Im2 ex2

313410497 [illegible]

1.1 Soulutions of the Normal Equations

הוכח $\ker(X) = \ker(X^TX)$

נניח בשלילה ש $\ker(X) \neq \ker(X^TX)$ ולכן קיים $0 \neq u \in \mathbb{R}^d$

כך ש- $X^TXu = y \in \mathbb{R}^n \in Im(X^TX)$ וגם $Xu = 0 \in \ker(X)$

אבל: $Xu = 0$ ולכן $X^T(Xu) = \langle X, 0 \rangle = 0$

כיון שהמכפלה הפנימית (של 0 וקטור) משמאלה X היא תמיד 0

[illegible]

1.2 $Im(A^T) = \ker(A)^\perp$

הינה כאשר המטריצה הנתונה A מקיימת

צריך להוכיח הכלה דו כיוונית

$Im(A^T) \subseteq \ker(A)^\perp$

יהי $u \neq 0$ כך ש $A^Tu = w$ כלומר $w \in Im(A^T)$

יהי $v \neq 0$ כך ש $Av = 0$ כלומר $v \in \ker(A)$

נראה ש [illegible] $\langle w, v \rangle = 0$ $w \in Im(A^T) \subseteq \ker(A)^\perp$

ולכן $w^Tv = \langle w, v \rangle = \langle A^Tu, v \rangle = (A^Tu)^Tv$

$= u^TAv = u^T0 = \langle u, 0 \rangle = 0$

[illegible] A^{nxm}

[illegible]

$\dim(Im(A)) + \dim(\ker(A)) = n \Leftrightarrow \dim(Im(A)) = \dim(\ker(A)^\perp)$

וסמן $\dim(Im(A))$ ב m [illegible]

$m = \dim(col(A)) = \dim(row(A^T)) = \dim(col(A^T)) = \dim(Im(A^T))$

1.3

[illegible] X [illegible] $y = Xw$

$y \perp \ker(X^T)$ [illegible]

X [illegible]

$Xw = y$ [illegible] $y \in Im X \Leftrightarrow y \in \ker(Y^T) \Leftrightarrow y \perp \ker(Y^T)$

$Im(A) = \ker(A^T)^\perp \Rightarrow Im(A^T) = \ker(A)^\perp$ [illegible] 2 [illegible]

(1.1.ד) נסתכל על המשוואות הנורמליות $X^TXw=X^Ty$
הוכחנו כי למשוואות הנורמליות יש פתרון יחיד (אם X^TX הפיכה)
אבל אם יש אינסוף פתרונות (אם X^TX לא הפיכה) במקרה?

נפרק למקרים: אם X^TX הפיכה אז $\hat{w}=[X^TX]^{-1}X^Ty$
עבור המקרה היחיד הזה, ...
אם X^TX לא הפיכה, נוכיח כי ל-X^TX ... מסעיף 1.3 נקבל ש:
ל-$X^TXw=X^Ty$ יש ∞ פתרונות $\iff X^Ty \perp \ker(X^T) = X^TXw \perp \ker(X)$
כלומר ניתן להראות שעבור $u\in\ker(X)$ מתקיים ש: $\langle X^TXw, u\rangle=0$ (*)
נשים לב שמסעיף 1.1 מתקיים: $\ker(X^TX)=\ker(X)$ לכן
$u\in\ker(X) \iff Xu=0 \Rightarrow w^TX^TXu=0 \iff (X^TXw)^Tu=0$
ואני יודע ש: $0=\langle w, X^TXu\rangle$ (*)
משני התנאים יש אינסוף פתרונות כי $\langle X^TXw,u\rangle=0$ כנדרש

1.1.2 Projection Matrices

$V\subseteq\mathbb{R}^d$, $\dim(V)=k$, ויהי $v_1,\dots,v_k$ בסיס אורתונורמלי של V
נגדיר את מטריצת ההטלה האורתוגונלית $P=\sum_{i=1}^{k}v_iv_i^T$ כאשר זוהי מכפלה חיצונית

(a) show that P is symmetric $\Rightarrow P^T=P$

$$P^T=\left(\sum_{i=1}^{k}v_iv_i^T\right)^T=\sum_{i=1}^{k}v_i^T\otimes v_i=\sum_{i=1}^{k}v_iv_i^T=P$$

