Theta(q \| p)^m}$

ולכן מראה לנו את (1), אי אשווין (2) מקבלים הסתברות אפס.

(11)

תרגיל בית

$$(1) \quad \Theta(p \| q) \geq 2(p-q)^2$$

(2) לעבור על ההשערה הסתברותית של λ שלבי מסתברות ונראה ש λ .



נוצר למידה PAC (Valiant 1984)

Probably approximately correct

הגדרה: יהי C מחלקת מושגים המוגדרת בין X ו- Y .

נאמר ש- C למידה PAC אם קיים אלגוריתם למידה L כך

התכונה הסתברותית: בהינתן $\frac{1}{2} > \delta, \epsilon > 0$ קיים אלגוריתם $m_0(\epsilon, \delta)$ *

הק, שבהינתן $m \geq m_0$ באופן Z באופן $m \geq m_0$ (הנצבים באקראי) מתקן

התפלגות Θ על $X \times Y$ בהסתברות לפחות $1-\delta$ האלגוריתם

מחזיר היפוטזה $h = L(Z) \in C$ הנקיימת

$$\text{err}(h) \leq \text{Opt}(C) + \epsilon$$

כאשר $\text{Opt}(C) = \min_{h \in C} \text{err}(h)$. m_0 $h^* = \arg \min_{h \in C} \text{err}(h)$

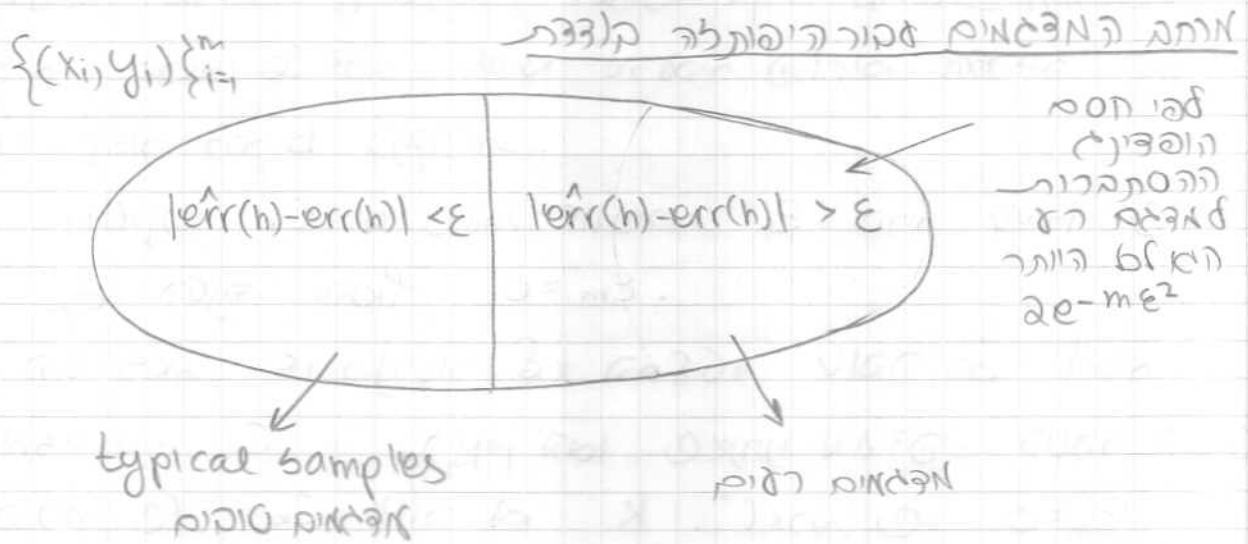
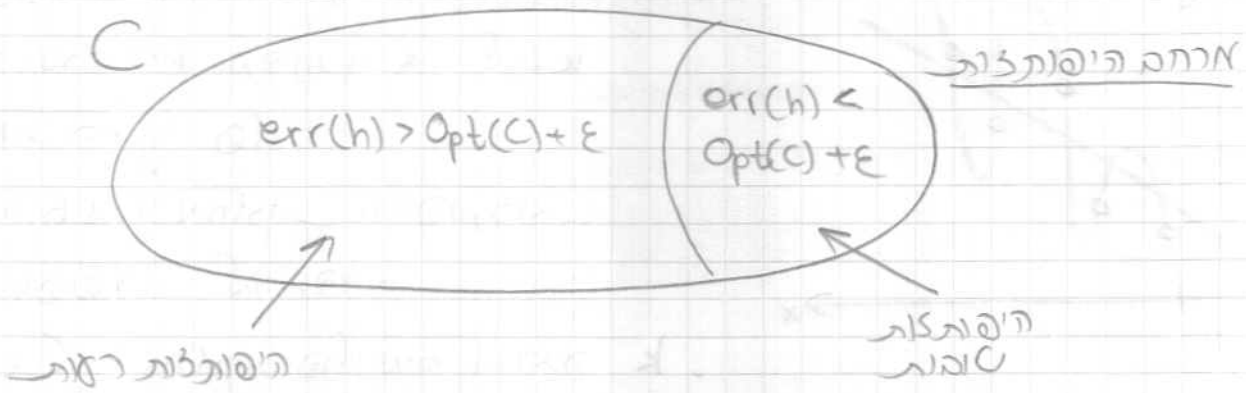
(נשים לב שהמקרה realizable $\text{Opt}(C) = 0$).

m_0 נקרא sample complexity של L .

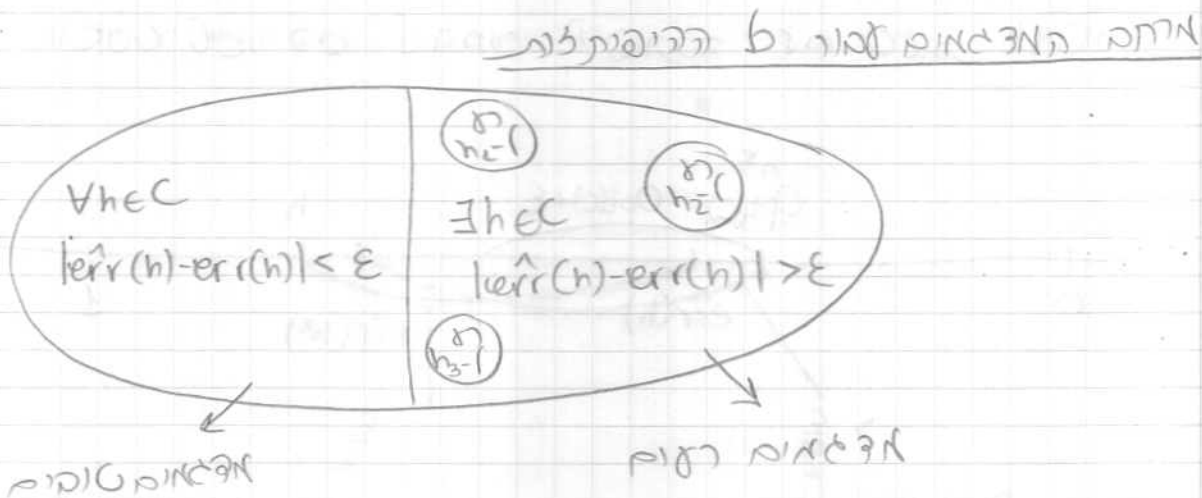
*דוגמה: m_0 δ תלוי ב- Θ . הוא צריך להיות פולינומי ב- $\frac{1}{\epsilon}$ וב- $\frac{1}{\delta}$ ולכן הכיזה של L צריך להיות פולינומי ב- $\frac{1}{\epsilon}$ וב- $\frac{1}{\delta}$ או חרא רגיל המייחד של X .

5

אילוסטרציה - PAC :



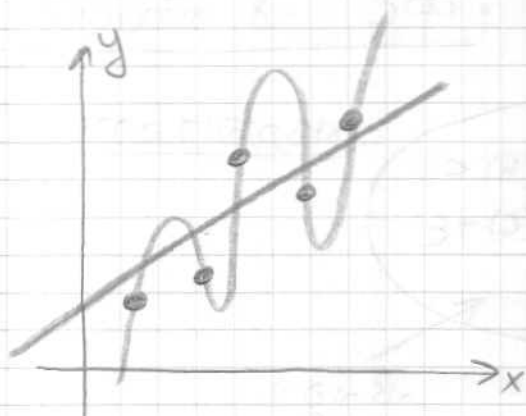
נשים $\mathcal{H}$ שמחשבת הופצ'ת התנוי שבתחת ההיפותזה נדחת תלויה במדגם.



הפעולה היחידה היא שיתן ערכים של ϵ ו- m ויחזיר את $\mathcal{H}$ ו- $\mathcal{C}$. אם $\mathcal{H}$ הוא מרחב המצבים

$$h(x) = \begin{cases} y_i & \text{אם } x = x_i \text{ והצג } (x_i, y_i) \text{ מופיע במדגם} \\ 0 & \text{אחרת} \end{cases}$$

דוג' אוילוסטריה השתתף האקטואציה:



נניח שהמדידה שלנו היא נקודות המישור
והמטרה היא להינתן x לנבא
את y - של.

יש סמני מחלקה היפותטית
שאפשר להגדיר

C_k - מחלקת פונקציות מדרגה k .

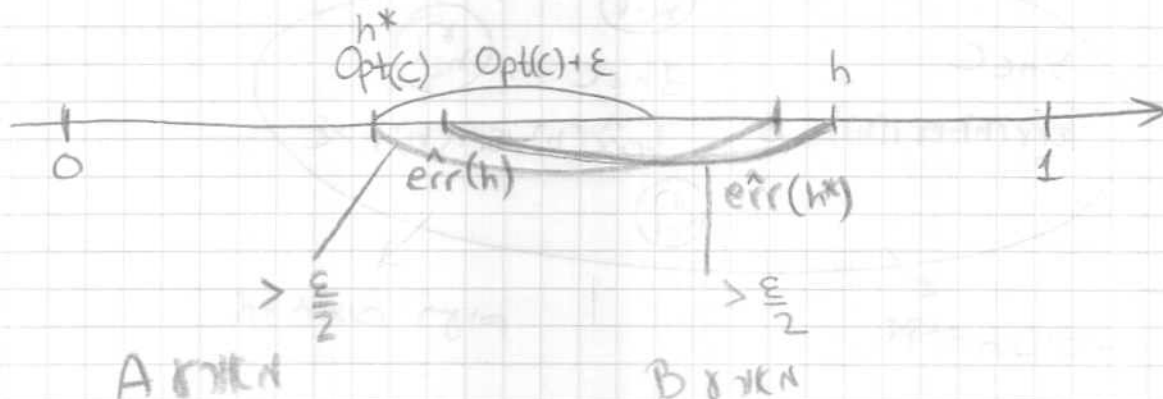
אפשר להחזיר קו ליניארי, ואפשר לקרב גזילה פונקציה
כיתות. הסופו של דבר אפשר להחזיר פולינום מדרגה m
והיא + חזק זכרם הנקודות.

על מחלקה C_k יש טעות אופיינית ϵ_k בהר שכל k -
היה ϵ_k קטן וחסר $\epsilon_m = 0$.

אם הם לא, זו כנראה לא התחלה טובה ר"ל לבחור
היפותזה m - C_m . העניין הוא שאנחנו מעצבים בטבר פשוט.
הסיבוכיות של C_k גדלה עם k . למרות $\epsilon_m = 0$,
 C_m היא פשוט מסובכת מדי.

עצמה: ה מחלקת מושגים סופית לאיזה - PAC.

הערה: האסטרטגיה היא לבחור היפותזה עם טעות אפסית
מינימלית.



$$P(A) = P(\hat{err}(h^*) > err(h^*) + \frac{\epsilon}{2}) \leq e^{-2m(\frac{\epsilon}{2})^2} = e^{-\frac{m\epsilon^2}{2}}$$

$$P(B) = P(\exists h: \hat{err}(h) < err(h) - \frac{\epsilon}{2}) \leq$$

$$\leq \sum_{\substack{h \in C \\ h \neq h^*}} P(\hat{err}(h) < err(h) - \frac{\epsilon}{2}) \leq (|C|-1) e^{-\frac{m\epsilon^2}{2}}$$

6

$$P(A \cup B) \leq P(A) + P(B) = |C| e^{-\frac{m \epsilon^2}{2}}$$

אם $P(A \cup B) < \delta$ אז $|C| e^{-\frac{m \epsilon^2}{2}} < \delta$ וזה אומר ש-
 $-\frac{m \epsilon^2}{2} < \ln \frac{\delta}{|C|}$

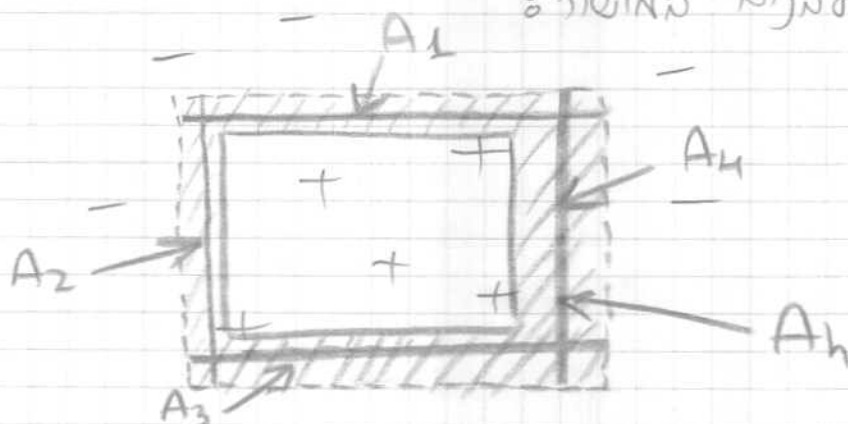
$$\Rightarrow m > \frac{2}{\epsilon^2} \ln \frac{|C|}{\delta}$$

$\Rightarrow$ אם נתנו $m > \frac{2}{\epsilon^2} \ln \frac{|C|}{\delta}$ ואנחנו בוחרים
 היפותזה שהכי טובה ביותר למצוא את $Opt(C)$
 אז מה שרצינו - שהסתברות קטנה δ היסטוריה
 של ההיפותזה קטנה ϵ - $Opt(C) + \epsilon$.



דאטה מתחלקת אושלים אינסופיות

דאטה של אלמנטים באינסוף



אם המדבן המקורי אנחנו לא רואים, אנחנו רואים את הנקודות
 הפנים ואם הנקודות בתוך המדבן רואים את המדבן.

האלמנטים: נבחר את ההיפותזה שטובה את המדבן
 המינימלי שתחס את ה- ϵ זרם. היסטוריה של ה- ϵ
 ההסתברות של הנקודות עיפוף במרווח הבין-ימי.

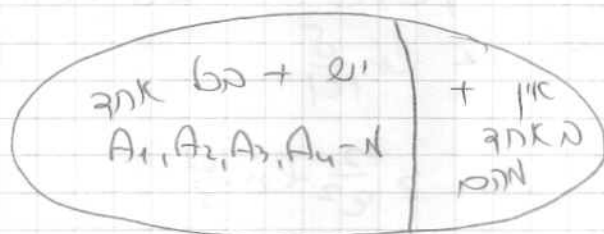
$$err(h) = P(x \in A_h) = \int_{A_h} \mathbb{1}(x) dx$$

בחרים את A_1, A_2, A_3, A_4 כך ש- $P(x \in A_i) = \frac{\epsilon}{4}$

אם יש + בט אחר A_i אז המסמן נקרא 'שגוי' ואלו
הפרישה.

אז אנתנו רוצים לחסום את ההסתברות שלא יופו + באחד מהם.

מרחב המצבים הוא



$$P(x_1, \dots, x_m \notin A_1) = \left(1 - \frac{\epsilon}{4}\right)^m$$

$$e^{-x} \geq 1-x \quad \text{C/ס} \quad e^{-x} \geq 1-x \quad \text{C/ס}$$

$$e^x \geq 1+x \quad \Leftarrow \quad e^x = 1+x + \sum_{i=2}^{\infty} \frac{x^i}{i!} \geq 1+x \quad \text{C/ס}$$

☺

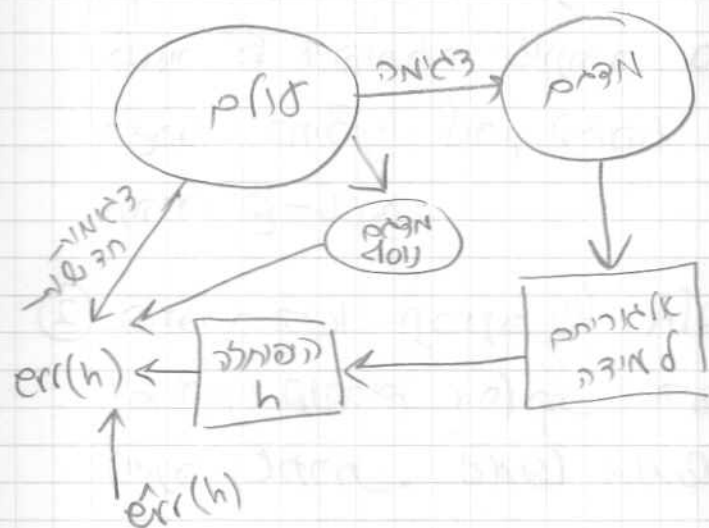
$$P(x_1, \dots, x_m \notin A_1) = \left(1 - \frac{\epsilon}{4}\right)^m \leq e^{-m\epsilon/4} \quad \text{מכאן ש-}$$

$$P(A_1 \cup \dots \cup A_4) < \delta \quad \text{ע-} \quad \text{ואנתנו רוצים שלא יהיה קטן } \frac{\delta}{4} \quad \text{כי אז וקרא}$$

$$m > \frac{4}{\epsilon} \ln \frac{4}{\delta} \quad \text{בטבלה צריך}$$

תלבוט:

שיעור 1 (I) הרצתו מוסת בלי אקטיות למידה



(II) side trip
validation
נתן מודל נוסף
והפונקציה באמצעות
משערים את
 $\hat{err}(h)$ ו- $err(h)$

שיעור 2 הרצתו מודל למידה פורמלי . PAC

PAC לא מנסה למצוא נוסף אלא רק במצב המקורי.

איתותם למידה PAC מודל והפונקציה יחד עם

המטרה של הביצועים של ההפונקציה ($err(h)$) במודלים

של דוגמ ופונקציה -1

הרצתו של מודל מודלים סוסות למידה PAC.

הראנו שמודל מודלים של מודלים מודלים (ה- R^2)

למידה PAC.

4

היום למידה PAC של מודל של הפונקציות אפילו.

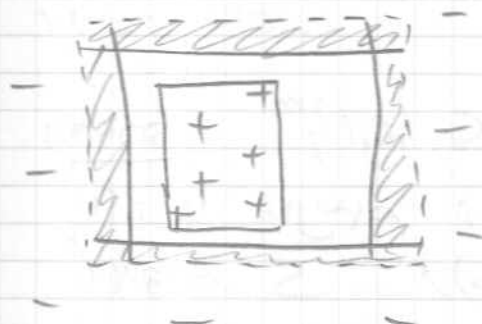
מה היה "הקסם" של המודלים שאפשר למידה PAC?

והרצתו איתותם איתותם של איתותם

רק שאלה המודלים "פונקציה" הם איתותם

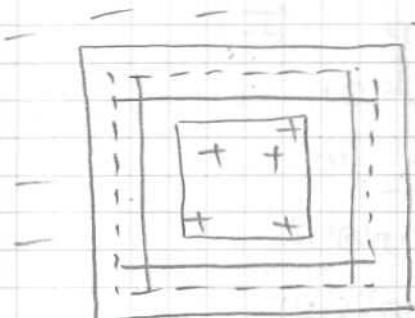
האיתותם מודלים איתותם של ההפונקציה

שנותר לפי האיתותם תהיה ע-אבה.



נשים ופרטים הסאבים:

① אם אנחנו נגדיר ϵ אז אבדע אזור של $\frac{\epsilon}{4}$ ונבדוק שהמדיה יפגע בה שמורת האזורים. אז ב היפותזה עקבית עם המדיה תהיה ϵ -אבה.



לומר ב היפותזה שתהיה בין הגבול החיצוני לבין הגבול הפנימי תהיה ϵ -אבה.

② התמונה היא תכונה של מלבנים.

אם היינו בוחרים מתחת היפותזה אחת, אזי התמונה הייתה אחרת. למשל המשלים היינו פתלים קו.



מסקנה: מצא סופי מאובל $m \geq \frac{4}{\epsilon} \ln \frac{8}{\epsilon}$ בהסתברות $1-\delta$ קלע ϵ כפי ש התנהגות של אינסוף היפותזות שמסכימות עם המדיה.

קובעו שמדיה האובל $\frac{4}{\epsilon} \ln \frac{8}{\epsilon}$ יודר ϵ -רשת זכור מתקת היפותזות של המלבנים.

מאנה: לאפין תכונות של מתקת של היפותזות וקלקה את אובל המדיה הנדרש כפי אזור ϵ -רשת.

לכל דרך חקר של התנהגות של מתקת ויפותזות C של מדיה סופי.

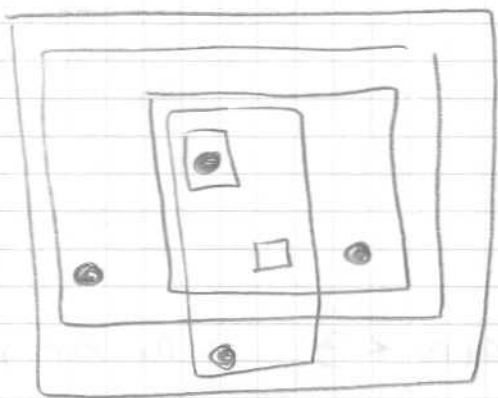
$$\Pi_C(S) = \{h(x_1), \dots, h(x_m) : h \in C\} \subseteq \{0,1\}^m$$

אזה ב התוצאה האפשרית שהיפותזות C יכולות להתחלק למדיה S $\Pi_C(S)$ (קלקה) $\Pi_C(S)$ dichotomies of S .

8

נשים $|\Pi_C(S)| \leq 2^m - 1$

המשפט: אם $|\Pi_C(S)| = 2^m$ נאמר ש- S מנותלת (shattered)
ו"י C . המושג אחרות, S מנותלת "ע"י C אם ההיפרתלות
ה- C יכולות לנתב כל תיבט אפלטו לנקודות ה- S .



