שיטת הריבועים הפחותים

בהינתן n נקודות במישור $(x_1,y_1),\dots,(x_n,y_n)$ כש-x_i שונים זה מזה, יש פולינום אחד ויחיד P ממעלה $\le n-1$ המקיים $\forall i P(x_i)=y_i$ לפולינום הזה קוראים פולינום האינטר-פולציה של הנקודות הנ"ל. (במקרה הפשוט ביותר $n=2$ מדובר בישר המוגדר ביחידות ע"י שתי נקודות במישור). יש גם נוסחה מפורשת לפולינום הזה - נוסחת האינטרפולציה של לגרנז':

$$P(x)=\sum_{i=1}^{n} y_i \prod_{j\ne i} \frac{x-x_j}{x_i-x_j}$$

לעיתים קרובות, במגוון רחב של יישומים השאלה המעניינת היא אחרת. מעוניינים בפולינום Q ממעלה $\le m$ כש-m קטן (אולי בהרבה) מ-n, מובן שלא ניתן יהיה לקיים עוד $P(x_i)=y_i$ לכל i. אולם היינו רוצים ש-$Q(x_i)$ יהיו קרובים למספרים y_i. איך נכמת את מושג הקירבה הזה?

מטרתנו היא שהווקטור $Q(x_1),\dots,Q(x_n)$ יהיה קרוב לווקטור $(y_1,\dots,y_n)$. לכן, עלינו להגדיר תחילה איך מודדים מרחק בין ווקטורים. לשם כך דרוש לנו מושג של אורך של ווקטורים. המושג הטכני הוא נורמה. מסמנים ב-$\|z\|$ את הנורמה ("האורך") של הווקטור $z\in\mathbf{R^n}$. הדרישות מנורמה הן:

1. לכל $z\in\mathbf{R^n}$ מתקיים $\|z\|\ge 0$ ושויון אם"ם $z=0$.
2. לכל λ ממשי $\|\lambda z\|=|\lambda|\|z\|$.
3. $\forall u,v \quad \|u+v\|\le\|u\|+\|v\|$

יש מגוון רחב של נורמות חשובות ומעניינות. החשובה ביותר (ומוכרת לכם) ללא ספק היא הנורמה האוקלידות המסומנת $\|\cdot\|_2$ ומוגדרת כך: $\|z\|_2=\sqrt{\sum z_i^2}$ עוד שתי דוגמאות חשובות הן נורמת l_1 המוגדרת כ-$\|z\|_1=\sum|z_i|$ ונורמת l_∞ שהגדרתה: $\|z\|_\infty=\max_i|z_i|$

באופן כללי יותר נורמת l_p מוגדרת כך: $\|z\|_p=(\sum|z_i|^p)^{1/p}$ לכל $1\le p<\infty$. שימו לב שהנורמה האוקלידית זהה לנורמת l_2 (ולפעמים גם נקראת כך).

יש עוד נורמות מעניינות אחרות. על התורה של מרחבים נורמיים ניתן ללמוד בקורסים מתקדמים באנליזה מתימטית ובמיוחד באנליזה פונקציונלית. נחזור אם כן לשאלתנו. אנו מחפשים כזכור פולינום Q ממעלה $\le m$ כך שהווקטור $(Q(x_1),\dots,Q(x_n))$ קרוב לווקטור הנתון $(y_1,\dots y_n)$. מדד הקירבה שבו נשתמש יהיה זה של הנורמה האוקלידית. אפשר לשאול כאן מדוע העדפנו את נורמת l_2 על פני נורמות אחרות. הסיבה העיקרית היא שזהו המקרה הקל ביותר לפתרון. ניתן לפתור את הבעייה באופן יעיל גם בנורמת l_∞ אולם לשם כך צריך לפתח את התורה של תכנון ליניארי (שאותה ניתן ללמוד בקורס אופטימיזציה דיסקרטית). מלבד שני מקרים אלה ווריאציות פשוטות שלהם אין מקרים רבים שניתן לפתור באופן יעיל.

נשוב, אם כן לשאלתנו היסודית ונחפש פולינום כנ"ל. פולינום ממעלה $\le m$ ניתן לרשום כ: $Q(x)=c_0+c_1x+\dots c_mx^m$ ולכן הבעיה היא למצוא את המקדמים הטובים ביותר $c_0,\dots,c_m$. את השאלה ניתן אם כן לנסח מחדש כך:

תהיה $A_{n\times(m+1)}$ מטריצת ונדרמונה המוגדרת ע"י

$$a_{ij}=x_i^j \quad (1\le i\le n \ \ ,0\le j\le m)$$

אז הבעייה היא למצוא ווקטור c כך שהווקטור Ac קרוב ככל האפשר, בנורמת l_2 לווקטור y במילים אחרות, מכיוון שהפונקציה $t\to t^2$ עולה בתחום $t>0$, השאלה שקולה למציאת ווקטור c כך ש:

$$\min_c \|Ac-y\|_2^2$$

אין סיבה שנגביל את עצמנו לבעייה המקורית, אם כן. ניתן לעיין בשאלה הכללית יותר: נתונה מטריצה ממשית A ווקטור y ואנו רוצים לפתור את הבעייה

$$\min_c \|Ac-y\|_2^2$$

אם A מטריצה ריבועית לא סינגולרית ניתן לפתור את הבעייה בכך שמוצאים c המקיים $Ac=y$ והמינימום הוא 0. המקרה המעניין הוא זה שבו יש ב-A יותר שורות מעמודות.

ניתן להתבונן בשאלה זו גם מנקודת מבט נוספת, גיאומטרית. יהיה L המרחב הליניארי הנפרש ע"י העמודות של המטריצה A. מטרתנו למצוא את הנקודה הקרובה ביותר (במרחקי l_2) ב-L לנקודה הנתונה. נקודת המבט הגיאומטרית מבהירה מייד את התשובה - הפתרון יהיה ההיטל של הנקודה y על המרחב L.

לאחר כל הדיון שניהלנו ניגש עתה לפתור את הבעייה. אנו מעוניינים במינימום של

$$||Ac-y||_2^2=\sum_i(\sum_j a_{ij}c_j-y_i)^2$$

נגזור לפי c_r ונשווה את הנגזרת ל-0 נקבל:

$$\forall r \quad 2\sum_i a_{ir}(\sum_j a_{ij}c_j-y_i)=0$$

כלומר

$$\forall r \quad \sum_i a_{ir}\sum_j a_{ij}c_j=\sum_i a_{ir}y_i$$

אבל $\sum_i a_{ir}y_i$ שווה לקואורדינטה ה-r ב-A^Ty. את אגף שמאל נרשום כ:

$$\sum_j c_j\sum_i a_{ir}a_{ij}=\sum_j c_j\sum_i a^T_{ri}a_{ij}=\sum_j c_j(A^TA)_{rj}$$

דהיינו, הקואורדינטה ה-r בווקטור A^TAc. קיבלנו אם כן: $A^TAc=A^Ty$
כלומר:

$$c=(A^TA)^{-1}A^Ty$$

למטריצה $(A^TA)^{-1}A^T$ קוראים לכן הפסאודו-הופכית של המטריצה A (שימו לב ש-A יכולה להיות מלבנית וממילא אין לה אז הפיך).

את החשבון הנ"ל ניתן לבצע גם ללא עיסוק מופרז באינדקסים, כדלקמן:

$$||Ac-y||_2^2=(Ac-y)^T(Ac-y)$$

(מכפלה פנימית של הווקטור $Ac-y$ בעצמו.)
ולכן

$$\frac{\partial}{\partial c}(||Ac-y||_2^2)=2(Ac-y)^TA=0$$

אם A מטריצה ריבועית, הערכים העצמיים והווקטורים העצמיים שלה מספקים אינ-פורמציה מועילה רבה על המטריצה ותיאורה כטרנספורמציה ליניארית. אם A מלבנית עלינו לחפש תחליף. ואכן, יש מושג דומה המהווה תחליף לניתוח ספקטרלי-(ניתוח של בעיות באמצעות ערכים עצמיים ווקטורים עצמיים). מדובר בערכים סינגולריים. המטר-יצה AA^T היא סימטרית ומוגדרת חיובית ולכן יש לה ערכים עצמיים $\sigma_1\geq\sigma_2\geq\ldots\geq 0$ המספרים $\sqrt{\sigma_i}$ נקראים הערכים הסינגולריים של A. זה יהיה מרכיב עיקרי בדיון שנערוך בשאלה הבאה: בהינתן המטריצה A ומספר טבעי t, מהו הקירוב הטוב ביותר של A ע"י מטריצה מדרגה $\leq t$? שוב עלינו להבהיר תחילה איך נמדוד קירבה בין מטריצות. כמקודם, התשובה תיגזר ממושג של נורמה מטריציאלית. כאן הדרישות הן כדלקמן:

1. $|||A|||\geq 0$ ושוויון אם"ם $A=0$
2. $|||cA|||=|c|\cdot||A||$ לכל מטריצה A ולכל c ממשי
3. $|||A+B|||\leq|||A|||+|||B|||$ לכל שתי מטריצות
4. $|||AB|||\leq|||A|||\cdot|||B|||$ (זה תנאי שכמובן ייחודי לנורמות של מטריצות ואין לו אנלוג בהקשר של נורמות ווקטוריות).