מכפלה חיצונית היא קומוטטיבית

(b) הוכח שהערכים העצמיים של P הם 0 או 1 ויש $v_1,\dots,v_k$
הם הו"ע המתאימים לערכים העצמיים של 1
עבור λ ע"ע (נראה) שלכל v_j נקבל שהוא ו"ע
מתקיים: $Pv_j=\lambda_j \Leftarrow \sum_{i=1}^{k}v_i\otimes \underbrace{v_i^T\cdot v_j}\iff\sum_{i=1}^{k}v_i\delta_{ij}=v_j$ (*)

$\delta_{ij}\begin{cases}1 & i=j\\ 0 & \text{אחרת}\end{cases}$

ממשיך קיבלנו ש: $Pv_j=\lambda v_j=v_j \Rightarrow \lambda=1$ ע"ע

לכן $\lambda=1$ ו-v_j הו"ע המתאים לו

(c) <u>show that $\forall v \in V$, $Pv = v$</u>

$v_1 \dots v_k$ בסיס אורתונורמלי של V אז לכל $v \in V$ קיים $a_i \in \mathbb{R}$ ש

$$v = \sum_{j=1}^{k} a_j v_j$$

לכן

$$Pv = \sum_{i=1}^{k} v_i v_i^T \sum_{j=1}^{k} a_j v_j = \sum_{i=1}^{k}\sum_{j=1}^{k} a_j v_i v_i^T v_j = \sum_{i=1}^{k}\sum_{j=1}^{k} a_j v_i \delta_{ij} =$$

[illegible]

$$= \sum_{j=1}^{k} a_j v_j = v$$

(d) $P^2 = P$

אם אנחנו יודעים ש P סימטרית ולכן היא לפי פירוק ספקטרלי
קיים U מטריצה אורתוגונלית, ו D מטריצה אלכסונית ש
זה תלוי הערכים של A. $A = UDU^T$ לכן:

$$P^2 = (UDU^T)(UDU^T) = U\underbrace{D D}_{}U^T = UDU^T = P$$

$U^T = U^{-1}$ אורתוגונלית, אלכסונית ו-0 או 1

(e) Prove that $(I-P)P = 0$

מספיק ש:

$$(I-P)P = 0 \Leftrightarrow P - P^2 = 0 \Leftrightarrow P = P^2$$ — הוכחנו ב-d

1.3 Least Squares

6. Show that if X^TX is invertible, the general solutions we derived in the recitation = the solution we saw in class

הנחה ש X^TX הפיכה נשתמש בה
מהנוסחה אז X^TX הפיכה אז:

$$X^TX\hat{w} = X^Ty \Leftrightarrow (X^TX)^{-1}(X^TX)\hat{w} = (X^TX)^{-1}X^Ty \Leftrightarrow \boxed{\hat{w} = (X^TX)^{-1}X^Ty}$$ (1)

$$\hat{w} = V\Sigma^{+}U^Ty$$

קיבלנו הנוסחה X^TX הפיכה אז
נגיע תמיד עם פתרון היחיד, נשתמש ב-SVD

$$X = U\Sigma V^T \rightarrow X^TX = (U\Sigma V^T)^T(U\Sigma V^T) \Leftrightarrow V\Sigma^TU^TU\Sigma V^T \Leftrightarrow V\Sigma^T\Sigma V^T =$$

$$= VDV^T$$

כאשר D מטריצה אלכסונית עם [illegible] V מטריצה אורתוגונלית.
נשים לב ש X^TX סימטרית ⇒ $X^TX = VDV^T$ פירוק ספקטרלי

מטריצה

(3) נזכר שמתקיים $X^TX = VDV^T$ [illegible]

$$[X^TX]^{-1} = [VDV^T]^{-1} = VD^{-1}V^T$$

כאשר X^TX הפיכה [illegible]

[illegible] D^{-1} [illegible]

כעת, נזכר מה שקיבלנו ב-(A) ונקבל $X^T = (U\Sigma V^T)^T = V\Sigma U^T$

$$[X^TX]^{-1}X^T = VD^{-1}\underbrace{V^TV}_{I}\Sigma U^T = VD^{-1}\Sigma U^T = V(\Sigma^T\Sigma)^{-1}\Sigma U^T =$$

$$V\Sigma^{+T}\Sigma^{+}\Sigma U^T$$

ומכך ש- $\Sigma^{+T} = \Sigma^{+}$ [illegible] $\Sigma^{+}\Sigma = I_n$ [illegible]

$$\hat{w} = [X^TX]^{-1}X^Ty \Rightarrow \hat{w} = V\Sigma^{+}U^Ty$$

נקבל ש- $\blacksquare$

(7) Show that X^TX הפיכה $\Leftrightarrow$ $\mathrm{span}\{x_1, \dots, x_m\} = \mathbb{R}^d$