מנותלת

ציון מה: מאבנים במישור:



עם מנותלת
(התיוג והמצג לא ייתנו
שהתקדם)

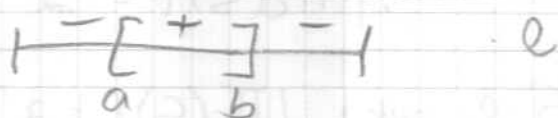
VC-dimension (Vapnik-Chervonenkis)

הגדרה: ממד VC של מחלקת היפותזות C הוא
ו"י $VCdim(C)$ הוא הגודל של קבוצת נקודות S
המקסימלית ביותר המנותלת "ע"י C . אם C יכולה לנתב
קבוצות קבוצות בגודל לא חסום של נקודות (אמר
ע- $VCdim(C) = \infty$

אבחנה: כפי שהראנו $VCdim(C) \geq d$ מספיק למצוא
קבוצה אחת של נקודות S בגודל d ק-ע- C מנותלת
א- S .

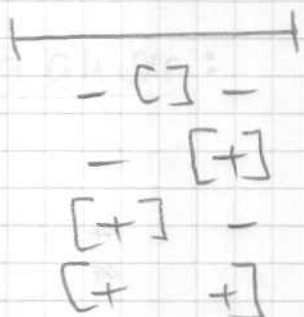
כפי שהראנו $VCdim(C) < d+1$ זניק להראות
ממצא בגודל $d+1$ עם מנותלת "ע"י C .

המרחב $[a, b]$ קטנים $\leq$ הישר הממשי $\mathbb{R}$



מסלול $VCdim = 2$

מכאן $VCdim \geq 2$ כי אפשר לנתק נקודה



למרחב $VCdim < 3$ כי בכל מקרה שלוש נקודות
 בו משלו מהצורה $+$ $-$ $+$ לא ניתן לנתק.

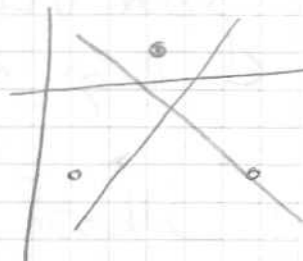
מרחב $VCdim = 4$ ראוי שיהיה אפשר לנתק
 ואילו אם יהיו נקודות ונקודות עם אותה

מ-4 נקודות מהכמה תהיה נקודה
 במקור של 4 נקודות אחת אחת
 או אפשר לנתק:



מרחב $VCdim = 3$ מרחב מרחבי

מרחב שלוש נקודות ונקודות:



אם אפשר להשתמש
 שאם אפשר לנתק 4 נקודות.
 4 נקודות הן באותה מהקוויים 3 יחד



ואם התוצאות
 המשותפים או אפשר
 לנתק.

(9)

איחוד בין מניה של קטעים של הישר

$$VCdim = \infty \quad \underbrace{[-] - [-] - [-] - \dots}_{\infty}$$

בהרבה מקרים $VCdim$ שווה למספר הפרמטרים של היפותזה באפדת, אך לא תמיד.

נניח שאנחנו יוצרים שרשרת היא סופי d , אז התיוג של הקובץ $d+1$ נקבע ע"י d הקובץ האחרון. ואז אם יש מספר הרבה יותר גדול $d-n$ של קובץ אז כוח התיוגים הם נקבעים ע"י אשהו שיש לנו. אז אינסופיות זה מראה לנו כמה אם המרחב היא סופי אז אפשר לאחד את המחקר.

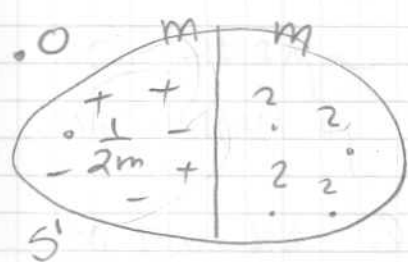
לענין: מחלקת מוסטים C אם $VCdim = \infty$ אז PAC לא חיצה.

הוכחה: נניח בשלילה ש- C חיצה PAC . יהי L אלמנטרית חיצה n - m מספר גדול מאד הנדרש וזו להיות ϵ דיוק $\epsilon = 0.1$ במיטתון $1-\delta = 0.9$. כיוון של- C חיצה VC אינסופי קיים מרחב S בגודל m שמכיל ע"י C .

נבדוק התפלגות על X להיות אחודה על S (הסתברות $\frac{1}{m}$ על $S \in X$) - 0 מחוץ ל- S .

א תיוג של S אפשרי (לפיה הנחה) ונחזי מושג מטה ע"י המרחב

$$P(C(x)=0) = \frac{1}{2} \quad \forall x \in S$$



אלמנטרית חיצה L צוים מרחב S' בגודל m לפי ϵ , סומך מתק S . $L \rightarrow S$ כואב על היות m נקודות מתק S ? על הקובץ מרחב S מתקיים:

$$\forall x_i \notin S' \quad P_{\substack{\text{בחירה} \\ \text{של } c}} (c(x_i) \neq h(x_i)) = \frac{1}{2}$$

עם היגדתי בחירה של c וכדי תואר ה- h .

$$\mathbb{E}_{\substack{\text{בחירה} \\ \text{של } c}} (\text{err}(h)) \geq \underbrace{m \cdot 0 \cdot \frac{1}{2m}}_{\text{בתוך } S'} + \underbrace{m \cdot \frac{1}{2} \cdot \frac{1}{2m}}_{\text{מחוץ ל-} S'} = \frac{1}{4}$$

$$\mathbb{E}_{\substack{\text{בחירה} \\ \text{של } S'}} \mathbb{E}_{\substack{\text{בחירה} \\ \text{של } c}} (\text{err}(h)) \geq \frac{1}{4} \quad \text{ולכן זה}$$

מצד שני, עם ההקטנה של L , δ מושג מטרה

$$\mathbb{E}_{\substack{\text{בחירה} \\ \text{של } S'}} (\text{err}(h)) \leq \underbrace{\varepsilon(1-\delta)}_{\substack{\text{בהסתקרות} \\ \text{השיאה קטנה} \\ \varepsilon - m}} + \underbrace{1 \cdot \delta}_{\substack{\text{בא מקרה} \\ \text{אחר השגאה} \\ \text{חסונה ע"י } L}} = 0.19$$

וזאת סתירה כי $0.19 < 0.25$ כי אכן

$$\mathbb{E}_{\substack{\text{בחירה} \\ \text{של } c}} \mathbb{E}_{\substack{\text{בחירה} \\ \text{של } S'}} = \mathbb{E}_{S'} \mathbb{E}_c \text{err}(h) = 0.19 < 0.25$$

(11)

מה אנחנו אומרים מזה? כל שהמימד קטן יותר
אם העצומות והאפיונים של ההיפותזה של האחרים
החצר קטנה יותר יהיום אחרים בטוחים יותר שהם
אם קרובה לעצומות האחרים של ההיפותזה.
כל שמחלקת המושגים קטנה יותר היא גמישה פחות אבל
מה שאנחנו כן מצליחים לעמוד הוא כנראה כן טוב.

כלומר מה שצוינו עליו הוקדו והקירוב ע"י פולינומים.
אם יש מן הקצוות והיחסים פולינים מצדד מ זה לא באמת
"איך אנחנו מבינים תכנים של הקצוות חדשות שיכולים.

10) 13.11.07
תאריך אחרון

בשיעור הקודם הוכחנו שאם $VCdim(C) = \infty$ אז C לא למידה
PAC. היום נראה את ההפך: אם $VCdim(C)$ סופי אז
 C למידה PAC.

נבחר S - $\Pi_C(S)$ הם כל התוצאות האפשריות לריבועות מתוך
 C נותנות למידה S .

נרצה לחסום את מספר הריבועים בהן מתקרת הריבועות - בעזרת
מימד VC השווה d נראה לתייג m נקודות.
(יציור פונקציות יצורים):

$$\Pi_C(m) = \max_S \{ |\Pi_C(S)| : |S| = m \}$$

$$\begin{cases} \phi_d(0) = \phi_0(m) = 1 \\ \phi_d(m) = \phi_d(m-1) + \phi_{d-1}(m-1) \end{cases}$$

נצייר את פונקציה הקומבינטורית

לענין 1: אם $VCdim(C) = d$ אז $\Pi_C(m) \leq \phi_d(m)$
אנטיאיציב לדבריה: צומח דבריה הקומבינטורית - d

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

S • $m-1$ נבחר S - $m-1$ נקודות אחת!
מהנחת האנדוקציה את $m-1$ הנקודות אפשר לתייג בכל
היותר $\phi_d(m-1)$ דרכים. כעת, יש אפשרות ללחוץ את התיוג
של הנקודה שמתנו רק אם לא השתמשנו בה d המימדים
של $m-1$ הנקודות האחרות אז מספר התיוגים הנוסף הוא
כל היותר $\phi_{d-1}(m-1)$ מה"כ לפי ההנחה לה $\phi_d(m)$.

$$\phi_d(m) = \sum_{i=0}^d \binom{m}{i} \quad \text{לענין 2}$$

הוכחה: באנדוקציה (תרגיל בית)

$$\phi_d(m) \leq \left(\frac{e m}{d}\right)^d = O(m^d) \quad \text{לענין 3} \quad d < m$$

בסעיף: אם מספר הנקודות קטן מהמימד אז יש 2^m אפשרויות
לתייג וזה אקספוננציאלי הכעם היא להשלים יותר נקודות

מהימנע את מספר האפשרויות הקיימות פולינומיאלית ולא אסימטרית.

הוכחה: $\left(\frac{d}{m}\right)^d \phi_d(m) = \left(\frac{d}{m}\right)^d \sum_{i=0}^d \binom{m}{i} \leq \sum_{i=0}^d \left(\frac{d}{m}\right)^i \binom{m}{i} \leq$

כי $\frac{d}{m} < 1$ ולכן $\left(\frac{d}{m}\right)^d \leq \left(\frac{d}{m}\right)^i$

$\leq \sum_{i=0}^m \left(\frac{d}{m}\right)^i \binom{m}{i} = \left(1 + \frac{d}{m}\right)^m \leq e^d$

$\Rightarrow \phi_d(m) \leq \left(\frac{m}{d}\right)^d e^d = \left(\frac{me}{d}\right)^d$

☺

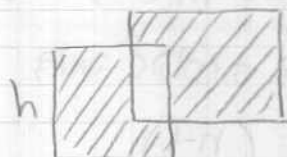
סיכום:
$$\Pi_C(m) = \begin{cases} 2^m & m \leq d \\ \leq \left(\frac{em}{d}\right)^d & m > d \end{cases}$$

100) $C \in \mathcal{C}$ ממש המטרה! $h \in \mathcal{C}$ ממש בלבד.

הגדרה: $C \Delta h = h \Delta C = \{x : C(x) \neq h(x)\} \subseteq X$

זהו אזור בו C ו- h אינם הסכימו.

לנד:



הגדרה: $\Delta(C) = \{h \Delta C : h \in \mathcal{C}\} \subseteq 2^X$ אוסף אזורי שטח.

(= אזורי אי הסכמה). (ניתן להסתכל על $\Delta(C)$ כאוסף היפותזות)

נניח שאם $x \in h \Delta C$ אז התוצאה של x היא 1 ואחרת התוצאה היא 0.

טענה: $VCdim(C) = VCdim(\Delta(C))$

הוכחה: נניח $S \subseteq X$ מתקיים $|\Pi_{\Delta(C)}(S)| = |\Pi_C(S)|$

כי $h \Delta C$ הוא XOR, ולכן אם h_1, h_2 מתאימים

אם S מאופן שונה את $h_1 \Delta C$ שונה $h_2 \Delta C$ וכן

$h_1 \Delta C \neq h_2 \Delta C$ כי $h_1 \neq h_2$ זכיכות אחרת שונות על S .

☺

(11)

הגדרה: (סמן) $\Delta_\varepsilon(C) = \{h \in C : \text{err}(h) \geq \varepsilon\}$

זהו אוסף אזורי השגיאה הגדולים מ- ε (אם $\emptyset$)

הגדרה: לכל $0 < \varepsilon$ נגד שמות S יוצר ε -רשת עבור $\Delta_\varepsilon(C)$

אם S "פוגע" באזור $r \in \Delta_\varepsilon(C)$, נומר לכל

מתקיים $S \cap r \neq \emptyset$.

נשים לב שאם S יוצר ε -רשת עבור $\Delta_\varepsilon(C)$ אזי S היפותזה

קונסיסטנטית לכל S היא ε -טובה.

ההסתברות שמות S באחד מ- m "פוגע" באזור גודל

ε - m חסומה $\varepsilon^m (1-\varepsilon)^m$ (כל נקודה בהסתברות לכל

כל תפגע האזור שיאף הוא $1-\varepsilon$ ויש m נקודות בלתי

תלויות).

ההסתברות שקיים אזור שא לא נפגע ε^m חסומה ε^m

$(1-\varepsilon)^m$. אבל $|\Delta_\varepsilon(C)|$ לא חסום באופן כללי.

(במשחק עם המלכנים הצלחנו לאחסן אזורי השגיאה חכמה).

שיטת שני המצבים: נגד שמות S_1 באחד m .

- מאיזה A (רע) - קיים אזור שיאף $r \in \Delta_\varepsilon(C)$ (והיפותזה

מתאימה ב- C) שלא נפגע ε^m . S_1 .

נצדף לחסום את ההסתברות של A .

נגד שמות נוסף S_2 באחד m . אז תוחלת מספר

הפגעות של S_2 ב- r היא $\mathbb{E}(|S_2 \cap r|) = m\varepsilon$

$$(*) \quad P(|S_2 \cap r| \geq \frac{m\varepsilon}{2}) = 1 - P(|S_2 \cap r| < \frac{m\varepsilon}{2}) \geq \frac{1}{2}$$

← עבור $m \geq \frac{8}{\varepsilon}$ (תנאי הור)

- מאיזה B - קרה A אם $|S_2 \cap r| \geq \frac{m\varepsilon}{2}$. (אם $(*)$)

$$P(B) = P(A) \cdot P(B|A) \geq \frac{1}{2} P(A)$$

$$P(A) \leq 2P(B) \Leftarrow$$

אנחנו רוצים לחסום את $P(A) \Leftarrow$ מספר אחר $P(B)$.

ננסה את B מחדש: (נחלק את $S = S_1 \cup S_2$ ל- $2m$ תתי-קבוצות מאונות B כאשר שקיים עבור r רק ϵ -
 $r \cap S_1 = \emptyset$ אם $|r \cap S| \geq \frac{\epsilon m}{2}$ -! רואים כי עבור r יש
 של היותה של r ששאינה יכולה להיות ϵ

נקבע $(S) \in \Pi_{\Delta(\epsilon)}(S)$ שמקיים $|r| \geq \frac{\epsilon m}{2}$ ונחסום את
 ההסתברות לחלק את S לשתי קבוצות S_1, S_2 רק
 ש- $r \cap S_i = \emptyset$.

יש $2m$ כדורים שמותכם $\ell \geq \frac{\epsilon m}{2}$ אדומים.
 מה ההסתברות שבתחילה שלהם לשתי קבוצות כל
 האדומים יהיו בקבוצה אחת?

ניסוח שקוף: קחם לחלק $2m$ כדורים לשתי קבוצות ואז
 לציבם באופן מקרי ℓ כדורים. מה הסיכוי שהם חלם
 בקבוצה אחת?

$$P(r \cap S_i = \emptyset) = \frac{\binom{m}{\ell}}{\binom{2m}{\ell}} = \frac{\prod_{i=1}^{\ell-1} \frac{m-i}{2m-i}}{\prod_{i=1}^{\ell-1} 1} \leq \left(\frac{1}{2}\right)^\ell \leq 2^{-\frac{\epsilon m}{2}}$$

$$\Rightarrow P(B) \leq \underbrace{|\Pi_{\Delta(\epsilon)}(S)|}_{\text{חומר ראשוני}} 2^{-\frac{\epsilon m}{2}} \leq$$

מספר אזורי השטח
 של ϵ השונים בהסתברות

$$\leq |\Pi_{\Delta(\epsilon)}(S)| 2^{-\frac{\epsilon m}{2}} = |\Pi_C(S)| 2^{-\frac{\epsilon m}{2}} \leq$$

$$\leq \phi_d(2m) 2^{-\frac{\epsilon m}{2}} = \left(\frac{2em}{d}\right)^d 2^{-\epsilon m}$$

$$\text{באור סיכוי } P(A) \leq 2 P(B) \leq 2 \left(\frac{2em}{d}\right)^d 2^{-\epsilon m}$$

$$P(A) \leq \delta \quad \text{אז } m \geq K \left(\frac{1}{\epsilon} \log \frac{1}{\delta} + \frac{d}{\epsilon} \log \frac{1}{\epsilon} \right)$$

משפט: תהי C מחלקת היפותזות בעלת מידת VC d . יהי

L אלמנטים בלבדו שמקבלים אדום מתוך ϵ היפותזת מטרה

$(C \in C)$ באופן m ומחבר היפותזה $h \in C$ שעקבות

אם המצגם. אזי L הוא אדמיתם PAC עבור C ראש

$$m \geq K \left(\frac{1}{\epsilon} \log \frac{1}{\delta} + \frac{d}{\epsilon} \log \frac{1}{\epsilon} \right)$$

(12) 20.11.04
ת' טל' אולמבר

(הזכור) ההוכחה ל- $\Pi_d(m) \leq \Phi_d(m)$ יש m נקודות ונוציא אחת

$m-1$

עם הנחת האנדוקציה ל- $(m-1)$ יש לכל היותר $\Phi_d(m-1)$ תוצים. השאלה היא מה קורה עקרונית שמוצאנו, אם יש ה- $(m-1)$ נקודות שקיבלו את כל התוצים האפשריים אז אנחנו שלטים על התוצה של הנקודה. אחרי אפשר להסתכל על $(m-1)$ כמחלקת היפותזות. זה מראה $d-1$ וכל עם הנחת האנדוקציה לה לכל היותר $\Phi_{d-1}(m-1)$ תוצים אפשריים ומה' $\Pi_d(m) \leq \Phi_d(m-1) + \Phi_{d-1}(m-1) = \Phi_d(m)$

Generalization Bounds

שתי פנים של תסם הופציונג.

$$P(|err(h) - \hat{err}(h)| \geq \varepsilon) \leq 2e^{-2m\varepsilon^2} \leq \delta$$

(1) sample complexity סבוכות נדגים

נתונים δ, ε . אחר ה- m המינימלי הנדרש כדי להשיג

$$m \geq \frac{1}{\varepsilon^2} \ln \frac{2}{\delta}$$

(2) generalization bound תסם הכלל

נתון m (המציא שקיבלנו) ונתון δ . אחר $err(h)$?

בהסתברות גדולה $1-\delta$

$$err(h) \leq \hat{err}(h) + \sqrt{\frac{\ln \frac{2}{\delta}}{2m}}$$

תסם הקריצינליות

נתונת מחלקת מושגים C באפס סופי $|C|$.

$$P(\exists h: \text{err}(h) - \hat{\text{err}}(h) \geq \varepsilon) \leq |C| e^{-2m\varepsilon^2} \leq \delta$$

חסם האיחוד

הסתברות אפולה $1-\delta$

$$\text{err}(h) \leq \hat{\text{err}}(h) + \sqrt{\frac{\ln \frac{|C|}{\delta}}{2m}}$$

Occam razor (PAC-Bayesian approach)

לפי דפוס היפותזה $h \in C$ פריור $p(h)$ (לפי חזק סוג של האמונה שלנו במשלבו דפני שלם פנו א'זה מ'זה ע'יו) $P(h)$ רק $-$ $\sum_{h \in C} P(h) = 1$ וכל h $P(h) \geq 0$

$$P(\text{err}(h) - \hat{\text{err}}(h) \geq \varepsilon) \leq e^{-2m\varepsilon^2} \leq p(h) \cdot \delta - \text{ל} \quad (*)$$

ה'אפ'ן

$$P(\exists h: \text{err}(h) - \hat{\text{err}}(h) \geq \varepsilon) \leq \sum_{h \in C} P(h) \cdot \delta = \delta \quad \text{אז}$$

$$\varepsilon \geq \sqrt{\frac{\ln \frac{1}{p(h)} + \ln \frac{1}{\delta}}{2m}} \quad \text{ל} \quad (*) \quad \text{והסתברות}$$

אפולה $1-\delta$

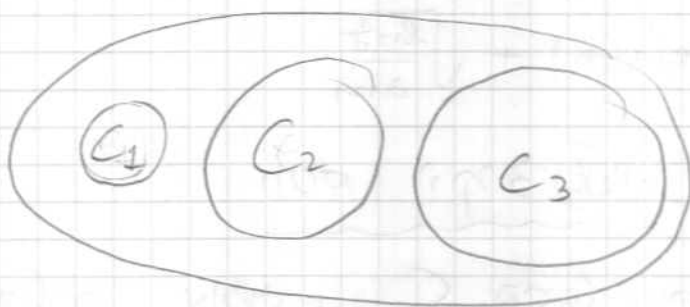
$$\text{err}(h) \leq \hat{\text{err}}(h) + \sqrt{\frac{\ln \frac{1}{p(h)} + \ln \frac{1}{\delta}}{2m}}$$

אח'יקה $p(h) = \frac{1}{|C|}$ אזי נקדם חצה את חסם הקרובין א'יו

$$\text{err}(h) \leq \hat{\text{err}}(h) + \sqrt{\frac{\ln |C| + \ln \frac{1}{\delta}}{2m}}$$

חשיו, למה חשיו א δ

נ'יו שחלק היפותזות C מוכנה K מ'חלקה מ'חלקה ש'יו.



13

נניח $h \in C_j$ כך $P(h) = \frac{1}{K} \cdot \frac{1}{|C_j|}$
 אם $h \in C_j$ אז $h \in C$ בהסתברות $1-\delta$

$$\text{err}(h) \leq \text{err}(h) + \sqrt{\frac{\ln |C_j| + \ln K + \ln \frac{1}{\delta}}{m}}$$

אם יש אישה trade-off בין $\text{err}(h)$ לבין
 הסימיות של החלקה שלנו. אופטימלי היינו נוצרים
 דמיון עם $\text{err}(h)$ זה לא הסימיות.

נקודה חשובה: $\rho(h)$ (קבץ) דפני שלטאות את המצבים
 (אחרת האנליזה פה לא עובדת)

דבור מחלקת היפותזות בשל d VC-dim

$$\text{err}(h) \leq \text{err}(\hat{h}) + \sqrt{\frac{32(d \cdot \ln \frac{em}{d} + \ln \frac{4}{\delta})}{m}}$$

(התחל בזה) ניתן גם עדיין דאיתוך בן מניה של
 מחלקת עם VC-dim סופי (זה תחיל בזה)
 זה דאיתוך לא בן מניה (אבל זה לא נחשב עדיין)

למוד עם מפרקים עקרוניים

הקצה: יש נקודות במישור מתחלקות ב- / +
 מחלקת היפותזות: ישנים



יש אלגוריתם פופולרי בשם Perceptron
 ואם לא נמצא אחרון

SVM - support Vector Machines

יש קבוצת אימון $\{(\vec{x}_1, y_1), \dots, (\vec{x}_m, y_m)\}$ $\vec{x}_i \in \mathbb{R}^n$

הפרדה ליניארית פירושה קבוצת $\vec{w} \in \mathbb{R}^n$ ו- $b \in \mathbb{R}$!

כך שאם $\vec{w}^T \cdot \vec{x} - b > 0$ אז התיוג הוא "+" ואם $\vec{w}^T \cdot \vec{x} - b < 0$

אם התווית של $\vec{x}$ הוא "-"

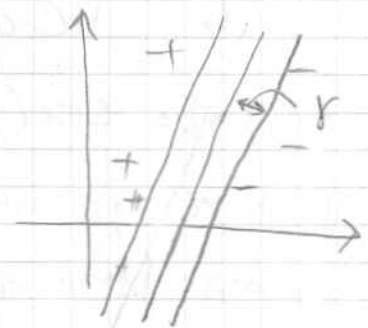
אם $y_i \in \{\pm 1\}$ אז אם $y_i(\vec{w} \cdot \vec{x} - b) > 0$ אז התווית נכון ואם $y_i(\vec{w} \cdot \vec{x} - b) < 0$ אז התווית לא נכון.
ממד VC של המרחב $\mathbb{R}^n$ הינו $n+1$.
זה אומר ש-

$$\text{err}(h) \leq \text{err}(h) + \sqrt{\frac{32(n+1) \ln \frac{em}{n+1} + \ln \frac{4}{\delta}}{m}}$$

זה טוב עבור n קטן אבל לא טוב עבור n גדול.
באמצעות משתנה יותר טוב,

מרחק של וקטור $\vec{x}$ מאל-מישור $(\vec{w}, b)$ הינו
(גודל margin) (נורמליזציה) של הפרדה

$$\gamma = \min_i \left\{ \frac{y_i(\vec{w} \cdot \vec{x}_i - b)}{\|\vec{w}\|} \right\}$$



טאבו היפרפלין
הוא אופן משיק
אלאגוריתם
שפותר את
שאלת האם
ייתכן שיש
קצוות.

מרחק C_r - מרחק היפרפלין
של המרחב $\mathbb{R}^n$ עם נורמליזציה
ו ונניח שכל הוקטורים חסומים
ברדיוס R

נומף C_r מממד VC של C_r הוא

$$\text{VC-dim}(C_r) = \min \left\{ \frac{R^2}{r^2}, n \right\} + 1$$

נשים $\heartsuit$ שאם $\vec{w} \cdot \vec{x} - b > 0$ אז $\vec{x}$ הוא $\heartsuit$
אם $\vec{w} \cdot \vec{x} - b < 0$ אז $\vec{x}$ הוא $\clubsuit$
אם $|\vec{w} \cdot \vec{x} - b| = 1$ אז $\vec{x}$ הוא $\heartsuit$ או $\clubsuit$

$$\gamma = \frac{|\vec{w} \cdot \vec{x} - b|}{\|\vec{w}\|} = \frac{1}{\|\vec{w}\|} = \frac{1}{\sqrt{\vec{w} \cdot \vec{w}}}$$

אנחנו רוצים למקסם את γ .

נשים $\heartsuit$ ש-

$$\arg \max_{\vec{w}} \frac{1}{\sqrt{\vec{w} \cdot \vec{w}}} = \arg \min_{\vec{w}} \frac{1}{2} \sqrt{\vec{w} \cdot \vec{w}}$$

קריטריון מינימום אופטימיזציה

$$\min \frac{1}{2} \vec{w} \cdot \vec{w} \quad \text{s.t.} \quad y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad i=1, \dots, n$$

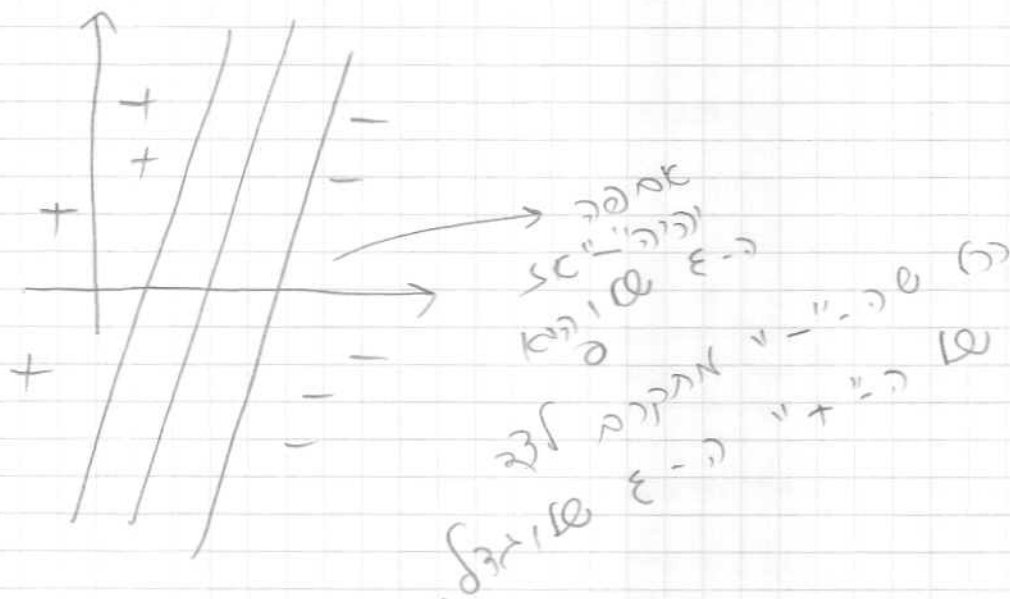
(14)

quadratic linear programming זה
 הניסוח מניח הפרדה מושלמת. כלומר, δ ו- ϵ הם
 מספרים חיוביים

$$\min \frac{1}{2} W W + \nu \sum_{i=1}^m \epsilon_i \quad y_i (W x_i - b) \geq 1 - \epsilon_i \quad \epsilon_i \geq 0$$

ν זה ה-tradeoff הקבוע מראש.

! $\text{err}(h) = |\{i : \epsilon_i > 1\}|$ היא מספר הטעויות
 $\sum_{i=1}^m \epsilon_i = \|\vec{\epsilon}\|_1 \geq \text{err}(h)$



(16) 27.11.04
→ 13411 '0N' A

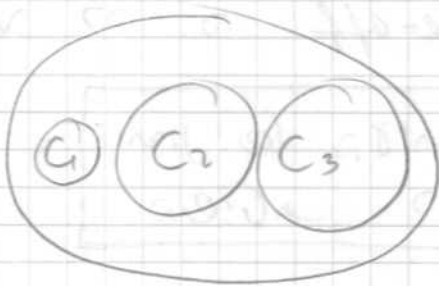
תלמוד

- חסם הקרדן ע"ה:

$$\text{err}(h) \leq \text{err}^*(h) + \sqrt{\frac{\ln |C| + \ln \frac{1}{\delta}}{2m}}$$

Occam's Razor -

$$\text{err}(h) \leq \hat{\text{err}}(h) + \sqrt{\frac{\ln \frac{1}{\delta(n)} + \ln \frac{1}{\delta}}{2m}}$$



$$\text{err}(h \in C_j) \leq \hat{\text{err}}(h) + \sqrt{\frac{\ln |C_j| + \ln K + \ln \frac{1}{\delta}}{2m}}$$

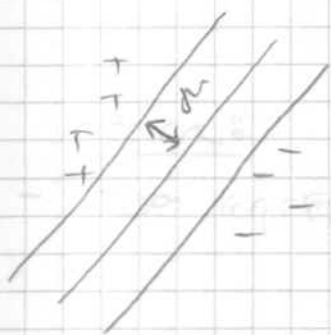
SVM -

$$\gamma = \min_i \left\{ \frac{y_i (w^T x_i - b)}{\|w\|} \right\}$$

$$\min_{w, b, \xi} \frac{1}{2} w \cdot w + \gamma \sum_{i=1}^n \xi_i \quad \text{s.t.} \quad y_i (w^T x_i - b) \geq 1 - \xi_i \quad \xi_i \geq 0$$

$$\text{err}(h) \leq \hat{\text{err}}(h) + \sqrt{\frac{32(d \ln \frac{em}{d} + \ln \frac{4}{\delta})}{m}}$$

נניח שאנחנו פועלים מחדש ויש לנו שני אפשרויות: עיבודים
ורביעים. נניח שאנחנו בעלי עיבודים אחדים, את העיבוד
הכי טוב וזה עמנו בעלות כבידוע ומצאנו את הכי טוב
הכי טוב. ואז נסתכל על העיבוד והכבידוע הכי טובים ונבחר
את זה שהיה הקטנה יותר. זה לא טוב!
זה לא כמו עם אחד אחד מתקן שחורבתיים מעיבודים
ואם מריבועים - זינג עבוסל את אהל (בן 2=א)
יו למשה אנחנו פועלים מתקנה שהיא איתנו
הריבועים והעיבודים



יש נקודות במישור ואנחנו מחפשים
 לאנחנו מפריצים אותן עם שוליים $\pm$.
 אנחנו נרצה לחפש אפיק עם שוליים כמה
 שיותר גדולים וטעות אמפירית יותר קטנה.
 כמה אנחנו חוצים שוליים גדולים? - כי אולי

מימד VC יותר קטן (אם הוקדקו המימד R
 אם המימד VC הוא $\pm 1 + \min\{\frac{R^2}{\gamma^2}, \gamma\}$ אולי אולי
 לאנחנו עוזרים טוב יותר.
 נראה trade-off בין השניים לבין הטעות.

פיתרון של בעיה אופטימיזציה עם אינדיקטורים
 בשיטה כופלי לזרנס

אנחנו מחפשים $\min_x f(x)$ רק $-e$
 כאשר $f(x): \mathbb{R}^n \rightarrow \mathbb{R}$!
 $h(x): \mathbb{R}^n \rightarrow \mathbb{R}^k$ $h(x) = 0$

$$\begin{pmatrix} h_1(x_1, \dots, x_n) \\ \vdots \\ h_k(x_1, \dots, x_n) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

ונניח $-e$ $\vec{\lambda} = (\lambda_1, \dots, \lambda_k)$

לצורך אחר הדיפרנציאל של (*) :

$$\mathcal{L}(x, \lambda) = f(x) + \vec{\lambda} \cdot \vec{h}(x) = f(x) + \sum_{i=1}^k \lambda_i h_i(x)$$

משפט כופלי לזרנס : יהי x^* נקודת מינימום לוקאלית

של f תחת אינדיקטורים $h(x) = 0$ ונניח שהגרדיאנטים

של האינדיקטורים $\nabla h_1(x^*), \dots, \nabla h_k(x^*)$ בת"ס

אזי קיים וקטור יחיד $\lambda^* = (\lambda_1^*, \dots, \lambda_k^*)$ הוקרא וקטור

כופלי לזרנס רק $-e$

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i=1}^k \lambda_i^* \nabla h_i(x^*) = \vec{0}$$

אם בנוסף f !- h חזירו פעמיים בהצטרף לטוב

$$\forall v \in V(x^*) \quad v^T \underbrace{\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*)}_{\text{הסימטר}} v \geq 0$$

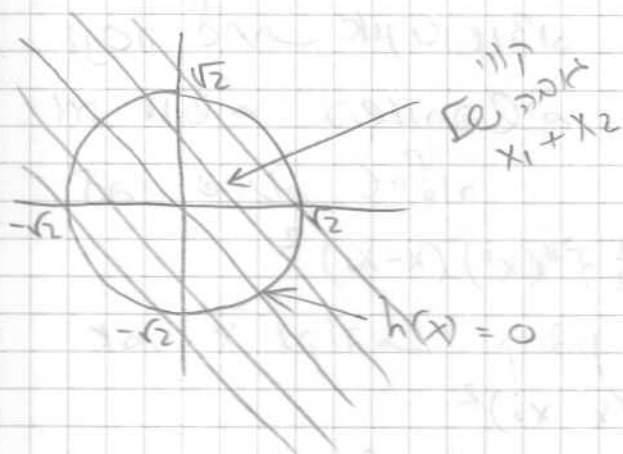
(17) $V(x^*) = \{v: \nabla h_i(x^*)v = 0 \quad i=1, \dots, k\}$ כאשר
 הוא תת-מרחב של המרחב $\mathbb{R}^n$ וזוהי קבוצת הווקטורים
 (first order feasible variations) δ שלה

מטריצה סימטרית M היא positive semi-definite
 אם B הערכים העצמיים שלה אינם שליליים $\delta \geq 0$.

תכונה: אם $f: \mathbb{R}^n \rightarrow \mathbb{R}$ אז

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)^T$$

$$\min (x_1 + x_2) \quad \text{s.t.} \quad x_1^2 + x_2^2 = 2$$
 דוגמה



אם δ הוא וקטור
 אז $\delta^T \nabla f(x^*) = 0$
 כלומר δ הוא וקטור
 המשיק למישור $x_1 + x_2 = c$
 הנורמל $\nabla h(x^*)$ הוא
 וקטור המשיק למישור $x_1 + x_2 = c$

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^k \lambda_i h_i(x) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2)$$

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) + \sum \lambda_i \nabla h_i(x) = \left(\frac{\partial \mathcal{L}}{\partial x_1}, \frac{\partial \mathcal{L}}{\partial x_2} \right)^T = (0, 0)^T$$

$$\Rightarrow \begin{pmatrix} 1 + 2\lambda x_1 \\ 1 + 2\lambda x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{matrix} x_1 = -\frac{1}{2\lambda} \\ x_2 = -\frac{1}{2\lambda} \end{matrix}$$

$$2 = x_1^2 + x_2^2 = \frac{1}{4\lambda^2} + \frac{1}{4\lambda^2} = \frac{1}{2\lambda^2} \Rightarrow \lambda = \pm \frac{1}{2}$$

$$\Rightarrow (x_1, x_2) = (1, 1), (-1, -1)$$

כלומר נקודות אלה הן נקודות קיצון

$$\nabla_{xx}^2 \mathcal{L}(x) = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial x_1^2} & \frac{\partial^2 \mathcal{L}}{\partial x_1 \partial x_2} \\ \frac{\partial^2 \mathcal{L}}{\partial x_2 \partial x_1} & \frac{\partial^2 \mathcal{L}}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Rightarrow \nabla_{xx}^2 \mathcal{L}(x, \lambda) = \nabla_{xx}^2 f(x) + \lambda \nabla_{xx}^2 h(x) = \begin{pmatrix} 2\lambda & 0 \\ 0 & 2\lambda \end{pmatrix}$$

positive semi-definite מצבון חצי חיובי

$\lambda = \frac{1}{2} \Leftrightarrow$ חצי מה שיש

$$(-1, -1) - \lambda = \frac{1}{2} \rightarrow \text{לא, נחמץ סומ' נמוך} \Leftarrow$$
$$(1, 1) - \lambda = -\frac{1}{2} \Rightarrow \text{לא נקודה קיצונית}$$

משה דוד אינטראקציה עם הילדה קמיל:

מחזרים משתמשים מנימוס של פונקציה ראשית.

דעם פילן שטעלן

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2$$

$$p^3 \mid p \rightarrow 3 \mid p \mid x \approx k$$

$$f(x) = f(x_0) + \frac{1}{2} f''(x_1) (x - x_0)^2$$

16. $f''(x)$ is not a polynomial

• $PIN'O, PN$

ב- $\mathbb{R}^n$ נמצא נקודה x ויש לה ויכוח

$$\Delta x^T \nabla_{xx}^2 f(x') \Delta x > 0 \quad \text{r/c}$$

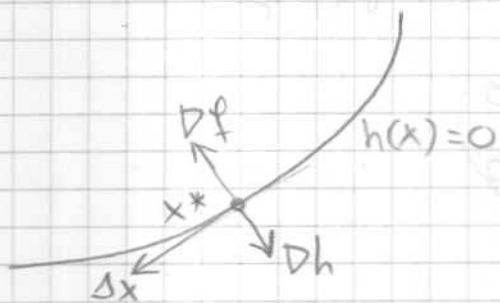
$$(\Delta x_1 \dots \Delta x_n) \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} = \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} \Delta x_i \Delta x_j$$

5. הביטוי מתאר $f''(x)(x-x_0)^2$ במידה דבר - נא

Conductivity

ווייזט ער x^* אן מיט אים זיך אים

$$h_i(x^* + \Delta x) \approx \underbrace{h_i(x^*)}_{=0} + \underbrace{\nabla h_i(x^*) \Delta x}_{\text{yik'os}} = 0$$



$$(19) (*) \quad \nabla h_i(x^*) \Delta x = 0 \quad i = 1, \dots, k \quad \Delta x \Leftarrow$$

(feasible variation) Δx - כיוון השתנות היות Δx (feasible variation)
 אם x^* מינימום מקומי של f ושל h_i בכיוונים האלה
 הסובבים Δx אז $\Delta x^T \nabla f(x^*) = 0$

$$f(x^*) \leq f(x^* + \Delta x) = f(x^*) + \underbrace{\nabla f(x^*) \Delta x}_{=0} + \Delta x^T \nabla_{xx}^2 f(x^*) \Delta x$$

נראה, $\nabla f(x^*) \in \text{Span}\{\nabla h_1(x^*), \dots, \nabla h_k(x^*)\}$ -
 קיים λ^* כך $\nabla f(x^*) + \sum_{i=1}^k \lambda_i^* \nabla h_i(x^*) = 0$

Karush - Kuhn - Tucker אי שוויונים והתא

$$\min f(x) \text{ s.t. } h(x) = 0, g(x) \leq 0 \quad (*)$$

כאן $g: \mathbb{R}^n \rightarrow \mathbb{R}^s, h: \mathbb{R}^n \rightarrow \mathbb{R}^k, f: \mathbb{R}^n \rightarrow \mathbb{R}$

$$\bar{g}(x) = \begin{pmatrix} g_1(x_1, \dots, x_n) \\ \vdots \\ g_s(x_1, \dots, x_n) \end{pmatrix} \leq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

השוואה לפי קאלברצ'ו

יציר $i(x)$ -

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^k \lambda_i h_i(x) + \sum_{j=1}^s \mu_j g_j(x)$$

נקודה x_0 היא נקודה ראשית אם הקבוצה
 $\{ \nabla h_j(x_0) \}_{j=1}^k$ היא בת"ס.

כל נקודה היותה מהשגה (בלא מקיימת) את
 (אנאלוגים) נצדי קבוצת אי שוויונים בעי"ה

$$A(x) = \{ j : g_j(x) = 0 \}$$

Karush - Kuhn - Tucker תנאי x^* מינימום

פונקציה $\mathcal{L}$ (כאשר f, h_i, g_j עכ"ל) -
 מרבית. ונראה x^* היא נקודה ראשית.

$\mu^* = (\mu_1, \dots, \mu_S)$ כוונת המכנים הנפלאים והחידים
 $\lambda^* = (\lambda_1, \dots, \lambda_k)$ - ע"י

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0$$

$$\mu_j \geq 0 \quad j = 1, \dots, S$$

$$\mu_j = 0 \quad \forall j \notin A(x^*) \iff \mu^{*T} g(x^*) + \sum_{j=1}^S \mu_j g_j(x^*) = 0$$

אם נתונים f, h, g פונקציות קונקסיות, אזי

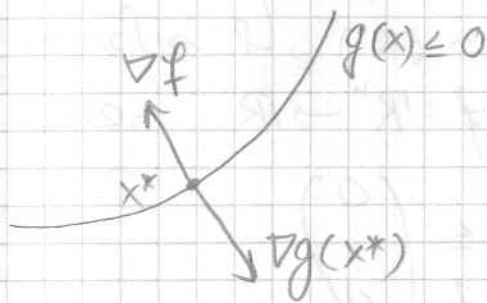
$$v^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*) v \geq 0$$

ובסדר $v \in \mathbb{R}^n$ המקיים

$$\nabla h_i(x^*) v = 0 \quad i = 1, \dots, k$$

$$\nabla g_i(x^*) v = 0 \quad \forall j \in A(x^*)$$

$$\mu_j^* \geq 0$$



$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum \lambda_i h_i(x) + \sum \mu_j g_j(x)$$

אם $g_j(x^*) < 0$ אז $\mu_j = 0$!
 חלקיק ונוסף להיותם מסומן

$$(\mu_j = 0 \quad \text{אם})$$

אם $g_j(x^*) = 0$ אז μ_j יכול להיות שווה ל-0 או לא-0.
 אם $\mu_j > 0$ אז $\nabla g_j(x^*) = 0$ (אם לא אז לא יהיה נוקטת קיצון).