שאלה 1.1 ; X^TX הפיכה [illegible]

X^TX הפיכה $\Leftrightarrow$ $\mathrm{Ran}(X^TX) = d$ $\Leftrightarrow$ $\{0\} = \ker(X^TX)$ $\Leftrightarrow$ $\{0\} = \ker(X)$

$\Leftrightarrow \mathrm{Rank}(X) = d \Leftrightarrow \mathbb{R}^d = \mathrm{span}(x_1, \dots, x_m)$

כאשר נזכור שהמרחב של X = $\mathrm{span}\{x_1, \dots, x_m\}$ [illegible]

(8) [illegible] X^TX לא הפיכה [illegible] ∞ פתרונות (או 0)

[illegible] $\hat{w} = X^{+}y$ הוא הפתרון [illegible] L_2 מינימלית.

[illegible] $\hat{w}$ [illegible] $\|\hat{w}\| \le \|w\|$

נסמן: $X = U\Sigma V^T$ פירוק ה-SVD של X

U ו- V הן מטריצות אורתוגונליות [illegible]

[illegible] $v \in V$ [illegible] $\|v\| = \|Uv\|$

[illegible] $U^T = U^{-1}$, $V^T = V$ [illegible] $U^TU = Id$, $V^TV = Id$

[illegible] $\hat{w} = \hat{y} =$ [illegible]

$$\hat{w} = \sum_{i=1} x_i y_i = X^Ty$$

[illegible] $\hat{w}$ [illegible]

$$\hat{w} = \sum_{i=1}^{d} u_i \sigma_i v_i^T y \in \arg\min \|y - Xw\|^2 = \arg\min \|Xw - y\|^2$$

נניח: $\mathrm{ran}(X^TX) = \mathrm{rank}(X) = r$

המשך $\rightarrow$

לכן:

$$\hat{w} = \|U\Sigma V^T \hat{w} - y\|^2 \overset{(*)}{=} \|U^T(U\Sigma V^T(w)) - U^T(y)\| = \|\Sigma V^T w - U^T(y)\|$$

כאשר המעבר נכון בגלל [illegible] אורתו.

$$= \|\Sigma V^T w - U^T(y)\| = \underbrace{\sum_{i=1}^{r}(\sigma_i v_i^T w_i - u_i^T y)^2}_{(B)} + \sum_{j=r+1}^{m} u_j^T y$$

$\sigma_{i>m} = 0$ [illegible]

נסמן ב-(A), ניתן להסיק מכאן ש-$\hat{w} \in \arg\min \|Xw - y\|^2$ "מסתמך" על r הקורדינטות הראשונות של w ב-(B) מכיוון ש[illegible] השאר [illegible] הפתרון היחיד?

[illegible] שדרגת X היא r ולכן לא נוכל לומר שיהיו לו r שורות (ועמודות) בלתי תלויות לינארית.

[illegible] עבור (A) יהיה הפתרון היחיד. [illegible]

[illegible] m עד $\hat{w}$ ה- אפסים [illegible] (*) [illegible] הקורדינטות של y [illegible] ($u_{r+1}, \dots, u_m$)

[illegible] Σ [illegible] ($\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & 0 & \end{bmatrix}$)

שממנו $\hat{w}$ ייראה כך:

$$\begin{bmatrix} \sigma_1 v_1 \\ \vdots \\ \sigma_r v_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

ולכן: [illegible] $\bar{w}$ אחר מקיים:

$$\|\bar{w}\| = \sum_{i=1}^{d} \bar{w}_i^2 = \sum_{i=1}^{r} \bar{u}_i^2 + \sum_{i=r+1}^{d} w_i^2 \geq \sum_{i=1}^{r} \bar{w}_i^2 + 0 = \sum_{i=1}^{r} \hat{w}_i^2 = \|\hat{w}\|^2$$

IML -Ex 2

אביב אוחיון 313410458

שאלה 12:

בשאלה זו הורדתי את הפיצרים של id, date, long, lat.

הסיבור לכך הם שעבור הפ'יצר של הid, הערך הזה לא משחק חשיבות בשובי של הדירות אלא רק מייצג מי הבעלים של הדירה. עבור ה פ'יצרים long, lat אלו פ'יצרים אשר מתארים בסופו של דבר קורדינטיה במרחב, אשר שקולה פחות או יותר למה שמתאר הפ'יצר zipcode רק שבמקרה זה זה פ'יצר בודד ועל כן יותר יעיל להשתמש בו מאשר בהם. לגבי הפ'יצר של התאריך, אחרי מעבר על ערכי התאריכים ראיתי שהם רק עבור השנים 2014 2015, מהסתכלות זריזה באינטרנט מחירי הדירות לרוב משתנים (ברמה המייחסת חשיבות) אחת ללפחות 7 שנים, ועל כן העדפתי להוריד את העמודה הזו גם גם על מנת להקל על החישובים.

שאלה 13:

הפיצ'ר שהחלטתי לבצע עבורו one hot encoding היה הפ'יצר של ה zipcode. לפ'יצר זה אין יחס סדר ועל כן חישובים ניומרים התלויים בו צפויים לקבל תוצאות משונות, עם זאת עקב מה שהסברתי להעיל בשאלה 12, יש לו עדיין חשיבות אז לא רציתי להוריד אותו לגמריי כמו פ'יצרים אחרים.

שאלה 15:

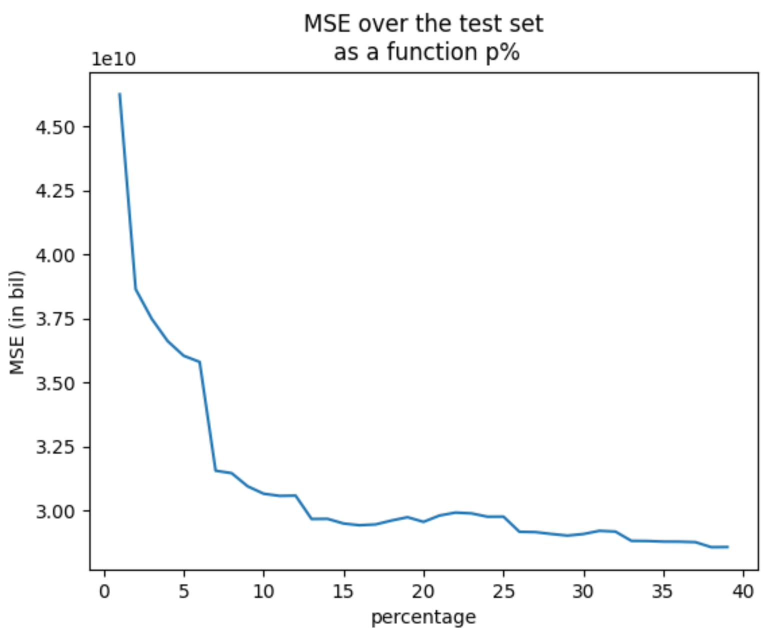
מה שנוכל ללמוד מהקשר בין כמות הפיצרים לבין ערכים הסינגולרים הוא על היחס הלינארי ביניהם, כלומר, ככל שהערכים הסינגולרים של מטריצה הפיצרים, דגימות יהיו יותר קרובים ל 0, זה יצביע על כך שעמודות הטבלה יהיו תלויות לינארית אלו באלו. מצב של תלות לינארית מושלמת קשה להשגה עקב הרבה גורמים חיצוניים ולכן גם "כמעט תלות" תצביע לנו על יחס הדוק בין עמודות הטבלה (הפיצ'רים התלוים למודל).

על כן, ככל שיותר דגימות יתנו לנו ערכים סינגולרים יותר קטנים השואפים ל אפס, כמות המידע שנוכל להסיק תקטן, כיוון כאילו פיצ'ר אחד לא כזה השפיע על התוצאה של הפ'יצר השני ולהיפך (זה בדיוק התלות הלינאירת), לעומת זאת, פיצ'רים בעלי ערכים סיגנולרים מאוד גדולים ורחוקים זה מזה יצביע לנו על קשר

הרבה יותר חזק אחד לשני כי תזוזה של תוצאת פ'יצר אחד תשפיע מאוד על תוצאת הפ'יצר השני וליהפך.