19 4.12.07
תרגיל 8' סימולטן

דוגמה אופטימיזציה עם אינאליזים

$$\min \frac{1}{2}(x_1^2 + x_2^2) \quad \text{s.t.} \quad x_1 + x_2 \leq -2$$

נתון את הפונקציה של המטרה:

$$\mathcal{L}(x, \mu) = \frac{1}{2}(x_1^2 + x_2^2) + \mu(x_1 + x_2 + 2)$$

$$\nabla_x \mathcal{L}(x, \mu) = \begin{pmatrix} \frac{\partial \mathcal{L}}{\partial x_1} \\ \frac{\partial \mathcal{L}}{\partial x_2} \end{pmatrix} = \begin{pmatrix} x_1 + \mu \\ x_2 + \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{matrix} x_1 = -\mu \\ x_2 = -\mu \end{matrix}$$

$$\left. \begin{matrix} -2\mu = -2 \\ \mu \geq 0 \end{matrix} \right\} \Rightarrow \mu = 1 \Rightarrow x_1^* = x_2^* = 1$$

אם האופטימום היה $x_1 + x_2 \leq 2$ היינו צריכים

$$\begin{cases} -2\mu = 2 \\ \mu \geq 0 \end{cases} \quad \text{דבר זה אי אפשר}$$

ואם לא יכלו להיות דבריהם $\mu = 0$ ואז $x_1^* = x_2^* = 0$ והפונקציה היא

$$\left(\begin{matrix} \text{כאן נקבע קיצון} \\ g(x) = 0 \text{ או } \mu = 0 \end{matrix} \right)$$

דוגמה

נתון את הפונקציה

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_i \lambda_i h_i(x) + \sum_j \mu_j g_j(x)$$

$$\theta(\lambda, \mu) = \min_x \mathcal{L}(x, \lambda, \mu) \quad \text{נתון}$$

נצטרך את הפונקציה הדואלית

$$\max_{\lambda, \mu} \theta(\lambda, \mu) \quad \text{s.t.} \quad \mu \geq 0$$

נתון את הפונקציה $\mathcal{L}(x, \lambda, \mu)$ ב- x בלבד

$$\theta(\lambda, \mu) = \min_{x'} \mathcal{L}(x', \lambda, \mu) \leq \mathcal{L}(x, \lambda, \mu) =$$

$$= \underbrace{f(x) + \sum_i \lambda_i h_i(x)}_{\substack{=0 \\ \text{הכנסת התנאים}}} + \underbrace{\sum_j \mu_j g_j(x)}_{\substack{\leq 0 \\ \text{אילוץ} \\ \text{כיוונית}}} \leq f(x)$$

אסטרטגיה הכללית

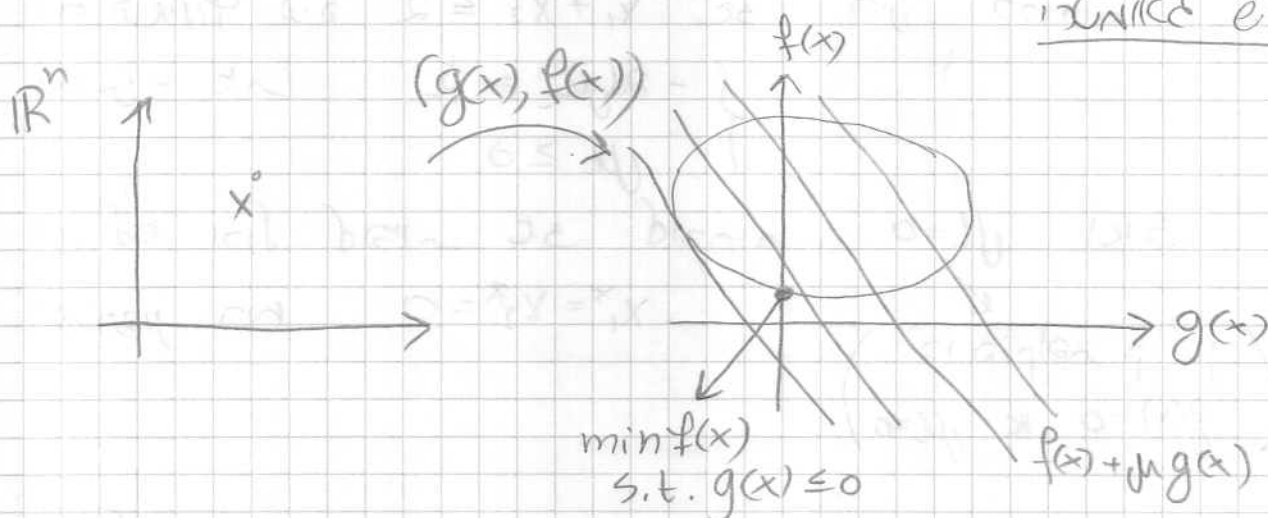
$$\max_{\substack{\lambda, \mu \\ \text{s.t.} \\ \mu \geq 0}} \theta(\lambda, \mu) \leq \min_{\substack{x \\ \text{s.t.} \\ h(x)=0 \\ g(x) \leq 0}} f(x)$$

אסטרטגיה הכללית

אם f - פונקציה קמורה h - פונקציה קמורה
 (אם $A \in M_{k,n}(\mathbb{R})$ ו- $b \in \mathbb{R}^k$ אז $h(x) = Ax + b$)

$$\max_{\substack{\lambda, \mu \\ \text{s.t.} \\ \mu \geq 0}} \theta(\lambda, \mu) = \min_{\substack{x \\ \text{s.t.} \\ h(x)=0 \\ g(x) \leq 0}} f(x)$$

הוכחה



$$\theta(\mu) = \min_x [f(x) + \mu g(x)]$$

הישר $g(x) \leq 0$
 $z + \mu g$

$$\mathcal{L}(x, \mu) = f(x) + \mu g(x)$$

$$\Rightarrow \max_{\substack{\mu \\ \text{s.t.} \\ \mu \geq 0}} \theta(\mu)$$

אם הפונקציה f היא קמורה
 אז הישר $z + \mu g$ שמשק $(g(x), f(x))$
 הוא זכר של f והוא
 התחתון.

הוא למעשה זכר של f
 והוא הישר $z + \mu g$ שמשק
 $(g(x), f(x))$ הוא זכר של f
 והוא הישר $z + \mu g$ שמשק
 $(g(x), f(x))$ הוא זכר של f

$$\min_{w, b, \epsilon_i} \frac{1}{2} w \cdot w + \nu \sum \epsilon_i$$

s.t.

$$y_i(w x_i - b) \geq 1 - \epsilon_i$$

$$\epsilon_i \geq 0$$

לחשב את הפונקציה

$$\mathcal{L}(w, b, \epsilon_i, \mu, \delta) = \frac{1}{2} w w + \nu \sum \epsilon_i - \sum \mu_i (y_i (w x_i - b) - 1 + \epsilon_i)$$

x μ

$$- \sum \delta_i \epsilon_i$$

$$\theta(\mu, \delta) = \min_{w, b, \epsilon} \mathcal{L}(w, b, \epsilon_i, \mu, \delta)$$

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^m \mu_i y_i x_i = 0 \Rightarrow w = \sum \mu_i y_i x_i$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^m \mu_i y_i = 0$$

$$i=1, \dots, m \quad \frac{\partial \mathcal{L}}{\partial \epsilon_i} = \nu - \mu_i - \delta_i = 0 \Rightarrow \nu - \mu_i \geq 0 \Rightarrow 0 \leq \mu_i \leq \nu$$

$\delta_i \geq 0$ $\delta_i = \nu - \mu_i$

$\mu_i \geq 0$

$$\theta(\mu, \delta) = \frac{1}{2} (\sum \mu_i y_i x_i) (\sum \mu_j y_j x_j) +$$

$$+ \nu \sum \epsilon_i - \sum \mu_i (y_i (\sum \mu_j y_j x_j) x_i - b - 1 + \epsilon_i)$$

$$- \nu \sum \epsilon_i + \sum \mu_i \epsilon_i =$$

$$= \frac{1}{2} \sum_{i,j} \mu_i \mu_j y_i y_j x_i x_j - \sum_{i,j} \mu_i \mu_j y_i y_j x_i x_j +$$

$$+ b \underbrace{\sum \mu_i y_i}_{=0} + \sum \mu_i - \sum \mu_i \epsilon_i + \sum \mu_i \epsilon_i =$$

$$= \sum \mu_i - \frac{1}{2} \sum_{i,j} \mu_i \mu_j y_i y_j x_i x_j$$

לחשב את הפונקציה

$$\max_{\mu_1, \dots, \mu_m} \theta(\mu) = \sum \mu_i - \frac{1}{2} \sum \mu_i \mu_j y_i y_j x_i x_j$$

s.t. $0 \leq \mu_i \leq \nu$

$$\sum y_i \mu_i = 0$$

התיבה יותר קומפקטית (היא)

$$\phi(\mu) = \mu^T \mathbf{I} - \frac{1}{2} \mu^T M \mu$$

$$(M)_{i,j} = y_i y_j x_i x_j \quad \text{צאט}$$

M היא positive semi definite, כלומר $\forall v \neq 0$

$$v^T M v \geq 0$$

$$M = \underbrace{\begin{pmatrix} (y_1 x_1) \\ \vdots \\ (y_n x_n) \end{pmatrix}}_{A^T} \underbrace{\begin{pmatrix} (y_1 x_1) & \dots & (y_n x_n) \end{pmatrix}}_A \quad \text{Gram Matrix}$$

$$\Rightarrow v^T M v = v^T A^T A v = \|A v\|^2 \geq 0$$

סיום של הקצות הצליל

$$\begin{aligned} f(x) &= \text{sign}(w x - b) = \text{sign}\left(\left(\sum \mu_i y_i x_i\right) x - b\right) = \\ &= \text{sign}\left(\sum \mu_i y_i \langle x_i, x \rangle - b\right) \end{aligned}$$

נשים לב ש- M תלויה רק במסלול פנימי בין x_i -ים
(אין תלות מפורשת ב- x וצדדים) ורק זה סיום של
הקצות הצליל.

The Kernel Trick

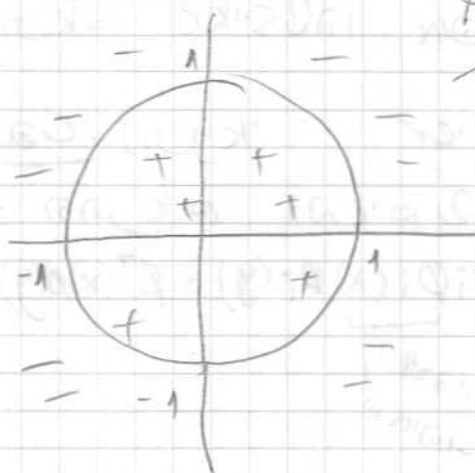
נניח למשל את ϕ הקצות למרחב אחר "אופייני"
 $\phi: \mathbb{R}^n \rightarrow \mathcal{F}$ כזה ש- $\mathcal{F}$ הוא מרחב פנימי
ועליו פערדה פנימי ב- $\mathcal{F}$.

אם אנו רוצים למצוא חוקים למרחב אחר הקצות ב- $\mathcal{F}$
(שיכול להיות מרחב אינסופי), אז נניח למשל
פנימי צדק $K: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ (kernel) ואז
 $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$

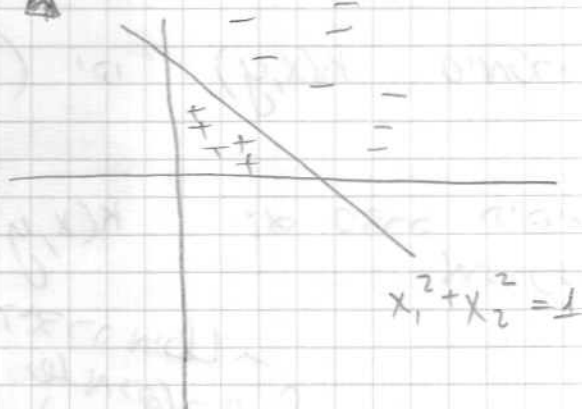
(21)

: הפונקציה

$$\phi(x_1, x_2) = (x_1^2, x_2^2)$$



האם יש אפשרות
למצוא פונקציה
הפונקציה



האם אפשר
למצוא פונקציה

הפונקציה

$$K(x, y) = (x^T y)^d$$

$$\phi(x) = \left(\sqrt{\binom{d}{n_1, \dots, n_k}} x_1^{n_1} \dots x_k^{n_k} \right)_{n_1 + \dots + n_k = d}$$

$d=2$, $k=2$ - הפונקציה, למשל

$$K(x, y) = (x^T y)^2 = (x_1 y_1 + x_2 y_2)^2 = x_1^2 y_1^2 + 2x_1 x_2 y_1 y_2 + x_2^2 y_2^2 = \phi(x)^T \phi(y)$$

$$\phi(x) = (x_1^2, x_2^2, \sqrt{2} x_1 x_2)$$

הפונקציה

$$K(x, y) = (x^T y + \theta)^d$$

הפונקציה

הפונקציה, הפונקציה, הפונקציה

Radical Basis Function

RBF

$$K(x, y) = e^{-\|x-y\|^2 / 2\sigma^2}$$

$$\phi(x) = \left(\frac{e^{-\frac{\|x\|^2}{2\sigma^2}}}{(\sqrt{j!})^{1/2}} \binom{j}{n_1, \dots, n_k} x_1^{n_1} \dots x_k^{n_k} \right)_{j=0, \dots, \infty, \sum n_k = j}$$

הפונקציה, הפונקציה, הפונקציה

22 11.12.07
אוריאל '88

SVM תלכורת

primal

$$\min_{w, b, \epsilon} \frac{1}{2} w w + \nu \sum_{i=1}^m \epsilon_i$$

$$\text{s.t. } y_i (w x_i - b) \geq 1 - \epsilon_i \\ \epsilon_i \geq 0$$

$$\mathcal{L} = \frac{1}{2} w w + \nu \sum \epsilon_i - \sum \mu_i (y_i (w x_i - b) - 1 + \epsilon_i) - \sum \delta_i \epsilon_i$$

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum \mu_i y_i x_i = 0 \Rightarrow w = \sum \mu_i y_i x_i$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum \mu_i y_i = 0$$

$$\frac{\partial \mathcal{L}}{\partial \epsilon_i} = \nu - \mu_i - \delta_i = 0 \Rightarrow 0 \leq \mu_i \leq \nu$$

dual

$$\max_{\mu} \theta(\mu) = \sum \mu_i - \frac{1}{2} \sum \sum \mu_i \mu_j y_i y_j \underbrace{x_i x_j}_{K(x_i, x_j)}$$

$$\text{s.t. } \sum y_i \mu_i = 0 \\ 0 \leq \mu_i \leq \nu$$

סיווג על הקצרות וקצרות

$$f(x) = \text{sign}(w x - b) = \text{sign}(\sum \mu_i y_i k(x_i, x) - b) =$$

$$= \text{sign}(\underbrace{\sum_{i: \mu_i > 0} \mu_i y_i k(x_i, x)}_{\text{support vectors}} - b)$$

בתקצרות סיווג השל"ם או התקצרות השל"ם מקצרות $y_i (w x_i - b) = 1 - \epsilon_i$
תמוך $\mu_i > 0$ וכן בתקצרות שלם שלם שלם שלם.

SVM מ'חושבים על

gradient ascent שיטה יסודית

נניח ש- b נקבע מראש (נניח $b=0$). מאחר שהם קבועים אנו זורקים את התנאי $\sum y_i \mu_i = 0$ להתקנה. מניסיון אפ- b ואם מתקבלת בעיה אופטימלית. נסו להתנאי $\sum y_i \mu_i = 0$. נסבך הכיוון הגרדיאנט

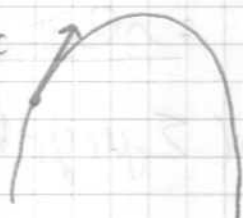
$$\frac{\partial \theta(\mu)}{\partial \mu_i} = 1 - \frac{1}{2} \sum_{j \neq i} \mu_j y_i y_j k(x_i, x_j) - \frac{1}{2} \sum_{j \neq i} \mu_j y_i y_j k(x_i, x_j) - \mu_i y_i^2 k(x_i, x_j) = 1 - y_i \sum_{j=1}^m \mu_j y_j k(x_i, x_j)$$

$$\mu_i^{t+1} = \mu_i^t + \eta \frac{\partial \theta(\mu)}{\partial \mu_i} \bigg|_{\mu_i^t} = \mu_i^t + \eta (1 - y_i \sum_{j=1}^m \mu_j^t y_j k(x_i, x_j))$$

learning rate
הצעד של μ אחרי $t+1$ צעדונים אפ-גרדיאנט

הגרדיאנט אומר האם כיוון הפונקציה יעלה או ייפול. חזרים על הכיוון הגרדיאנט וזה יגדיל או הפונקציה אולי ייפול. חזרים על זה יותר ויותר כי אולי נגיע לפסס :

אם נאק יותר מדי הכיוון יהיה נגד אדג'קט וזה לא מה שצריך.



אם עושים תהליך איטרטיבי שבו μ נעם צדים קצת, ומשלים שוב את הגרדיאנט וכלים עוד ...

איתותים: נתונה קבוצת אימון S וקבוצת אימון $\mu_1, \dots, \mu_m, 0 < \eta$

המשקל
המאזן

(13)

$$\vec{\mu} = 0$$

repeat

for $i=1$ to m

$$\mu_i = \mu_i + \eta_i (1 - y_i \sum_j \mu_j y_j K(x_i, x_j))$$

if $\mu_i < 0$ then $\mu_i = 0$ if $\mu_i > \nu$ then $\mu_i = \nu$

end for

until stopping criterion

return $\vec{\mu}$ מהו קריטריון העצירה?

יש 3 דרכים להגדיר תנאי עצירה

(1) הפיכת קצת יצוא של פונקציה המטרה $\theta(\mu)$
 אם $\theta(\mu^{t+1}) - \theta(\mu^t) < 10^{-9}$ עצור (איתחל אלמן)

(2) נאמקם אחרי תנאי Karush-Kuhn-Tucker

$$y_i f(x_i) = y_i \left(\sum_j y_j \mu_j K(x_i, x_j) - b \right) = \begin{cases} \geq 1 & \mu_i = 0 \\ = 1 & 0 < \mu_i < \nu \\ \leq 1 & \mu_i = \nu \end{cases}$$

(3) הפיכת סך בין Dual - δ Primal

$$W = \sum \mu_i y_i f(x_i) \quad WW = \sum \mu_i y_i y_j K(x_i, x_j)$$

$$\begin{aligned} \text{primal} - \text{dual} &= WW + \nu \sum \varepsilon_i - \sum \mu_i = \\ &= \sum \mu_i y_i y_j K(x_i, x_j) + \nu \sum \varepsilon_i - \sum \mu_i \end{aligned}$$

$$\varepsilon_i = \max(0, 1 - y_i (\sum_j y_j \mu_j K(x_i, x_j) - b)) \quad \text{כאשר}$$

קביעת קצת הראשוני

$\theta(\mu)$ היא פונקציה ריבועית בס אתר אה-מל-ים.

ניתן לקבוע יש מאקסימום של הפרבולה



$$\mu_i^{t+1} = \mu_i^t + \eta_i (1 - y_i \sum_j \mu_j^t y_j K(x_i, x_j))$$

$$\mu = (\mu_1^t, \dots, \mu_m^t)$$

: NO)

$$\hat{\mu} = (\mu_1^t, \dots, \mu_i^{t+1}, \dots, \mu_m^t)$$

$$0 = \frac{\partial \theta(\hat{\mu})}{\partial \mu_i} = \frac{\partial \theta(\hat{\mu})}{\partial \mu} \frac{\partial \mu}{\partial \mu_i} =$$

$$= (1 - y_i \sum_j \hat{\mu}_j y_j K(x_i, x_j)) (1 - \eta_i \underbrace{y_i y_i}_1 K(x_i, x_i))$$

$$\Rightarrow \eta_i = \frac{1}{K(x_i, x_i)}$$

טפוח ה-bias $\frac{b}{\text{טפוח}}$ $x \in \mathbb{R}^n$ וקצב $\hat{x} \in \mathbb{R}^{n+1}$ קצב -0

$$\hat{x} = (x, \tau)$$

יהי $\hat{w} = (w, -\frac{b}{\tau})$ גפכיה ה- $\mathbb{R}^{n+1}$. τ קצב

$$\langle \hat{w}, \hat{x} \rangle = \langle w, x \rangle - b$$

אם ההפכיה ה- $\mathbb{R}^{n+1}$ ה- $\mathbb{R}^n$ $\text{bias} = -b$ $\text{bias} = b$ τ קצב

והחבר יצאן טפוח :

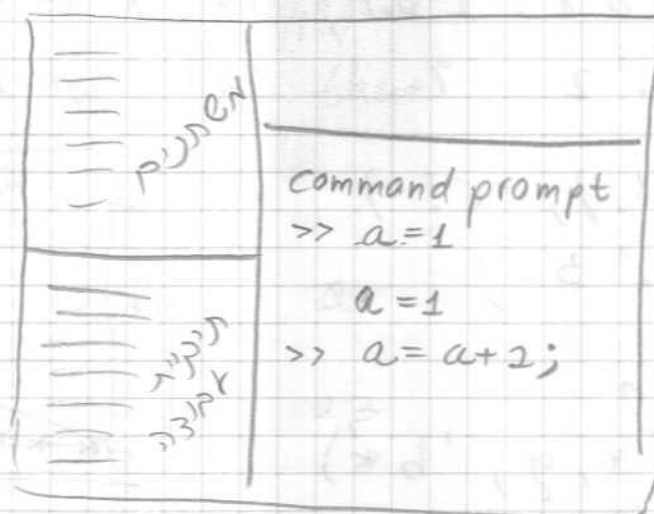
$$\hat{K}(x, y) = K(x, y) + \tau^2$$

בחירה טפוח τ ה-1 :

$$\tau = R = \max_{x_i} \|x_i\|$$

(אשר ערואל הספר)

Matlab



by value המשנים מוחזרים ב פונקציות

function [y1, y2] = myf(x1, x2)

% comment

y1 = x1

y2 = x1 + x2

matlab מספק מספר פונקציות למטרה של חישוב וקריאה

A = zeros(3,2); % A = $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

A(1,2) = 1; % A = $\begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

B = [0, 1; 2, 3]; % B = $\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$

X1 = [1, 2]'; % X1 = $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

X2 = [3, 4]; % X2 = $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$

X1 + X2; % $\begin{bmatrix} 4 \\ 6 \end{bmatrix}$

יש בוא לראות מטריצות וחיבור הוא אשר אשר

3 פרטים נוספים

figure ; % open new window

x = 0:0.1:1-0.1; % x = (0 0.1 0.2 ... 0.9 1)

y = x.^2 % נקודות בלבד

plot(x, y, 'r') % קו אדום

y = x.^3 ;

hold on

plot(x, y, 'b*') % קו כחול עם כוכבים

xlabel('x')

ylabel('y')

legend('y1', 'y2')

title('my Graph')

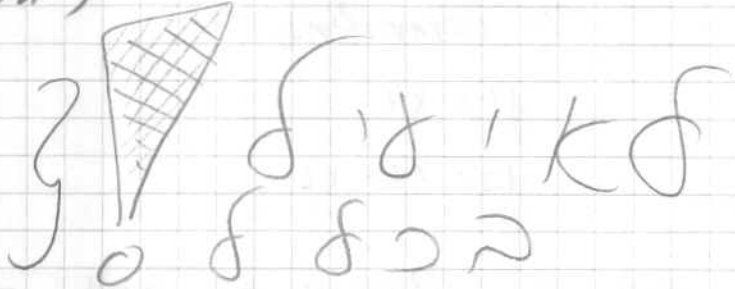
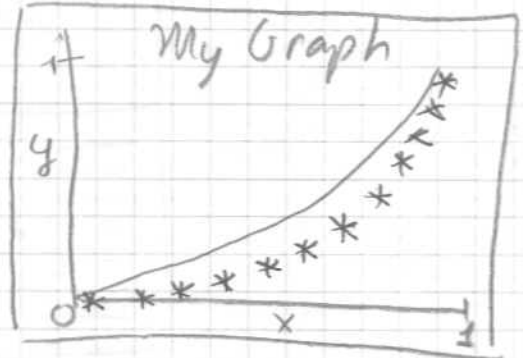
x=0;

for i=1:3

x(1,i)=i;

end

exp([1 2 3]) % [e¹, e², e³]