נסיק מכך שכיוון שככל שיש יותר פיצרים (במקרה של איך שאני ממשתי 15 עמודות) ככה הערכים הסינגורלים ישאפו ל 0 כלומר תיהיה בניהם תלות לינארית. מהגרף שמתאר כך ניתן לראות שכבר מהפ'צר ה 3 או 4 הערכים שואפים ל 0 ולכן כבר יש "כמעט תלות לינארית" בניהם משמע זה יהיה המקרה הסינגורלי (המקרה הלא הפיך)

<u>שאלה 16</u>

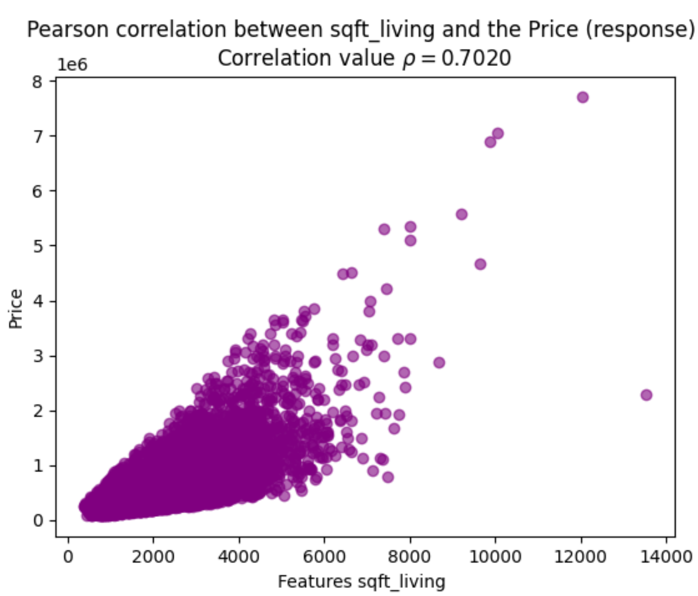


ניתן לראות שככל שאנחנו מגדילים את אחוז השורות שאנחנו "מאמנים" ככה ערך הסטייה MSE יורד גם כן ומהר מאוד מתייצב על סביבות B25. ההתכנסות הזו קוראת כבר לאחר שאנחנו דוגמים כרבע מהשורות מהחלק במודל שעליו אנחנו מתאמנים, נסיק מכך שנוכל כבר מכמות יחסית קטנה של דגימות עבור הערכים האלו לקבל מידע יחסית נאמן ובעל סטייה קבוע מהערכי ה y המקוריים.
הערה, עקב חוסר מקום לאלוקציה של זיכרון, האיטרציה רצה עד 40 אחוז, עם זאת לאחר שניסתי כמה פעמים להרציץ את החישובים עם אחוזים התחלים מ 50אחוז ולא מ 1 וכו, נוכחתי לגלות שההתיצבות נישארת זהה והגרף יתכלס פחות או יותר לMSE שהוא B25.

שאלה 17

ראשית, אנחנו נוכל להסיק מהגרפים המתארים את קורולציית פרסון מידע על כמה מידע נוכל להסיק (אם בכלל) בין פיצ'ר מסויים לעמודת המחירים. הדרך שבא זה הקורולציה מתבטאת בגרף על ידי זה שמצד אחד המטריצת השונות המשותפת מתארת לנו "לאיזה כיוון" הדגימות מתקדמות ביחס לy וקטור ומצד שני על ידי חלוקה בכפל בין השונות הנורמאלית של כל וקטור "מנרמלת" את היחסים בניהם ומדגימה עד כמה "חזק" הקשר בניהם. מצב זה מבטא בדיוק את הקשר הלינארי בין בכמה חזק פ'יצר כלשהו לוקטור הresponse. כאשר הערכים נעים בין [-1,1] (הקשר החזק ביותר 1 והחלק ביותר -1]
על כן, גרפים בהם ניתן לראות שככל שציר ה x עולה ככה עולה ה y (או יורד) בצורה יחסית רציפה הם גרפים אשר נוכל להסיק מהם מידע כי הקשר קיים שם. מנגד, גרפים אשר לא מראים בהכרח קשר (לינארי) בין הגדילה של x לבין איך מושפע ה y לא יעזרו לנו כל כך כי לא נוכל להסיק אם בכלל קיים קשר בין הזוג הנ"ל.

קשר טוב – בעל ערך 0.7 ניתן לראות שאפשר להעביר קו ישר יחסית בקלות בין כל הערכים הכי גבוהים בעל שיפוע "חזק" מאוד



קשר רע - בעל ערך 0.39, ניתן לראות שאין יחס לינארי בין התקדמות ה ערכי ה x לבין ערכי ה y הם פחות או יותר זהים לא משנה מה הפי'צר x יהיה ולכן הישר המחבר בניהם יהיה יחסית מאוזן