SVM פתרון

SMO - Sequential Minimal Optimization

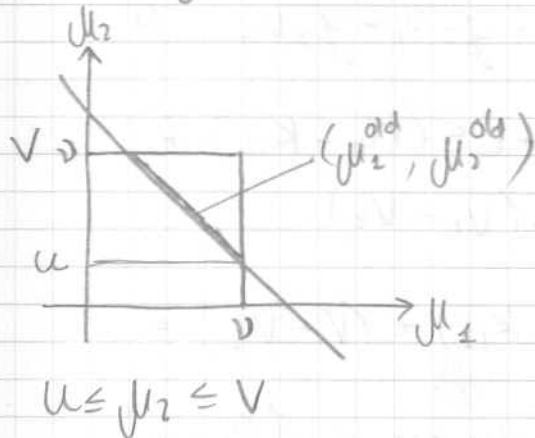
$$\max_{\mu_1, \dots, \mu_m} \theta(\mu) = \sum_{i=1}^m \mu_i - \frac{1}{2} \sum_{i,j=1}^m \mu_i \mu_j y_i y_j K(x_i, x_j)$$

$$\text{s.t.} \quad \sum y_i \mu_i = 0$$

$$0 \leq \mu_i \leq 1$$

פתרון מדויק עבור שני נקודות (אנחנו נניח $\mu_1, \dots, \mu_m$ כמיל קבועים)
מרחיבים מניחים $\mu_1 = 0$ נבדל שלב ושאר אחרים

$$\mu_1 y_1 + \mu_2 y_2 = \text{const} = \mu_1^{\text{old}} y_1 + \mu_2^{\text{old}} y_2$$



$$y_1, y_2 = \{\pm 1\}$$

$$\Rightarrow \frac{y_2}{y_1} = y_2 \cdot y_1$$

$$\Rightarrow \mu_1 = \frac{\text{const}}{c} - \frac{y_1 y_2}{s} \mu_2 = C - s \mu_2$$

$$\Rightarrow y_1 C - y_1 \mu_1^{\text{old}} = y_2 \mu_2^{\text{old}}$$

אילוץ μ_2 :

$$y_1 \neq y_2 : u = \max(0, \mu_2^{\text{old}} - \mu_1^{\text{old}})$$

$$v = \min(v, v - \mu_1^{\text{old}} + \mu_2^{\text{old}})$$

$$y_1 = y_2 : u = \max(0, \mu_1^{\text{old}} + \mu_2^{\text{old}} - v)$$

$$v = \min(v, v - \mu_1^{\text{old}} + \mu_2^{\text{old}})$$

μ_1, μ_2 על ציר $\theta(\mu)$ כמיל קבועים

$$V_i = \sum_{j=3}^m y_j \mu_j K(x_i, x_j) = f(x_i) - \sum_{j=1}^2 y_j \mu_j^{\text{old}} K(x_i, x_j) \quad \text{אנחנו}$$

$$f(x_i) = w x_i = \sum_{j=1}^m y_j \mu_j^{\text{old}} K(x_i, x_j) \quad \text{לכן}$$

$$\begin{aligned}\theta(\mu) &= \sum \mu_i - \frac{1}{2} \sum_{i,j} \mu_i \mu_j y_i y_j \underbrace{K(x_i, x_j)}_{K_{ij}} = \\ &= \mu_1 + \mu_2 - \frac{1}{2} \mu_1^2 K_{11} - \frac{1}{2} \mu_2^2 K_{22} - y_1 y_2 K_{12} \mu_1 \mu_2 - \\ &\quad - y_1 \mu_1 V_1 - y_2 \mu_2 V_2 + \text{const} =\end{aligned}$$

$$\begin{aligned}\checkmark \quad \mu_1 = c - s \mu_2 &= c - s \mu_2 + \mu_2 - \frac{1}{2} K_{11} (c - s \mu_2)^2 - \\ &\quad - \frac{1}{2} K_{22} \mu_2^2 - s K_{12} (c - s \mu_2) \mu_2 - \\ &\quad - y_1 (c - s \mu_2) V_1 - y_2 \mu_2 V_2 + \text{const}\end{aligned}$$

$$\begin{aligned}\frac{\partial \theta(\mu)}{\partial \mu_2} &= 1 - s + s K_{11} (c - s \mu_2) - K_{22} \mu_2 - \\ &\quad - s K_{12} (c - s \mu_2) + K_{12} \mu_2 + y_2 V_1 - y_2 V_2 = 0\end{aligned}$$

$$\Rightarrow \mu_2 \left(\frac{K_{11} + K_{22} - 2 K_{12}}{K} \right) = 1 - s + c s (K_{11} - K_{12}) + y_2 (V_1 - V_2)$$

$$\begin{aligned}\Rightarrow \mu_2 K y_2 &= y_2 - y_1 + c y_1 (K_{11} - K_{12}) + (V_1 - V_2) = \\ \downarrow & \\ y_2 s = y_1 & \\ y_2 y_2 = 1 & \\ &= y_2 - y_1 + f(x_1) - \sum_{j=1}^2 y_j \mu_j K_{ij} - f(x_2) + \\ &\quad + \sum_{j=1}^2 y_j \mu_j K_{2j} + c y_1 K_{11} - c y_1 K_{12} =\end{aligned}$$

$$\begin{aligned}&= y_2 - y_1 + f(x_1) - f(x_2) - y_1 \mu_1^{\text{old}} K_{11} - \\ &\quad - y_2 \mu_2^{\text{old}} K_{12} + c y_1 K_{11} + y_1 \mu_1^{\text{old}} K_{21} + \\ &\quad + y_2 \mu_2^{\text{old}} K_{22} - c y_1 K_{12} =\end{aligned}$$

$$\begin{aligned}&= y_2 - y_1 + f(x_1) - f(x_2) + y_2 \mu_2^{\text{old}} K_{22} - \\ &\quad - y_2 \mu_2^{\text{old}} K_{12} + y_2 \mu_2^{\text{old}} K_{11} - y_2 \mu_2^{\text{old}} K_{12} =\end{aligned}$$

$$\begin{aligned}y_1 c - y_1 \mu_1^{\text{old}} &= y_2 \mu_2^{\text{old}} \\ (\text{max 0}) \quad K_{12} &= K_{21}\end{aligned}$$

$$\begin{aligned}&= y_2 \mu_2^{\text{old}} \underbrace{(K_{11} + K_{22} - 2 K_{12})}_K + \underbrace{(f(x_1) - y_1)}_{E_1} - \underbrace{(f(x_2) - y_2)}_{E_2}\end{aligned}$$

(26)

$$\Rightarrow \mu_2^{\text{new}} = \mu_2^{\text{old}} + \frac{y_2 (E_1 - E_2)}{K}$$

זרשונים כחלק אר μ_2^{new} פקט $u \leq \mu_2^{\text{new}} \leq v$

לכאמור נחש'מנו רק: (תחול N - ϵ - μ) וכל
פער נחתך באר μ -ים ונקטם ביחס אליהם. לכן
כל יום לן פיתרון מקובל.

(שאר השאלה איך לבחור את הצד הארוך טובה
די לכאמור ותנסו כמה שיותר מזה).

יוריסטיקה לבחירת μ , μ_2 לעצמם:

בחירת נקודה 1: מתוך הנקודות שאפרות את KKT ,

באור הנקודה שמתקן $0 < \mu < v$

בחירת נקודה 2: המטרה היא שיפור מקסימלי של קריטריון העצירה.

אם יש קשה לבחור את זה ואם לוקחים יוריסטיקה

אפסטרטגיה:

1) לבחור μ רק ש- $|E_1 - E_2| = |(f(x_1) - y_1) - (f(x_2) - y_2)|$
מקסימלי

2) אם זה לא עזר (באור ה $primal$ או $dual$ K

התקבל) לבחור $0 < \mu < v$

3) אם זה לא עזר אז עושים סיבוב ^{בסדר הקדמ} על הנקודות

עד בראשונה שמשפרת את קריטריון העצירה. אם מצאנו רצו

אז עושים את השיפור ובתחתית נחש'מנו נקודות (אם

השתפרנו ביותר מ- ϵ).

הכל האמירות הנכד הכלל לא הופיע ב. איקונסס אמור

זמר μ שמקיים $0 < \mu < v$ KKT ϵ פו

האי שוויונים $(b - w^T x) / y_i \geq 1 - \epsilon$ מקיימים בשוויונים

וניתן לבחור את b . (אם עושים את זה בשלבי קטנים

יכול להיות over estimation של b)

פחיתות עם הקוואר

תסמיט של חיבור מוצר

$$P(\exists h \in C: |err(h) - \hat{err}(h)| > \epsilon) \leq 8 \Phi_d(m) e^{-m\epsilon^2/32}$$

טווח $\Phi_d(m)$ מספר האפשרויות m (קצות) C - N זיהוי

חיון היתר מופלט:

$$S = S_1 \cup S_2 \quad m \text{ מופלט}$$

$$P(\exists h: |err(h) - \hat{err}_{S_1}(h)| > \epsilon) \leq 2 P(\exists h: |\hat{err}_{S_1}(h) - \hat{err}_{S_2}(h)| > \frac{\epsilon}{2})$$

$$P(\exists h: |\hat{err}_{S_1}(h) - \hat{err}_{S_2}(h)| > \epsilon) \leq \underbrace{2e^{-m\epsilon^2/2}}_{\text{הקצות}} \underbrace{\Phi_d(2m)}_{\text{איתור}}$$

$$X_i = \text{ind}(\underbrace{h(x_i^1) = y_i^1}_{S_1}) - \text{ind}(\underbrace{h(x_i^2) = y_i^2}_{S_2}) = \begin{cases} 1 \\ 0 \\ -1 \end{cases} \quad \begin{matrix} \mathbb{E} X_i = 0 \\ \text{הסתברות} \end{matrix}$$

$$8 \Phi_d(m) e^{-m\epsilon^2/32} < \delta \quad \text{איתור מוצר ש}$$

$$m \geq O\left(\frac{1}{\epsilon^2} \ln \frac{1}{\delta} + \frac{d}{\epsilon^2} \ln \frac{1}{\epsilon}\right) \Leftrightarrow \Phi_d(m) \leq \left(\frac{em}{d}\right)^d$$

יש מורכבות המציג

$$err(h) \leq \hat{err}(h) + \sqrt{\frac{32(d \ln \frac{em}{d} + \ln \frac{4}{\delta})}{m}} \quad \text{חסם הבטוח אחר ש}$$

נניח ש $C = C_1 \cup \dots \cup C_k$. איתור מוצר במסמך δ . כל תת מוצר C_j (יתר המסמך δ . $p(C_j)$. איתור מוצר

$$P(\exists h \in C_j: |err(h) - \hat{err}(h)| > \epsilon) < P(C_j) \delta$$

$$h \in C_j \quad \text{אז נרצה ש} \quad 8 \Phi_d(m) e^{-m\epsilon^2/32} < P(C_j) \delta$$

$$err(h) \leq \hat{err}(h) + \sqrt{\frac{32(d_j \ln \frac{em}{d_j} + \ln \frac{4}{p(C_j) \delta})}{m}}$$

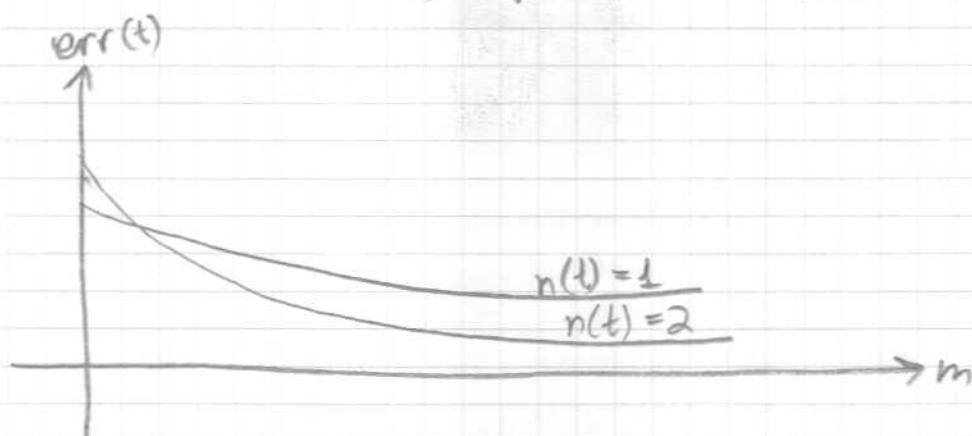
זה מראה שיש trade-off בין המסמך האיתור לסיוכמו המחקר.

24

האקרה של צב המלכה עם צבג פיצוץ

$$P(t) = \frac{1}{2^{n(t)+1}} \frac{1}{2^{2^{n(t)}}}$$

$$\Rightarrow \text{err}(t) \leq \hat{\text{err}}(t) + \sqrt{\frac{\ln 2 (\lambda^{n(t)} + n(t)+1) + \ln \frac{1}{\delta}}{2m}}$$

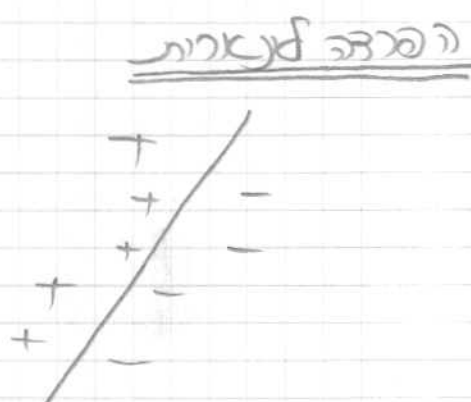


אם יש אופי קטן של פונקציות מורכבות

עם זאת מתחילת פונקציות

$$\text{VC-dim}(\mathbb{R}^n) = n+1$$

חלקנו את המערכות לבד
תת מרחב של המרחב



יש trade off בין השטח לסטיות ואנחנו רוצים למצוא

$$\min \frac{1}{2} \|w\|^2 + \nu \sum \epsilon_i$$

אם זה אנוני הנכנס לבחון בתור...

1. The 2nd order approximation is

$$f(x) \approx f(0) + f'(0)x + \frac{f''(0)}{2}x^2$$

$$f(x) = (1+x)^2 \approx 1 + 2x + x^2$$

$$f(0) = 1$$

$$f'(0) = 2$$

2. The 3rd order approximation is

$$f(x) \approx f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f'''(0)}{6}x^3$$

$$f(x) = (1+x)^2$$

$$f(0) = 1$$

$$f'(0) = 2$$

$$f''(0) = 2$$

$$f'''(0) = 0$$

3. The 4th order approximation is

$$f(x) \approx f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f'''(0)}{6}x^3 + \frac{f^{(4)}(0)}{24}x^4$$

4. The 5th order approximation is

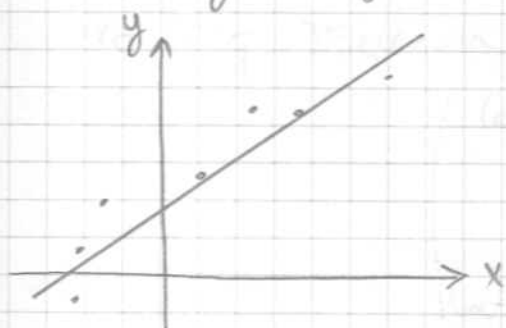
שיטת סבךטורטור כניולת של (תונ'ס)

Linear Least Squares

צורטור

① כרתאמה (fitting) של (תונ'ס ע' ק' ישר.

יש קרטר $\{(x_i, y_i)\}_{i=1}^m$ צורטור מרטר ישר $y = \alpha + \beta x$ שמתאם לרן.



אם $m > 2$ אז יש יותר אורטר (שלאים) וזה קרטר overdetermined system

אז זה שאננו אורטרטור $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ ק' ע.

$$\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

$A \quad x \quad b$

רטרטרטור מרטרטור א' ק' ->

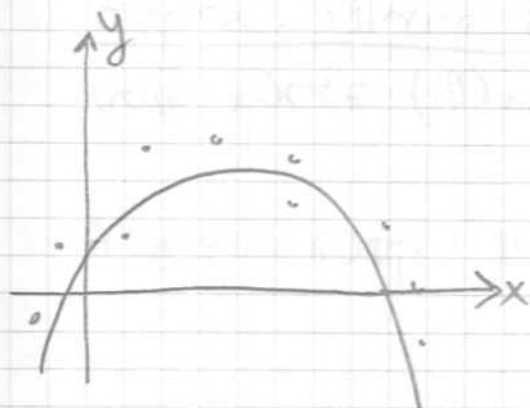
$$\begin{pmatrix} (x_1) \\ \vdots \\ (x_m) \end{pmatrix} W = y$$

$$y = w_0 + w_1 x + \dots + w_n x^n$$

② כרתאמה ע' פורטר (ר)

$$\begin{pmatrix} 1 & x_1 & \dots & x_1^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & \dots & x_m^n \end{pmatrix} \begin{pmatrix} w_0 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

(תונ'ס כרתאמה) $\{(x_i, y_i)\}_{i=1}^m$



(3) נכונה תהיה של תיבקים K -א סוגים שונים.
 תיבקי אסא: i אגורה הקרה אקספוננציאלית החומר ϵ''
 המשוואה
$$\underline{n_i(t)} = n_i(0) e^{\lambda_i t}$$

לחות תיבקים אסא: i הלמן t
 אל הכמות הקולל של התיבקים הם $n(t) = \sum_{i=1}^K n_i(t) = \sum_{i=1}^K n_i(0) e^{\lambda_i t}$
 נתון: קרה אצנל λ

- מדידות של $n(t)$ הלמים $t_1, \dots, t_m$
 צרכ, למצוא: תהיה של תהיה הלמן 0 , סומר $n_1(0), \dots, n_k(0)$
 אז צדיק לפעור אל: הקרה והסא:

$$\begin{pmatrix} e^{\lambda_1 t_1} & \dots & e^{\lambda_k t_1} \\ \vdots & & \vdots \\ e^{\lambda_1 t_m} & \dots & e^{\lambda_k t_m} \end{pmatrix} \begin{pmatrix} n_1(0) \\ \vdots \\ n_k(0) \end{pmatrix} = \begin{pmatrix} n(t_1) \\ \vdots \\ n(t_m) \end{pmatrix}$$

ושבים אנו מצה של יותר משוואות אנשלמים

המטח שלנו היא אלמור לפעור משוואות אנלורה $Ax = b$ היא יש יותר משוואות אנשלמים.

משוואה $Ax = b$ לקרא homogeneous אם $b = 0$, סומר אם
 המשוואה היא $Ax = 0$.

פתרון כללי של $Ax = b$ נותן אחרת הצורה

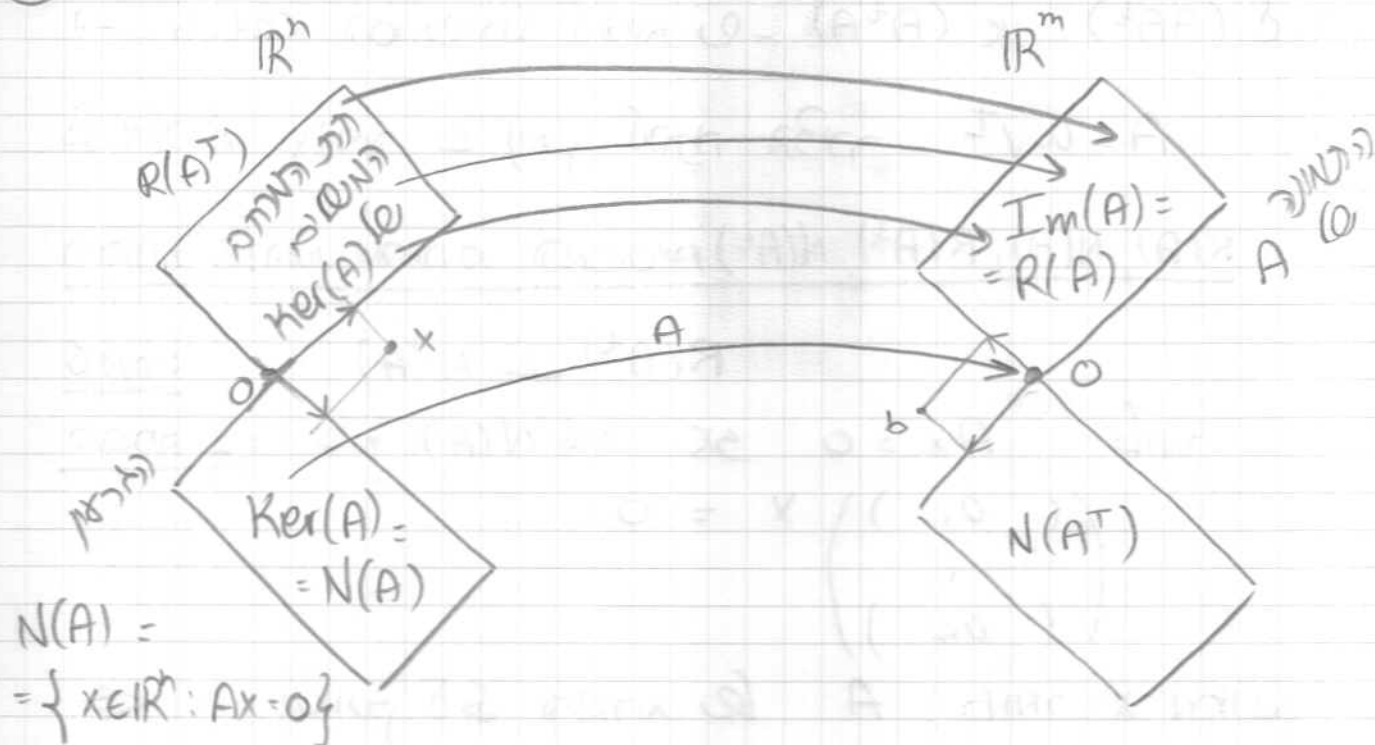
$$x_{\text{general}} = \underbrace{x_{\text{particular}}}_{\substack{\text{פתרון כללי של} \\ Ax = b}} + \underbrace{x_{\text{homogenous}}}_{\substack{\text{פתרון כללי של} \\ Ax = 0}}$$

תלכידת מאלמורה עין אכיות

$$\underbrace{\left\{ \begin{pmatrix} \vdots \\ \vdots \\ \vdots \end{pmatrix} \right\}}_{\substack{\text{מחר} \\ \text{מחר} \\ \text{מחר}}} = A \in M_{m \times n}(\mathbb{R})$$

A מדידות הקרה $\mathbb{R}^n - \mathbb{R}^m$.

29



$\text{Ker}(A)$ is the set of vectors in $\mathbb{R}^n$ that are mapped to the zero vector in $\mathbb{R}^m$. It is the null space of A .

$$\text{rank}(A) = \text{rank}(A^T) = r$$

$$\dim(R(A)) = \dim(R(A^T)) = r$$

rank(A) is the number of non-zero rows in the row echelon form of A.

Rank(A) is the number of non-zero rows in the row echelon form of A.

$BA = I_n$ - $B \in M_{n \times n}(\mathbb{R})$ is the inverse of A.

$AC = I_m$ - $C \in M_{m \times m}(\mathbb{R})$ is the inverse of A.

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

$$B = B(AC) = (BA)C = C$$

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

Let B and C be matrices such that $BA = I_n$ and $AC = I_m$.

ה- catch הם שיתל עברות ש- $(A^t A)$ או $(A A^t)$ לא הריכתי.

באמצעות אמצעי $\perp$ ניתן לכתוב הצורה $A = U V^t$

תכונות של הוגר מרחבים והמסוסים $R(A)$, $N(A)$, $R(A^t)$, $N(A^t)$

טענה: $R(A^t) \perp N(A)$

הוכחה 1: אם $x \in N(A)$ אז $Ax = 0$ עבור

$$\begin{pmatrix} (a_1) \\ \vdots \\ (a_m) \end{pmatrix} x = 0$$

$x \in N(A)$ מאונק עם המרחב של A , A מאונק x מאונק $R(A^t)$ (11)

הוכחה 2: אם נתון $v \in R(A^t)$ אז $v = A^t z$ עבור

$$v^t x = (A^t z)^t x = z^t \underbrace{Ax}_{=0} = 0 \quad \text{אם } z \in \mathbb{R}^m \text{ ו} \quad x \in N(A)$$

(12)

טענה: המרחב המשותף של $N(A)$ ו- $R(A^t)$ הוא

הוכחה: $\dim(R(A)) = \dim(R(A^t))$

$$\dim(R(A)) + \dim(N(A)) = n$$

$$\Rightarrow \dim(R(A^t)) + \dim(N(A)) = n$$

לכן $R(A^t) \perp N(A)$ נקטו ש- $R(A^t)$ ו- $N(A)$ יחד

אם $N(A)$ פורשים את $\mathbb{R}^n$ (13)

באנו אכן $R(A) \perp N(A^t)$ וקודם הם פורשים את $\mathbb{R}^m$

למשל $Ax = b$ יש פיתרון אם $b \in R(A)$ עבור

$$b \perp N(A^t)$$

אם $x = x_r + x_n$ כאשר $x_r \in R(A^t)$ ו- $x_n \in N(A)$ אז

$$Ax = Ax_r + Ax_n = Ax_r$$

טענה: $x_r \in R(A^t)$ היא פיתרון של $Ax = b$ אם x_r היא ותיק.

הוכחה: נניח בשאלה שיש שני x_r, x_r' כך ש- $Ax_r = b$ ו- $Ax_r' = b$

$$A(x_r - x_r') = Ax_r - Ax_r' = 0 \quad \text{אז}$$

(30)

$$x_r - x_{r'} \in N(A) \cap R(A^t) \iff x_r - x_{r'} \in N(A) \text{ מכ}$$

$$. x_r = x_{r'} \iff x_r - x_{r'} = 0 \iff \text{והכי הם אותו כי}$$

(31)

המשפט: לא יתכן להיות שני וקטורים ב- $R(A^t)$ שזכור

דאטה (קצה) ב- $R(A)$ מכ ההצקה מזווגות - $R(A^t)$

ביאור: ואם לא יתכן ההצקה (בוכה, בואו יש ההצקה

pseudo inverse של R^m של $R(A)$ הפכה של $R(A)$.

אז נהפך, ב (כאן אותה) $R(A)$ (יורה) אן הם עובר.

הכאן של וקטור של תת מרחב

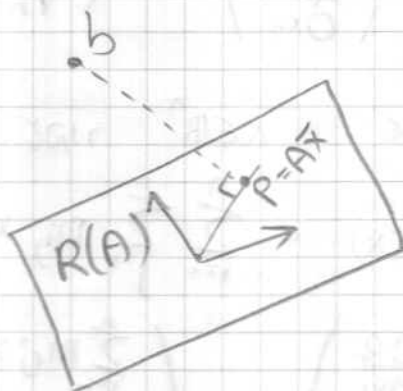
$$Ax = b$$

$$b \text{ של } p = A\bar{x} \text{ = ההכאן של } b$$

$$R(A) \text{ - וואת שפירש}$$

$$A \text{ של } \bar{x} \text{ דע מוכר של}$$

$$b - p = b - A\bar{x} \perp R(A) \text{ מכ}$$



$$A^t(b - A\bar{x}) = A^tb - A^tA\bar{x} = 0 \iff b - A\bar{x} \in N(A^t) \iff$$

$$A^tb = A^tA\bar{x} \iff$$

$$\underbrace{A^tA}_{\text{נורמל קטור}} \text{ של } A \text{ מוגדר כזו} - A^tA \text{ יש דופכי וקטור}$$

$$\bar{x} = (A^tA)^{-1}A^tb$$

$$N(A) = \{0\} \text{ ! } R(A^t) = R^n \text{ מכ } A \text{ מוגדר כזו } R^p$$

$$p = \underbrace{A(A^tA)^{-1}A^t}_{\text{מטריצה ההכאן}} b = Pb$$

$$b \in R^m \text{ מוגדר } P \text{ מטריצה אחר } R(A).$$

מקרה קצה

$$1) \text{ מכ } b \in R(A) \text{ אז } b = Ax \text{ של } x \text{ !}$$

$$p = A(A^tA)^{-1}A^tAx = Ax = b$$

$$2) \text{ מכ } b \perp R(A) \text{ מוגדר } b \in N(A^t) \text{ אז } A^tb = 0 \text{ !}$$

$$p = A(A^tA)^{-1}A^tb = 0$$

③ אם A מטריצה ריבועית הפיכה אזי A פורשת את $\mathbb{R}^n$ ומקבלים

$$p = A(A^T A)^{-1} A^T b = A A^T (A^T)^{-1} A^T b = b$$

↑
רק עבור מטריצת ריבועיות

איך אלו הם הוסיטוסים?

נניח $x \in \mathbb{R}^n$ ונבחר $f(x) = v^T x$

$$f(x) = v^T x = \sum_{i=1}^n v_i x_i$$

$$\frac{df}{dx} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

נניח $x \in \mathbb{R}^n$ ונבחר $g(x) = x^T M x$

$$g(x) = \sum_{i=1}^n \sum_{j=1}^n m_{ij} x_i x_j$$

$$\frac{dg}{dx} = \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n m_{1j} x_j + \sum_{j=1}^n m_{j1} x_j \\ \vdots \\ \sum_{j=1}^n m_{nj} x_j + \sum_{j=1}^n m_{jn} x_j \end{pmatrix} = Mx + M^T x$$

$$\frac{\partial g}{\partial x} = 2Mx \quad ! \quad M = M^T \quad \text{אם } M \text{ סימטרית אז}$$

הטלה ציבור אנליזה

נתון $p = A\bar{x}$ ונתון $\bar{x}$ רגור שמתחיל בנקודה $\bar{x}$ ונרצה למצוא את $\bar{x}$ המינימליזצור את $\|A\bar{x} - b\|^2$

$$E = \|A\bar{x} - b\|^2 = (A\bar{x} - b)^T (A\bar{x} - b) =$$

$$= \bar{x}^T A^T A \bar{x} - b^T A \bar{x} - \bar{x}^T A^T b + b^T b =$$

$$= \bar{x}^T A^T A \bar{x} - 2\bar{x}^T A b + b^T b$$

$$b^T A^T \bar{x} \in \mathbb{R}$$

(3)

אז (זכור

$$\frac{dE}{dx} = 0$$

הנקודות והמינוס

$$\frac{dE}{dx} = 2A^T A \bar{x} - 2A^T b = 0$$

כי $A^T A$ סימטרי

$$\Rightarrow A^T A \bar{x} = A^T b$$

$$\Rightarrow \bar{x} = (A^T A)^{-1} A^T b$$

ההנחה שיש הפיכה $A^T A$

ולו אותה תוצאה שקיבלנו קודם!

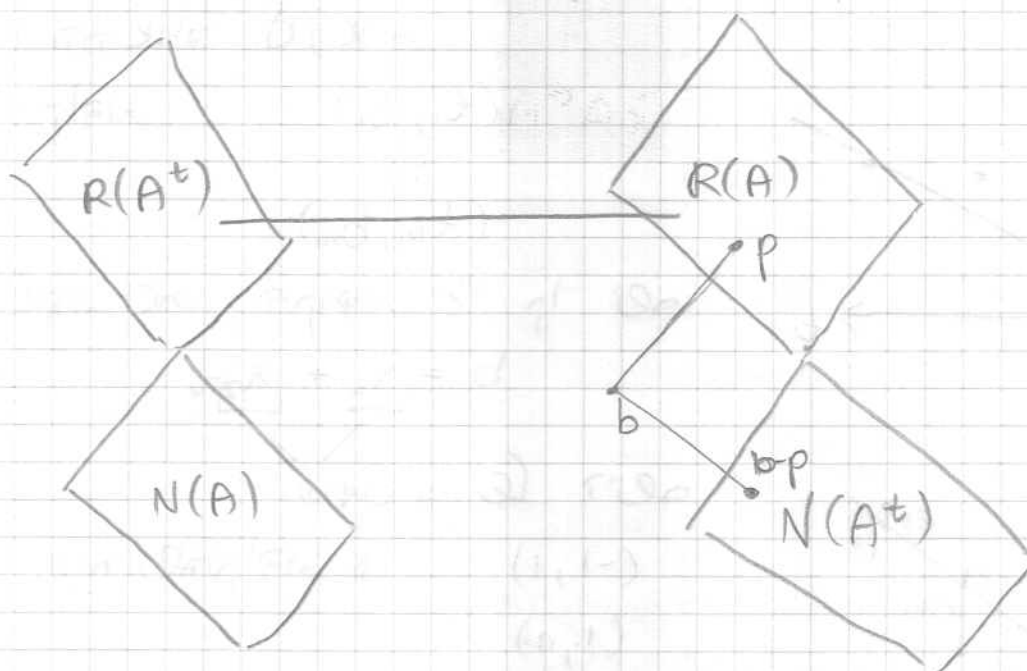
$$(A^T A)^T = A^T (A^T)^T = A^T A \quad \text{תכונה של } A^T A$$

$$\text{סימטרי } A^T A \quad \Leftarrow$$

$$!!! \text{ נכון } A^T A \neq A A^T \quad *$$

תלכורק :

$$\mathbb{R}^n \longrightarrow A \longrightarrow \mathbb{R}^m$$



$$R(A^t) \perp N(A)$$

ראינו שהצורה $R(A) = R(A^t)$ היא המרחב הריבועי.

ענה שרובים לפתור את $Ax = b$. ענה שההסתברות היא $\frac{1}{2}$

• $Ax = p$ - $\Leftrightarrow$ $p \in \text{Im}(A)$ - $\Leftrightarrow$ $p \in R(A)$ (p)

$$\Rightarrow b-p = b - A\bar{x} \in N(A^t)$$

$$\Rightarrow A^t(b - A\bar{x}) = A^t b - A^t A x = 0$$

$$\Rightarrow A^t A x = A^t b$$

$$\text{rank}(A^t A) = \text{rank}(A) \quad \text{8.7.8C}$$

$A^T A$ нь $n \times n$ хэмжээний $\text{rank}(A) = n$ нь эргэлт

$$x = (A^t A)^{-1} A^t b$$

הנורמ האוקלידי: אם A וריבועי - $N(A^t A) = N(A)$ (קריטריון)

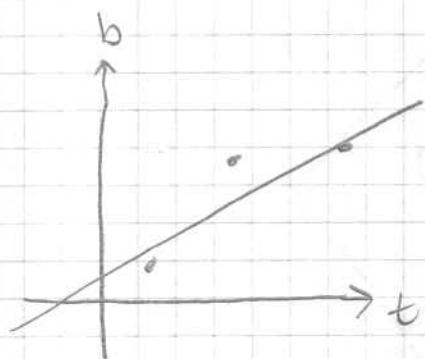
השטח (הכלל) של N (מ'צ'ים)

pc 1) $x \in N(A^t A)$ sk $x \in N(A)$ pc 1) callp ($\geq$)

$$A^t A x = A^t 0 = 0 \quad \text{so} \quad A x = 0$$

$$A^T A x = 0 \quad \text{ש"כ} \quad x \in N(A^T A) \quad - \text{ל} \quad (x) \\ x \in N(A) \Leftrightarrow Ax = 0 \Leftrightarrow \|Ax\|^2 = x^T A^T A x = 0 \Leftrightarrow$$

(iii)



דוגמה: נתונה סדרה
(t_1, b_1)
והנקודות

$$\vdots \\ (t_m, b_m)$$

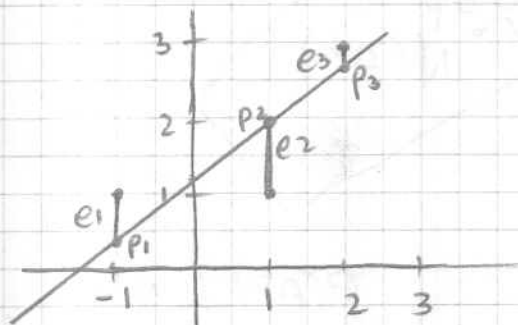
רוצים למצוא את הנקודות "ע" קו ישר

$$b = \underbrace{x_1}_{\text{אחידות}} + \underbrace{x_2 t}_{\text{אחידות}}$$

פרמטרים של הישר

דוגמה: נניח שנקודות הן

$$\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$



$$\begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

צריך לפתור את המשוואה

$\text{rank}(A) = 2$
אם ניקח מרחב כולל
שמישגנו.

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \quad \text{לפתור}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (A^T A)^{-1} A^T b = \begin{pmatrix} \frac{9}{7} \\ \frac{4}{7} \end{pmatrix}$$

$$b = \frac{9}{7} + \frac{4}{7}t$$

$$b = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$

$$p = A \bar{x} = \begin{pmatrix} 5/7 \\ 13/7 \\ 17/7 \end{pmatrix}$$

$$e = b - p = \begin{pmatrix} 2/7 \\ -6/7 \\ 4/7 \end{pmatrix}$$

(33) אד קורה האתה $\frac{1}{2} \text{rank}(A) \leq n$ - נ"ס -
 יש יחידים אלו x נסה x רק p - $Ax = p$ (נ"ס) אלו
 x $R(A^t)$ (ול) $N(A)$ ונניח $x = x_r + x_n$
 $y \in N(A)$ $A(x_r + y) = p$ אלו $x_r \in R(A^t)$ יש רק
 יחיד שמקיים זאת, ולכן (כזה) נמצא.
 אנו מחפשים x^+ רק p -

$$A\bar{x}^+ = p = \arg \min_{A\bar{x}} \|A\bar{x} - p\|^2 \quad (1)$$

(2) x^+ פיתרון של (1) אם ש"כ $R(A^t)$ -
 x^+ היא הפיתרון של (1) אם נורמה מינימלית

משפט SVD (Singular Value Decomposition)
 כל מטריצה A היא $m \times n$ ניתן לכתוב בצורה

$A = U D V^t$
 כאשר : U - אורתוגונלית $m \times m$
 D - אלכסונית $m \times n$ (ול) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$
 V - אורתוגונלית $n \times n$

ומתקיים : U - הוקטורים העצמיים של AA^t
 V - הוקטורים העצמיים של $A^t A$
 D - אלכסונית עם שורשים של הערכים העצמיים

של $A^t A$ (שהם גם ע"ע של AA^t)

ע"פ קובנציה הם מסודרים בסדר יורד
 (סדר העמודות של U ו- V מתאים לסדר הע"ע ב- D)

מטריצות אורתוגונליות

הצורה: קבוצה של וקטורים $e_1, \dots, e_n$ היא אורתוגונלית
 אם : $\langle e_i, e_j \rangle = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

הצורה: מטריצה אורתוגונלית ריבועית היא מטריצה ריבועית

עם עמודות אורתוגונליות

אין לה מומד למטריצה עם עמודות אורתוגונליות שפ"א
פ"א ריבועית.

נשים את לבם $Q \in M_k(\mathbb{R})$ מטריצה אורתוגונלית כל
 $Q^t Q = Q Q^t = I_k$

דבר $Q^t = Q^{-1}$.

אם Q מטריצה אורתוגונלית ויש לה רבועי-הבט
 $Q^t Q = I$.

טענה: כל מטריצה אורתוגונלית (היא לא רבועית) היא
 משנה את האורכים של הוקטורים.

הוכחה:
 $\|Qx\|^2 = (Qx)^t Qx = x^t Q^t Qx =$
 $= x^t I x = x^t x = \|x\|^2$

☺

טענה: יהי λ ערך של $A^t A$ אזי $\lambda \geq 0$

הוכחה: יהי v וֶקְטוֹר של $A^t A$ עם ערך λ כללי

$$0 \leq \|Av\|^2 = v^t A^t A v = v^t \lambda v = \lambda v^t v = \lambda \|v\|^2$$

☺

$$\lambda \geq 0 \quad \leftarrow \quad 0 \leq \|v\|^2$$

$r = \text{rank}(A)$ הוא מספר הוֶקְטוֹרִים הִרְצֵנִים של $A^t A$ שִׁבְצֵנִים

ממ 0 עד n . כל מחקים

$$A = U D V^t = \left(\begin{pmatrix} u_1 \end{pmatrix} \dots \begin{pmatrix} u_r \end{pmatrix} \begin{pmatrix} u_{r+1} \end{pmatrix} \dots \begin{pmatrix} u_m \end{pmatrix} \right) \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r & & 0 \\ & & & \ddots & \\ 0 & & 0 & & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

$u_1, \dots, u_r$ פורשים את $R(A)$ - הִתְחַבֵּר הֶעֱמִיזוֹת של A
 $u_{r+1}, \dots, u_m$ פורשים את $N(A^t)$ כי הם משלימים את הִתְחַבֵּר
 ומאונכים ל- $u_1, \dots, u_r$.

אם נסתכל על $A^t = V D^t U^t$ נקבל ש-

$v_1, \dots, v_r$ פורשים את $R(A^t)$

$v_{r+1}, \dots, v_n$ פורשים את $N(A)$

(34)

Linear Least Squares פתרונות

$$\min \|Ax - b\|^2$$

(1) נתונים

(2) נמצא x^+ כך שיהיה מינימום של $\|Ax - b\|^2$

פתרון:

$$\|Ax - b\|^2 = \|UDV^T x - b\|^2 =$$

$$\|U^T (UDV^T x - b)\|^2 = \|U^T U D V^T x - U^T b\|^2 =$$

$$= \|D V^T x - U^T b\|^2 = \|D y - U^T b\|^2 =$$

$$= \left\| \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r & & 0 \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_r \\ \vdots \\ y_m \end{pmatrix} - \begin{pmatrix} u_1^T b \\ \vdots \\ u_r^T b \\ \vdots \\ u_m^T b \end{pmatrix} \right\|^2 =$$

$$= \sum_{i=1}^r (\sigma_i y_i - u_i^T b)^2 + \sum_{i=r+1}^m (u_i^T b)^2$$

שורה 0- r עבורעל תלוי y

$$y_i = \frac{u_i^T b}{\sigma_i}$$

(1) $y_i = 0$ עבור $i > r$ וכן y רשום ב-
 נמצא מינימום של $\|Ax - b\|^2$

הערות: עבור $D = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r & & 0 \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix}^{m \times n}$ (עבור

$$D^+ = \begin{pmatrix} \sigma_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \sigma_r^{-1} & & 0 \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix}^{n \times m}$$

$$y = D^+ U^T b \quad \text{כא}$$

$$y = V^T x \Rightarrow V y = V V^T x = I x = x$$

$$\Rightarrow x = V y = V D^+ U^T b$$

y היה בסיס נמצא מינימום. הרבה V לא משנה את

הנורמה, כי V אורתוגונלית. אם הפתרונות לפי V הם

$$x^+ = V D^+ U^T b \in R(A^T)$$

הערה: $A^+ = V D^+ U^T$ - pseudo inverse (הרבה פעמים)

$$\begin{aligned} p = A x^+ &= A V D^+ U^T b = U D V^T V D^+ U^T b = \\ &= U D D^+ U^T b \end{aligned}$$

$$A = \omega D V^t$$

↓ אורח
↓ אורח
↓ אורח

$$Ax = b \quad \text{אנחנו}$$

הא

$$X^+ = V D^+ U^T b$$

$$p = Ax^+ =$$

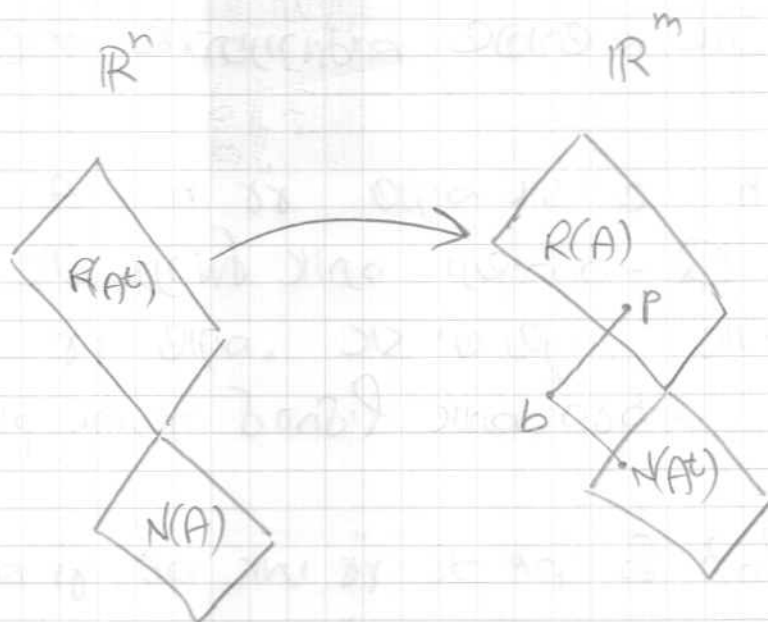
$$= \underbrace{U D V^T V D^+ U^T}_P b$$

נניח B - ארטימילר. SC היא מרחב של $\mathbb{R}^n$.

נניח X וקראר אירועים $B_1, B_2, \dots, B_n$

אזכור א' x אחרת המאובת ב B כחיסוס הסוגי + e

$\mathbb{R}^n$. ! - B' זוהי כדור, K - קוטר, N נקודות
הסגור $+ C$ וסדר B - B' נקודות.



פרק 10 (יה) SVD

צמח: זה של מחר צד סמכות כמתאים לז' שנים

אמאנוכא לך אלה.

הארכה: 99 ש"ס

$$Ax = \lambda_1 x$$

$$\lambda_1 \neq \lambda_2$$

$$Ay = \lambda_2 y$$

$$\lambda_1 x^t y = (Ax)^t y = x^t A^t y = x^t Ay = \lambda_2 x^t y \quad \triangleq \quad A = A^t$$

$$X^t Y = 0 \quad \text{ע"פ } (2), \quad \lambda_1 \neq \lambda_2 \quad \text{ע"פ } (1)$$

(ג) (נשים) אם שלה (סון) חם אם אמה ה- ל (הוא) אמה (ש)

משקלה: מטריצה סימטרית A יש מטריצה אורתוגונלית Q
כך ש- $A = Q \Lambda Q^{-1}$ ויש Q מטריצה של

על A : D - אלכסונית עם ע"ע של A על האלכסון,
 בלוח $AQ = QD$.

* אם יש שניים (או יותר) ו"ע נח"ם זה אומר ע"ע אס (יותר)
 ע"ע'ים אחרים בוקטורים אורתונורמליים שפורשים את אזור
 זה.

ה'רחה: אם יש ל- A ח ע"ע שונים אז יש ח ו"ע
 מאונכים זה לזה. אם נכנס אחרת ונשים ב- Q .
 יש בעיה אם אין ח ע"ע שונים. אז יש נק'ים שווים.
 במקרה זה נראה איך אפשר ע"ע'ים אחרים, בעזרת חלק
 גרם-שמית.

אם v_1, v_2 הם ו"ע עם אותו ע"ע אז הם ב-3
 ענפי שליה הם ו"ע עם אותו ע"ע.

$$q_i' = v_i - \sum_{j=1}^{i-1} (q_j^T v_i) q_j$$

הסילת v_i
מכיוון q_j

$$q_i = \frac{q_i'}{\|q_i'\|}$$

11

SVU בניה של

נניח ש- $n \geq m$. יהיו $\sigma_1^2, \dots, \sigma_r^2$ ע"ע של $A^T A$
 (יודע שהם חיוביים)

ע"פ ה- A קיימת מטריצה V אורתונורמלית (המורכבת
 מ"ע של $A^T A$) כך ש- $A^T A V = \Sigma V$ נא

$$A = QDQ^{-1}$$

$$A^T A = V \Sigma V^{-1}$$

$$V^{-1} A^T A V = \Sigma$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_r^2 & \\ & & & 0 \dots 0 \end{pmatrix}$$

$n \times n$ נא

$$\tilde{W} = A V \Sigma^{-\frac{1}{2}}$$

$m \times n$ $n \times n$ $n \times n$ ניקח

$$\Sigma^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \ddots & & \\ & & \frac{1}{\sigma_r} & \\ & & & 0 \dots 0 \end{pmatrix}$$

(36)

כ"ז

$$A = \tilde{U} \Sigma^{\frac{1}{2}} V^t = \tilde{U} \Sigma^{\frac{1}{2}} V^t$$

↓
אוטומטית
הפיכה

$$\tilde{U}^t \tilde{U} = \Sigma^{-\frac{1}{2}} V^t A^t A V \Sigma^{-\frac{1}{2}} = \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 & & \\ & & & 0 & \dots & 0 \end{pmatrix}$$

⌊ ⌋
נסתב על $\tilde{U}^t \tilde{U}$ כ"ז
נרחיב אותם לבסיס של $\mathbb{R}^m$
אוטומטית "א".

$$U = \left(\begin{pmatrix} u_1 \end{pmatrix} \dots \begin{pmatrix} u_r \end{pmatrix} \begin{pmatrix} \text{הרחבה} \\ \text{ל} \\ \text{בסיס} \\ \text{אוטומטית} \end{pmatrix} \right)$$

$$\Sigma^{\frac{1}{2}} \rightarrow D = \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & \dots & 0 \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \quad \text{נ"קח}$$

$m \times n$

$$\Rightarrow A = \tilde{U} \Sigma^{\frac{1}{2}} V^t = U D V^t$$

אם $n > m$ אז נעשה את הפירוק של A^T והפסק.



SV D של A

$$A = U D V^t = \begin{pmatrix} u \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r & & \\ & & & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} v^t \end{pmatrix} =$$

$$= \sum_{i=1}^r u_i \sigma_i v_i^t = \sum_{i=1}^r \sigma_i u_i v_i^t \approx \sum_{i=1}^q \sigma_i u_i v_i^t =: A_q$$

↓
 $q < r$ בחתך
 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$

זה משו שקטור עצמים של המטריצה - נווה ל- A
לו תמונה ואנחנו רוצים להעביר אותה את הפחית היטב,
אז בשביל להעביר את A_q צריך $q_m + q_n + q$ היטב.

ניתוח ספקטראלי של נתונים

אנטימציה - מכוונת את אימץ הקצוה ציבן ניתוח של מרחב הפוטנציאל.
יתרונות:

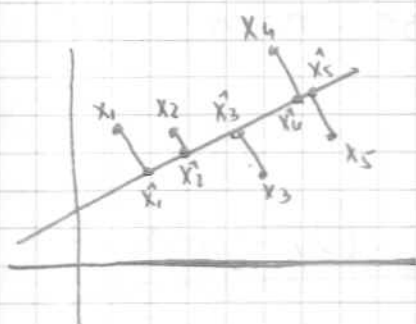
- (1) עמידה במימד יונו (מק) יותר ונוסף הכללה יותר טובה.
- (2) היבנה יותר טובה של הפוטנציאל של הקצוה.

דרכים להחלפת מידע

- (1) feature selection - בחירה של תת קבוצה של פוטנציאלים
- (2) feature extraction - יצירה של פוטנציאלים חדשים ממק
פוטנציאלים קיימים ועמידה עם מרחב מרחב של פוטנציאלים חדשים

PCA- Principal Component Analysis

יש מ קבוצת בלתי R^n וקצוה לראו אומן מרחב מרחב
(מק) R^q , $q < \min(m, n)$ והמרחק בין הנקודות
 $\hat{x}_1, \dots, \hat{x}_m$ והוקבוצת המקוריות יהיה מקינאלי:
$$\min \sum_{i=1}^m \|x_i - \hat{x}_i\|^2$$



קוצא + יש כאן אימץ מ'צו . אכן אומן
קוצים לאסד (מה שפחות אכן ממשללים
אם המרחקים .

נתונה מ'צו $A = \left(\begin{pmatrix} x_1 \end{pmatrix}, \dots, \begin{pmatrix} x_m \end{pmatrix} \right)$

מחפשים תת מרחב U של R^n שאליו נ'סל אה הוקבוצו .

נונת (מחזר) U עם וקטורים אורתונורמליים (ה'ים אלה ונורמלי) $(U - I)$

ה'סלסו ק $\hat{x} = Uy$ ע'ר $y \in R^n$ ע'לשו.

מחפשים מסד'צת הסלס P שנתנת $\min \|x - \hat{x}\| = \min \|x - Px\|$

נסתב על $x - \hat{x} = x - Uy \in N(U^t)$ $U^t(x - Uy) = 0 \Leftrightarrow$
 $\hat{x} = Uy = UU^t x \Leftrightarrow U^t x = U^t U y = y \Leftrightarrow$

(3)

: צדקה ל א

$$\min_u \sum_{i=1}^m \|x_i - \hat{x}_i\|^2 = \min_u \sum_{i=1}^m \|x_i - uu^t x_i\|^2 =$$

s.t. $u^t u = I$

$$= \min_u \|A - uu^t A\|_F^2$$

s.t. $u^t u = I$

צדקה ל Frobenious ל א

$$\|B\|_F^2 = \sum_{ij} b_{ij}^2$$

$$\|B\|_F^2 = \text{tr}(BB^t)$$

: צדקה: ל

$$BB^t = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{m1} \\ \vdots & & \vdots \\ b_{1n} & \dots & b_{mn} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m b_{1i}^2 & & \\ & \ddots & \\ & & \sum_{i=1}^m b_{ni}^2 \end{pmatrix}$$

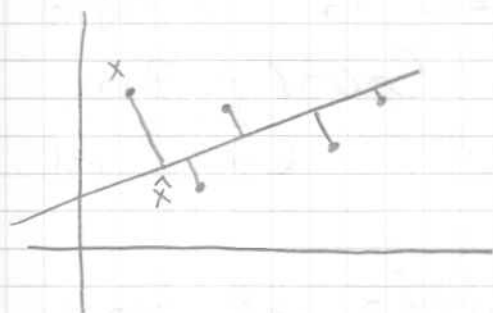
(11)

$$\arg\min_u \|A - uu^t A\|_F^2 = \arg\max_u \text{trace}(u^t A A^t u) = : \text{צדקה}$$

$$= \arg\max_u \frac{1}{2} \|u^t A\|_F^2 = \arg\max_u \frac{1}{2} \sum_{i=1}^n \|u_i^t A\|^2$$

PCA - Principal Component Analysis

תכלית: היו לנו וקטור אנחנו רוצים ישר את שדה הנתון
מ"ן הנקודה זה אולי שיהיה היש מניח.



$$A_{n \times m} = \left(\begin{pmatrix} x_1 \end{pmatrix}, \dots, \begin{pmatrix} x_m \end{pmatrix} \right)$$

ההסטה הן $\hat{x} = U U^T x$
ומכונה מטריצת וינג
מחפשים הסטה עם ע"ן מניח

$$\min_{s.t.} \|A - U U^T A\|_F^2$$

$$U^T U = I_{q \times q}$$

$$\arg \min_U \|A - U U^T A\|_F^2 = \arg \max_U \frac{1}{2} \|U^T A\|_F^2 = \text{C} \text{ (נכח)}$$

$$\left(= \arg \max_U \frac{1}{2} \sum_{i=1}^q \|u_i^T A\|^2 \right)$$

(F-נורמה של הריבוע של וינג)

הנחת: מתקיים $x = \hat{x} + (x - \hat{x})$
ההצבה של הסטה נכח $\hat{x} \perp x - \hat{x}$
פירוק $\|x\|^2 = \|\hat{x}\|^2 + \|x - \hat{x}\|^2$
מתקיים $\|A\|_F^2 = \sum \|x_i\|^2$
וכן $\|A\|_F^2 = \|U U^T A\|_F^2 + \|A - U U^T A\|_F^2 =$
 $= \|U^T A\|_F^2 + \|A - U U^T A\|_F^2$

כי U מטריצה עם
אנחות אורתונורמלית
אז ההפכה לא
משנה את אורכי הווקטורים

אם $\|A\|_F^2$ זמין אז U - ה- $\|A - U U^T A\|_F^2$ דומה למינימום של $\|U^T A\|_F^2$

זכשו נפתר את העזר האופטימלי.

$$\max_{\substack{u^t \\ u^t u = I}} \frac{1}{2} \|u^t A\|_F^2 = ?$$

לראות יש פה q^2 אופציות אבל למעשה יש פחות
אופציות כי $u_i^t u_j = 0$ זה אומר אופציות $u_j^t u_i = 0$
אז צריך להסתכל רק על הארנסון וכל מה שמעליו!
אז יוצא

$$\frac{q^2 - q}{2} + q = \frac{q^2 + q}{2}$$

נכתוב את הלינאר של ההערה

$$\mathcal{L} = \frac{1}{2} \sum_i \|u_i^t A\|^2 - \sum_{j \geq i} \lambda_{ij} (u_i^t u_j - \delta_{ij})$$

$$= \frac{1}{2} \sum_i u_i^t A A^t u_i - \sum_{j \geq i} \lambda_{ij} (u_i^t u_j - \delta_{ij})$$

$$\frac{\partial \mathcal{L}}{\partial u_i} = A A^t u_i - \sum_{j \geq i} \lambda_{ij} u_j = 0 \quad i = 1, \dots, q$$

$$A A^t u_q = \lambda_{qq} u_q \quad \text{אם } i = q \text{ אז}$$

λ_{qq} זה $A A^t$ על u_q ונצטרך לה

$$(*) \quad A A^t u_{q-1} = \lambda_{q-1, q} u_q + \lambda_{q-1, q-1} u_{q-1} \quad \text{אם } i = q-1$$

(כבר נשמע נכון) u_q^t

$$u_q^t A A^t u_{q-1} = \lambda_{q-1, q} \underbrace{u_q^t u_q}_{= \|u_q\|^2 = 1} + \lambda_{q-1, q-1} \underbrace{u_q^t u_{q-1}}_{= 0}$$

זה נשמע נכון
transpose
הערה

הם
אורתוגונליים

$$u_{q-1}^t A A^t u_q = \lambda_{qq} u_{q-1}^t u_q = 0$$

$$(*) \quad \text{אז } \lambda_{q-1, q} = 0 \quad \Leftarrow$$

$$A A^t u_{q-1} = \lambda_{q-1, q-1} u_{q-1}$$

$$\lambda_{q-1, q-1} \text{ זה } A A^t \text{ על } u_{q-1} \quad \Leftarrow$$

(39)

נשים האותו עשן :

$$AA^t u_i - \sum_{j \geq i} \lambda_{ij} u_j = 0$$

 $k > i$ ורק u_k^t - ו (הוא) אנון

$$\underbrace{u_k^t AA^t u_i}_{(כאילו)} = 0 - \sum_{j \geq i} \lambda_{ij} \underbrace{u_k^t u_j}_{= \delta_{kj}} = 0$$

$$\Rightarrow \lambda_{ik} = 0 \quad \forall k > i$$

$$\Rightarrow AA^t u_i = \lambda_{ii} u_i$$

אם λ_{ii} u_i הוא ו' על AA^t ו' λ_{ii} .

$$: \frac{1}{2} \sum_i \|u_i^t A\|^2 \quad \text{על} \quad \text{השיון}$$

$$\frac{1}{2} \sum_i \|u_i^t A\|^2 = \frac{1}{2} \sum_i u_i^t AA^t u_i =$$

$$= \frac{1}{2} \sum_i u_i^t \lambda_{ii} u_i =$$

$$= \frac{1}{2} \sum_i \lambda_{ii} \underbrace{\|u_i\|^2}_{=1} = \frac{1}{2} \sum_i \lambda_{ii}$$

אם פוקציה המטה שליון אלקס היא סכום העצ
 ו הוקטורים שמחנן. אם זה שלבים עצום זה
 אחרת א הוקטורים הם העצם העצמיים הכי גדולים
 בומר $\max_{\substack{U^t U = I \\ \text{s.t.}}} \frac{1}{2} \sum \|u_i^t A\|^2$ מתקדם עמר $u_1, \dots, u_g$ להם

ו' המתאמים g -י AA^t ו' העצמים ביותר.

נשים זה שלפרישה $U^t U = I$ הוא המצבה מסו' N

אמטיות א יחידה הפיתרון. אם יש ח' ע' שונים אם

ב מרחב עצמי N הוא ממוצע 1 והוא הנכחד הפיתרון

יחיד. אם יש ע' במרחב עצמי N -1 הוקדא א

יחידה אחרת יחידה...

קשר SVD - r

$A = U D V^t$ כאשר U ו' V של AA^t ונייג' - $\tilde{U}$ הפיתוח של קשיר האורטונורמליזציה כל

$$\tilde{U}^t A = \tilde{U}^t U D V^t = \tilde{U}^t \sum_{i=1}^r u_i \sigma_i v_i^t = \sum_{i=1}^r \tilde{u}_i \sigma_i v_i^t$$

$$\tilde{U} \tilde{U}^t A = \sum_{i=1}^r u_i \sigma_i v_i^t$$
 ההטענה הן

פרט סכני:

אם $m > n$ החיפוש של ו' x של AA^t יתכן
 עדיף יקר. אולם $A = U D V^t$ ולכן עבור
 r וקטורים ראשונים $U = A V D^{-1}$ כאשר
 V ו' x של $A^t A$ וזה יתיר קט לחישוב.

דוגמה לשימוש ה- PCA

ניקח 100 תמונות של פרצופים בגודל 50×50 . נ"ל
 אותם כווקטורים האורך 2500.

$$A = \begin{pmatrix} \begin{pmatrix} 100 \\ \vdots \end{pmatrix} & \dots & \begin{pmatrix} 100 \\ \vdots \end{pmatrix} \end{pmatrix}$$

2500x100

נרץ PCA של A ונקח מכיוונים עיקריים של הפיצולים.

פירוש סטטיסטי

חברה של אנשים ואינסיים הסטטיסטיקה:

$$E X = \mu_x = \sum_x x p(x) \approx \frac{1}{m} \sum_{i=1}^m x_i$$

 שיתוף
 אותם (תונים)
 זרימה x

(50)

הצורה מטריציונית: $A = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ בהינתן מטריצה

נורה A את התחלק μ ב μ סדר

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \approx \frac{1}{m} \sum_{i=1}^m \begin{pmatrix} x_i \end{pmatrix} = \frac{1}{m} A \vec{1}$$

שונות משותפת: Covariance

$$\text{Cov}(X_1, X_2) = \sigma_{12} = E((X_1 - EX_1)(X_2 - EX_2)) = E(X_1 X_2) - EX_1 \cdot EX_2$$

שונות: Variance

$$\text{Var}(X) = EX^2 - (EX)^2 = \text{Cov}(X, X) = \sigma_X^2$$

הצורה מטריציונית: נניח שהקצו n ממוכלול, $\vec{\mu} = 0$

$$\text{Cov}(X_1, X_2) = E(X_1 X_2) = \text{כא}$$

$$= \sum_{x_1, x_2} x_1 x_2 p(x_1, x_2)$$

covariance $\rightarrow$ צורה

$$\Sigma = \frac{1}{m} A A^t = \frac{1}{m} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \begin{pmatrix} x_1^t & \dots & x_m^t \end{pmatrix} =$$

$$= \frac{1}{m} \begin{pmatrix} \sum_{k=1}^m x_i^k x_j^k \end{pmatrix} \approx \left(\text{Cov} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right) = \begin{pmatrix} \sigma_{ij} \end{pmatrix}$$

הסקרה: הנתנה $A \cdot \vec{1} = 0$ - בומר שים הנתים הם

שונה שונה $A A^t$ היא קירוב Σ

מטריצה $\rightarrow$ Covariance - הן הקואורדינטות $1, \dots, n$ $\rightarrow$ x - וס.

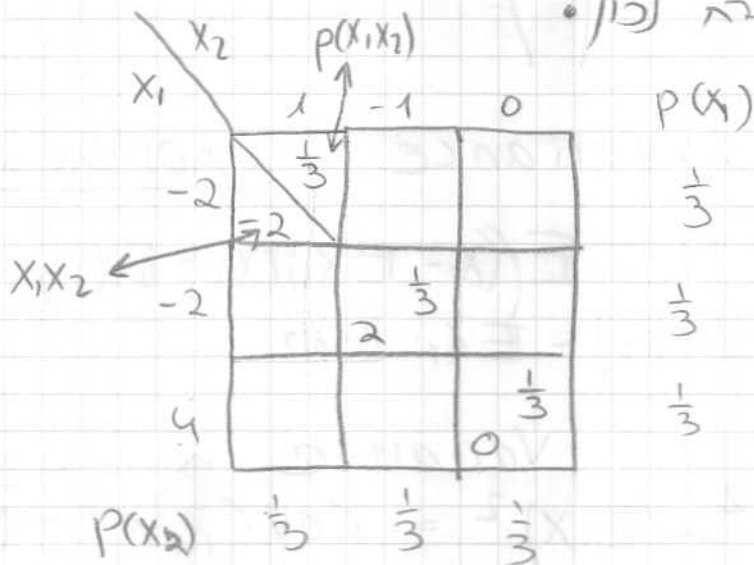
הסתברות

* אם $\text{Cov}(X_1, X_2) = 0$ אז X_1, X_2 נקראים

unrelated

* אם X_1, X_2 הם ללא $\text{Cov}(X_1, X_2) = 0$

* והפך לא נחשב ככזה:



* אם $P(X_i, X_j) = P(X_i)P(X_j)$ לכל $i \neq j$

אז $P(X_1, \dots, X_n) = P(X_1) \dots P(X_n)$

הם נקראים independent

דוגמה ים פירמידה: פאה אחת בחלוקה ב

אחת בצורה r

אחת יריקה g
אחת בשלושה הרים

אם הם independent אז

$$P(r) = P(b) = P(g) = \frac{1}{3}$$

והסתברותם $P(r, g) = \frac{1}{4} = P(r)P(g)$ הם ללא

אם הם ללא

$$P(r, g, b) = \frac{1}{4} \neq P(b)P(g)P(r) = \frac{1}{8}$$

(4)

PCA (פירוט)

ניקח $y_i = u^t x_i$ - קואורדינטה של הvektor של x_i במרחב u

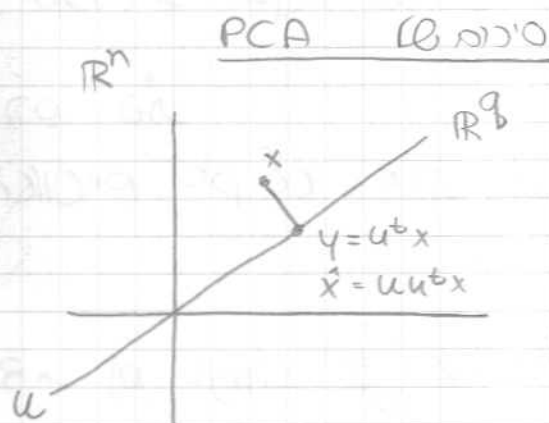
$$Y = U^t A = \begin{pmatrix} (y_1) & \dots & (y_m) \end{pmatrix}$$

: Y is covariance matrix

$$Y Y^t = U^t A A^t U = \begin{pmatrix} \sigma_1^2 & \dots & \sigma_g^2 \end{pmatrix}$$

$$A = \begin{pmatrix} | & | & | \\ x_1 & \dots & x_m \\ | & | & | \end{pmatrix}$$

$$Y = U^T A$$



PCA מתפל אחת הסלף u אחת q שיש k

התכונה הבאה שכן שקולה:

(1) המרחק בין ההסלף $\hat{x}$ למקור x יפיה קטן

$$\min_{s.t.} \|A - U U^T A\|_F^2$$

$$U^T U = I$$

(2) ההסלף כמה שיותר ארוכה

$$\max_{s.t.} \|U^T A\|_F^2 = \max_{s.t.} \|Y\|_F^2$$

$$U^T U = I$$

(3) (הנחה) הנקודות מאונכלות, סוגר סכום השורות של A שווה ל-0,

קואורדינטות של ההסלף יהיו סלף אחידות והשורות של ההסלף תהיה מקסימלית

דא $(2 \leftrightarrow 3)$?

covariance
ההסלף Y

$$= Y Y^T = U^T A A^T U = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_q^2 \end{pmatrix}$$

AA^T של u כי u ו u של

$$\|Y\|_F^2 = \text{trace}(Y Y^T) = \sum_{i=1}^q \text{Var}(Y_i)$$

קואורדינטה i של Y

אם כדי שהשורה תהיה מקסימלית שמתנו חצים לה

$$\max \text{trace}(Y Y^T)$$

ל (2)

$$s.t. U^T U = I$$

$$U^T A A^T U \leftarrow$$

כמה מקסימלית בחיים u כי u ו u של

אחריו ש-

מ"מ סלתי מתואמים ∇ מ"מ סלתי כזוהר ∇ מ"מ סלתי

טכניקה מ"מ גאומטרים סלתי מתואמים הם סלתי
 מתקדמת PCA אופטימלי עבור מ"מ גאומטרים אך לא אופטימלי
 באופן כללי.

התפלגות נורמלית (גאומטרי) זה מיוצגת עם תוחלת מן ומטריצה

Σ covariance

$$p(\vec{x}) = p\left(\begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix}\right) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\mu)^T \Sigma^{-1}(\vec{x}-\mu)}$$

אם הקואורדינטות סלתי מתואמות אז Σ אורסונלית,
 כלומר $\Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}$

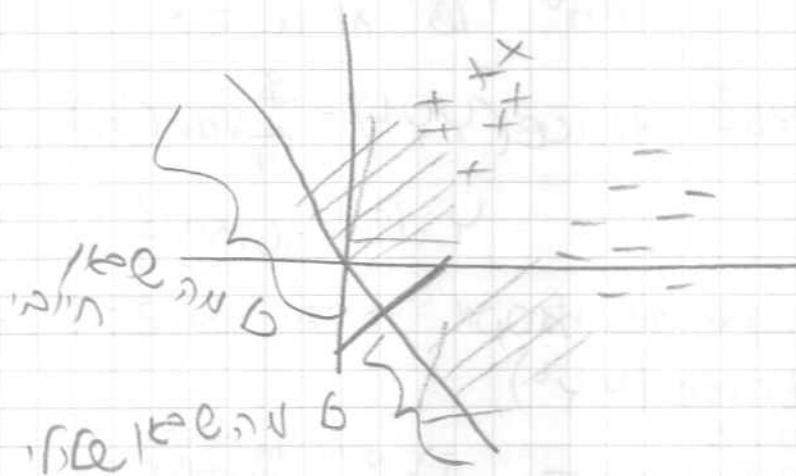
$$p(\vec{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} \sum \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2} =$$

$$= \prod_{i=1}^n \frac{1}{(2\pi)^{1/2} \sigma_i} e^{-\frac{1}{2} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2} = \prod_{i=1}^n p(x_i)$$

∇ $x_1, \dots, x_n$ הם סלתי תלויים.

LDA - Linear Discriminant Analysis

יש לנו קבוצה עם תיבות ואנחנו רוצים להכריז אותן למחלקות
 (אפשר למחלקות 1) כדי שסלתי ההכרזה עדיין
 אופטימלית והיה להפרדה ברורה.



(3)

הפרדה ליניארית

$$f(x) = w^T x + b$$

$$y = \text{sign}(f(x))$$

$f(x)$ = linear discriminant function

המשוואה $w^T x + b = 0$ מפרידה את המישור

$w^T x$ בזווית הנורמלית אל מישור w .

אם $w^T x \neq 1$, אז w אינו נורמלי, לכן

$$x^T w = \frac{w^T w}{\|w\|} x = (w^T w)^{-1} w^T x$$

הקו $f(x) = 0$ הוא המישור שממנו w .

יש כאן שני קווים מקבילים זה לזה.

$$f(x) = 0 \text{ מישור מקביל}$$

אם w הוא וקטור

שקבוע זהו קו

!- b הוא ה-

decision boundary
threshold

כל w שפירוט אסוף.

אנחנו מנסים מישור (קו) המפריד בין "+"

ל- "הם כמה שיותר טובה"

האלגוריתם למינימום

(1) המרחק בין המרכזים של "+" ו-" הוא אחר המרחק

הקטן ביותר יהיה כמה שיותר קטן.

(2) שונות של המסור ה- "+" ו-" הם סכום מרחקים.

הקטן ביותר יהיה כמה שיותר קטן.

$$\max_w \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$



יבין μ_i הממוצע של נקודות מחלקה C_i
 $\hat{\mu}_i$! מרכז הנקודות אחרי ההפצה
 יהי m_i מספר הנקודות במחלקה C_i .

$$\hat{\mu}_i = \frac{1}{m_i} \sum_{x_j \in C_i} x_j = \frac{1}{m_i} \sum_{x_j \in C_i} w^t x_j = w^t \left(\frac{1}{m_i} \sum_{x_j \in C_i} x_j \right) = w^t \mu_i$$

$$(\hat{\mu}_1 - \hat{\mu}_2)^2 = (w^t (\mu_1 - \mu_2))^2 = w^t (\mu_1 - \mu_2) (\mu_1 - \mu_2)^t w \quad \leftarrow$$

$$\hat{\sigma}_i^2 = \frac{1}{m_i} \sum_{x_j \in C_i} (\hat{x}_j - \hat{\mu}_i)^2 = \frac{1}{m_i} \sum_{x_j \in C_i} (w^t x_j - w^t \mu_i)^2 =$$

$$= \frac{1}{m_i} \sum_{x_j \in C_i} w^t (x_j - \mu_i) (x_j - \mu_i)^t w =$$

$$= w^t \left(\frac{1}{m_i} \sum_{x_j \in C_i} (x_j - \mu_i) (x_j - \mu_i)^t \right) w = w^t \Sigma_i w$$

↓
 covariance
 של נקודות מחלקה i

נשים לב שיש לנו כאן אינרציה בלבד (לא מסתכלים על הצורה)
 האופטימיזציה כי זה מנסה למצוא

$$S_b = (\mu_1 - \mu_2) (\mu_1 - \mu_2)^t \quad (\text{יציב})$$

$$S_w = \sum_{i=1}^2 \Sigma_i = \Sigma_1 + \Sigma_2$$

אם קטור האופטימיזציה הוא

$$\max_w \frac{w^t S_b w}{w^t S_w w} \quad \text{Rayleigh's quotient}$$

אם λ (צורה) $\max J(w)$ ו- λ צורה

$$\frac{dJ(w)}{dw} = \frac{2S_b w (w^t S_w w) - 2S_w w (w^t S_b w)}{(w^t S_w w)^2} = 0$$

$$\Rightarrow \overbrace{S_b w (w^t S_w w)}^{EIR} = \overbrace{S_w w (w^t S_b w)}^{EIR}$$

$$\Rightarrow S_b w = \frac{w^t S_b w}{w^t S_w w} \cdot S_w w = J(w) \cdot S_w w \quad (*)$$

כל נקודה w היא ו- λ (צורה) (S_b, S_w) (צורה)

44

אם $S_w^{-1} S_b W = J(W) W$ אז S_w מטריצה הפיכה
 בואו, W היא ו' של $S_w^{-1} S_b$ הממלאם דע"ע תבי גרמל
 (חשור ל) - (*):

$$\underbrace{(u_1 - u_2)(u_1 - u_2)^T W}_{\text{מספר (מטריצה)}} = J(W) S_w W$$

$$\Rightarrow J(W) S_w W = \gamma(u_1 - u_2)$$

אם S_w הפיכה אז

$$W = \underbrace{\frac{\gamma(W)}{J(W)}}_{\text{סקלר}} \underbrace{S_w^{-1} (u_1 - u_2)}_{\text{וקטור}}$$

אמרו שהיגד של W עם מספר על פתרון הקטור ולכן
 מה שמעבר וכן היא W בכיוון של $S_w^{-1} (u_1 - u_2)$
 חא משנה מה באורך שלו. בואו, דר רבי (נחמון) שלא משנה א
 פתרון הקטור $W = S_w^{-1} (u_1 - u_2)$

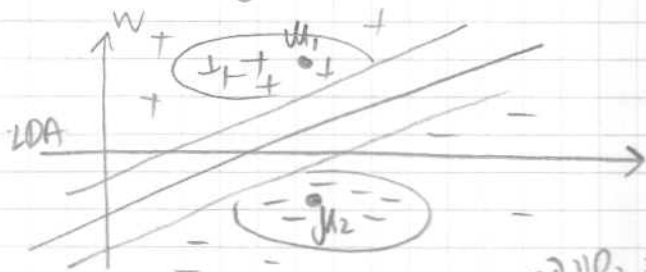
לכור שהמטרה שלנו ביותר למצוא את
 מציאו את W . נותן למצוא את b .

אבל (המטרה) ביום למחזור בלוי: ורזה $W^T (x - \mu) = 0$
 בואו $0 = W^T x - W^T \mu$ b היא $W^T \mu$
 (μ זה הנתון של $\underline{b}$ (ווקטור).
 אסרבים W (קטור)

$$x^T S_w^{-1} (u_1 - u_2) - \mu^T S_w^{-1} (u_1 - u_2) = 0$$

SVM מן LDA

הפתרון וכו'ים אחר שנים. זה ביטל LDA מנה התפלגות
 גאומטרי ו-SVM דע"ע!!
 אם אס התפלגות והתחלה
 הן צבוע להתפלגות והתחלה
 הנותנה שזה גאומטרי גרף תוצאה שונה.



LDA מניח התפלגות גאוסית אחת לכל מילה
SVM לא מניח התפלגות גאוסית אחת לכל מילה
פחיתות

אם להתפלגות היא גאוסית אז שני האלמנטים
הם נותנים את אותה התוצאה.

"הסב" הפאזיות" וקצות וסב' באמצע עינאית

נג' שיטתן מסויסה שמייצגת את השואר $Ax=b$ $\begin{pmatrix} \end{pmatrix}_n = \begin{pmatrix} \end{pmatrix}_m$ $\begin{bmatrix} \end{bmatrix}_{m \times n}$
אמ"ש יותר משואר אנקלוויס סבי' שאין פיתרון

אז מחפשים פיתרון קרוב שיתר טוב, בומר מחפשים $\min_x \|Ax-b\|^2$

ראינו לפיתרון הוא $x=(A^t A)^{-1} A^t b$ (אם $A^t A$ קיימת)

המטריצה $A^t=(A^t A)^{-1} A^t$ קראי pseudo-inverse. המטריצה

$A^t A(A^t A)^{-1} A^t$ היא מטריצה הסדר. היא מקדמת וקטורים ומסירה

אזת את-אחת שפרט ע"י המצדד של A (רבי ש- $A^t A$)

תהיה תפסה לרוב שפזמדות שלה אחרים בת"ס)

אם אמ"ש פתור משואר אנקלוויס $Ax=b$ $\begin{pmatrix} \end{pmatrix}_n = \begin{pmatrix} \end{pmatrix}_m$ $\begin{bmatrix} \end{bmatrix}_{m \times n}$ $Ax=b$ (undetermined) $\begin{bmatrix} \end{bmatrix}_{m \times n}$

המקרה זה הפיתרון (הוא לא יחיד) (אמקום לרוב) - כי יש פתור
משואר אנקלוויס

העיר פאזיות לה מאפיינים. נסתכל אמיל על $\mathbb{R}^3$ - לה אחת

אנטימאטיבי אבן. אפס. נניח שיש יש המרחב ומחפשים אז

המרחק שלו אחרת. המרחק שלו אחרת. לה המרחק המינימלי

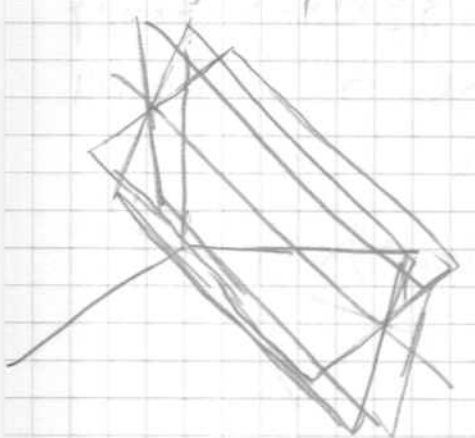
אחרת. מקדמת לשלבי הילס הלה. גרד

שני זה המרחק הכי גדול בין המרחק

למרחבים שזורים ביש זה (פשוט)

לרוב אפזמין אז לה. רשומים

המישורים מקדמים ומקדמים אחרת



מקרה על $A_{2 \times 3} = \begin{bmatrix} a_1^t \\ a_2^t \end{bmatrix}$ ומחפשים $Ax=b$ בומר

$$\begin{cases} a_1^t x = b_1 \\ a_2^t x = b_2 \end{cases}$$

המשואר (כזה מצדד) שני מישורים. אז
 x (מצד) המרחק שלהם

אנחנו נחפש את הפיתרון x עם הנורמה הקטנה ביותר (פשוט כי

$$\|x\| = \sqrt{x^T x} \quad \min_{Ax=b} \|x\| = \min_{Ax=b} \sqrt{x^T x}$$

זה מה שפשוט!)

אמנם תמונה הישנה של x יש את אותו אורך

אנחנו מחפשים את הקצרה ביותר שמתקנת $Ax=b$ וזהו הפתרון.

אנחנו מחפשים את הקצרה ביותר שמתקנת $Ax=b$ וזהו הפתרון.

$$\min_{Ax=b} \frac{1}{2} \|x\|^2 = \min_x \max_{\lambda} \left\{ \frac{1}{2} x^T x - \lambda^T (Ax-b) \right\}$$

אנחנו מחפשים את הקצרה ביותר שמתקנת $Ax=b$ וזהו הפתרון.

אם $Ax \neq b$ נניח $\lambda = 0$ אז

$$\max_{\lambda} \left\{ \frac{1}{2} x^T x - \lambda^T (Ax-b) \right\} = \infty$$

כי ניקח λ ככל שנקרא $Ax-b$ ונקרא λ ככל שנקרא $Ax-b$.

אם $Ax=b$ אז λ יכול להיות כל דבר וזה לא משנה.

אם $Ax=b$ אז λ יכול להיות כל דבר וזה לא משנה.

אם $Ax=b$ אז λ יכול להיות כל דבר וזה לא משנה.

אם $Ax=b$ אז λ יכול להיות כל דבר וזה לא משנה.

אם $Ax=b$ אז λ יכול להיות כל דבר וזה לא משנה.

$$\min_x \max_{\lambda} \left\{ \frac{1}{2} x^T x - \lambda^T (Ax-b) \right\} = \max_{\lambda} \min_x \left\{ \frac{1}{2} x^T x - \lambda^T (Ax-b) \right\}$$

אם $Ax=b$ אז λ יכול להיות כל דבר וזה לא משנה.

$$x - A^T \lambda = 0 \Rightarrow x^* = A^T \lambda$$

$$\left(\begin{aligned} \frac{\partial}{\partial x} \left(\frac{1}{2} x^T x \right) &= x, & \frac{\partial}{\partial x} (x^T A x) &= (A + A^T) x \\ \frac{\partial}{\partial x} (x^T a) &= a \end{aligned} \right)$$

$$\Rightarrow \min_x \left\{ \frac{1}{2} x^T x - \lambda^T (Ax-b) \right\} = -\frac{1}{2} \lambda^T A A^T \lambda + \lambda^T b$$

$$\max_{\lambda} \left\{ -\frac{1}{2} \lambda^T A A^T \lambda + \lambda^T b \right\}$$

$$x^* = (A A^T)^{-1} b \Leftrightarrow -A A^T \lambda + b = 0$$

$$x^* = A^T \lambda^* = A^T (A A^T)^{-1} b$$

$\Leftarrow$

(46)

אחרת

$$\min_{Ax=b} \frac{1}{2} \|x\|^2$$

אם פתרון היעיל

$$x = A^t(AA^t)^{-1}b \quad \text{אם } A \text{ זרוע } x \text{ זה מקיים את המשוואה}$$

$$(n \leq m) \quad AA^t \text{ הפיכה } (m \leq n)$$

זרוע אחרת יוצרים לפתור את $Ax=b$ כיון אם $m < n$ וזו

$$m < n$$

אם הפיתרון מצאנו בעזרת הפארוואל וזה היה חזק קטן לא תמיד אפשר

להשלים בפארוואל את זה היה אפשרי וזה טוב.

זרוע אחרת $\min_{Ax=b} \|x\|$ בשיטה אחרת - נש'נוש הוא למצוא

אם מחקים מנימוקים של ישירות ומרחקים למסלולים של אישורים.

כמו קופס, אם הפוסליו

$$\min_{Ax=b} \|x\| = \max_{\lambda} \min_x \{ \|x\| - \lambda^t(Ax-b) \}$$

בשיטה אחרת $\min_x \{ \|x\| - \lambda^t(Ax-b) \}$ אפשר למצוא את

$$\arg \min_x \{ \|x\| - \lambda^t Ax \} \quad (\text{שדור ההבדל הנגזר הוא רק בקבוע})$$

ואם אחרת לא יתאים וגזר (מחלקים נשים) שם ש.

$$(*) \quad \min_x \{ \|x\| - \lambda^t Ax \} = \begin{cases} -\infty & \|A^t \lambda\| > 1 \\ 0 & \|A^t \lambda\| \leq 1 \end{cases}$$

אם שנויה את זה (גידו איך) נש'נוש בזהו אם זה פשוט

$$\max_{\|A^t \lambda\| \leq 1} \{ \lambda^t b \} \quad \min_x \{ \|x\| - \lambda^t(Ax-b) \} = \lambda^t b$$

$$\max_{\|A^t \lambda\| \leq 1} \lambda^t b$$

וכן שווה ל-

$$\|A^t \lambda\| = 1 \quad (x) \quad (x) \quad (x)$$

$$x = \mu A^t \lambda \quad \mu > 1 \quad (x) \quad (x) \quad (x)$$

$$\|x\| - \lambda^t A^t \lambda = \mu \cdot \ell - \mu \ell^2 = \mu \ell (1 - \ell) \xrightarrow{\mu \rightarrow \infty} -\infty$$

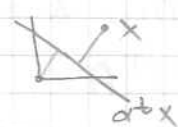
$$x^t A^t \lambda \leq \ell \|x\| \quad \text{אם } \ell \leq 1 \quad \text{אם } \ell > 1 \quad \text{אם } \ell > 1 \quad \text{אם } \ell > 1$$

$$\|x\| - \lambda^t A^t \lambda \geq \|x\| - \ell \|x\| = (1 - \ell) \|x\| \geq 0$$

$$(0) \quad \text{אם } \ell > 1 \quad \text{אם } \ell > 1 \quad \text{אם } \ell > 1$$

$$\min_{Ax=b} \|x\| = \max_{\lambda} \frac{\lambda^t b}{\|A^t \lambda\|}$$

$$\frac{\lambda^t b}{\|A^t \lambda\|} \quad \text{אם } \alpha^t x = \beta \quad \text{אם } \alpha^t x = \beta \quad \text{אם } \alpha^t x = \beta$$



למה $\max_{\|A^t x\| \leq 1} x^t b = \max_{\|A^t x\| \leq 1} \frac{x^t b}{\|A^t x\|}$ זה נובע מההומוגניות ב- λ . אנוני

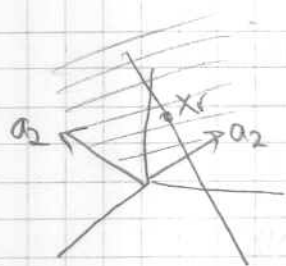
חזית ההקדון את המונה ולהקטין את המכנה. הבלם ההומוגניות ב- λ אם $\|A^t x\|$ לא שווה ל-1 אפשר יחסו אג λ כשצורה שקטין את הנדנה וההומוגניות זה נשאר אותו ביטוי.

L

זכשו נסמן ש- $\frac{x^t b}{\|A^t x\|}$ זה הביטוי של חיתק של על מישור מהבטול. חיתק של על מישור $\beta = x^t a$ נבחרת עם $\frac{|\beta|}{\|a\|}$. אק חיתק שור אנוני $Ax = b$ זה חיתק של מ על-אישורים: $a_i^t x = b_i$ $1 \leq i \leq m$ (הישורים של A). $x^t A x = x^t b$ זה על-מישור שנקרש ע"י b על-אישורים האלה. זו קומבינציה ענאית של $a_i^t x = b_i$ (מישור המקבילים הם λ). אז $x^t A x = x^t b$ הוא על מישור מחתק.

מהראשית הוא $\frac{x^t b}{\|A^t x\|}$ וכו' נחשש נוחה של מה שאנו בהתחלה: החיתק של קו מהבטול ב- $\mathbb{R}^m$ הם החיתק הוקסידל מהראשית על-אישורים שחורים בקו הזה. (נרשם זהו צאן הנצטר ז"ל שיקדים מאלהבה)

למה $x = A^t (A A^t)^{-1} b$ (סתם למשל ק) $\mathbb{R}^3$

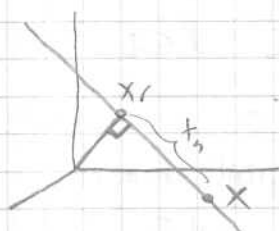


$A_{2 \times 3} = \begin{bmatrix} a_1^t \\ a_2^t \end{bmatrix}$ וסתם על מרחב הישורים $R(A^t)$. $Ax = b$ זה ישר (חיתק של של מישור ענאית) שכל אחת מהן (נונה מישור). אם היש הזה הם לא בחוק המישור אז הם חותק אותו בדיוק הנקודה אחת. בואו קיים $x \in R(A^t)$ ונצד רק ש- $Ax_r = b$.

יש לנו $x = A^t (A A^t)^{-1} b$ חתק אחד $x \in R(A^t)$ וחצו שני. $Ax = b$ אלא יש רק (קורה אחת שמקיימת את זה) וזה x_r . x בלי אפשר לחתק $x = x_r + x_n$ כול $x_n \in N(A)$ וכל $Ax = Ax_r = b$.

יפוג ע- $R(A^t) \perp N(A)$ בומר $x_r \perp x_n$

ע x זו בדיוק הנקודה שבה מחתקם החיתק מהבטול.

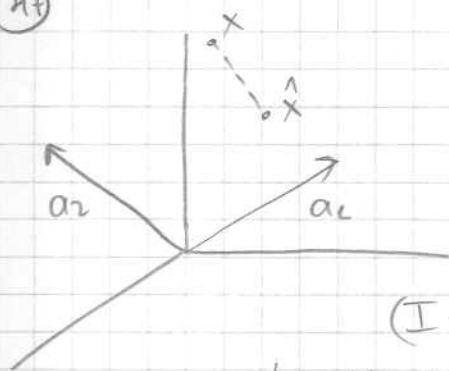


סימא $x = A^t (A A^t)^{-1} b$ ע"י הטרור

כוכים נמצא $x \in R(A^t)$ החקיים $Ax = b$.

(ניג שיש לנו x ונוכים נמצא את ההיטל שלו על $R(A^t)$ מחצו ההטלה היא $P = A^t (A A^t)^{-1} A$.

(47)



$$x - \hat{x} = (I - P)x$$

$x \neq 0$ אם $R(A^t)$ - שנייה - ווקטור

$$A(I - P)x = 0 \quad \text{אם}$$

$$Ax - APx = Ax - AA^t(AA^t)^{-1}Ax = 0$$

$$(I - P)x = x - \hat{x} = 0 \quad \text{אם } x \in R(A^t) \quad \text{אם}$$

$$x = A^t(AA^t)^{-1}b$$

$$\Leftrightarrow x - \underbrace{A^t(AA^t)^{-1}Ax}_b = 0 \quad \text{אם}$$

כלי להוכחה