להלן מספר דוגמאות לנורמות מטריציאליות

א. נורמת l_1: $|||A|||_1 = \sum |a_{ij}|$

ב. נורמת l_2 (הקרויה גם נורמת פרוביניוס, נורמת הילברט-שמידט או נורמת שור) $|||A|||_2 = \sqrt{\sum (a_{ij})^2}$.

ג. נורמת l_∞: $|||A|||_\infty = \max |a_{ij}|$

ד. בניגוד לנורמות הקודמות הדומות לנורמות הווקטוריות, כאן נדבר על הנורמה האופרטורית שעצם הגדרתה מחייבת לחשוב על A כטרנספורמציה ליניארית:

$$|||A||| = \max_{\|x\|=1} \|Ax\|$$

נדון אם כך בשאלה אחת לדוגמה:

בהינתן מטריצה A, לאו דווקא ריבועית. מחפשים B בדרגה 1 כך ש: $\min |||A-B|||_2$

פירוש הדבר הוא שאנו מחפשים שני ווקטורים x, y כך ש: $\min \sum (a_{ij} - x_i y_j)^2$

(אנו חושבים על x כווקטור שורה ועל y כווקטור עמודה). גוזרים לפי x_i ומקבלים:

$$\sum_j (a_{ij} - x_i y_j) y_j = 0$$

כלומר, הקואורדינטה ה-i של Ay שווה ל-$x_i \|y\|_2^2$. ז"א: $Ay = x^T \|y\|_2^2$

וכיו"ב $xA = y^T \|x\|_2^2$

נובע כי $xAA^T = \|x\|^2 y^T A^T = x \cdot \|x\|^2 \|y\|^2$ וכיו"ב ל-y

כלומר x הוא ו"ע שמאלי של AA^T ו-y ו"ע ימני של A^TA, שניהם עם אותו ערך עצמי, σ.

כלומר, $xAA^T = \sigma x; A^T A y = \sigma y$.

ואז

$$\sum (a_{ij} - x_i y_j)^2 = \sum a_{ij}^2 - xAy + \|x\|^2 \|y\|^2$$

אבל $xAy = \|x\|^2 \|y\|^2 = \sigma$ לכן $\sum (a_{ij} - x_i y_j)^2 = \sum a_{ij}^2 - \sigma$ המינימום מושג כאשר $\sigma = \sigma_1$ הערך הסינגולרי העליון של A.

באופן כללי מוכיחים את המשפט הבא: תהיה $A_{m\times n}$ מטריצה ממשית מדרגה k. אז ניתן לרשום $A = V \Sigma W^T$ באשר $V_{m\times m}$ ו-$W_{n\times n}$ מטריצות אורתוגונליות. האיברים היחידים ב-Σ שאינם 0 הם $(\Sigma)_{ii} = \sigma_i$ הערכים הסינגולריים של A. העמודות של V הם ו"ע של AA^T העמודות של W הן ו"ע של A^TA.

הקירוב הטוב ביותר של A במובן l_2 ע"י מטריצה מדרגה $\leq r$ היא $V\Sigma^{(r)} W$ כש-$\Sigma^{(r)}$ מתקבלת מ-Σ ע"י החלפת האיברים σ_{ii} ב-0 לכל $i > r$

הערה: זהו גם הקירוב הטוב ביותר בנורמה האופרטורית.

כדאי להדגיש כאן עוד עיקרון חשוב המופיע במגוון רחב של יישומים: נרשום את ההצגה $A = V\Sigma W^T$.

כך:

$$a_{ij} = \sum_t v_{it} \sigma_t w_{tj}$$

ולכן:

$$A = \sum \sigma_t V^{(t)} \cdot W_{(t)}$$

כש-$V^{(t)}$היא העמודה ה-t ב-V ו-$W_{(t)}$ השורה ה-t ב-W. שימו לב שכ"א מהמטריצות האלה מדרגה 1.

זוהי הצגה קנונית של A כצירוף של "מרכיבים אלמנטריים" - מטריצות מדרגה 1. פעמים רבות (ונראה עוד כאלה בהמשך) מעניין לראות כיצד נבנה הקירוב שבו מחברים רק חלק מן ה"מרכיבים האלמטריים" - אלה שמתאימים לערך סינגולרי גדול. שכיח מאוד הוא המצב שקירובים כאלה מועילים בפועל יותר מהמטריצה עצמה. כך למשל, קירוב מסדר נמוך יותר עשוי להיות האות (סיגנל) המתואר במטריצה כשהוא מנוקה מרעשי מדידה- נתון המעניין אותנו הרבה יותר מאשר האות המדוד כשלעצמו. יש לציין גם שאם מספר המחוברים קטן, מושגת כאן גם דחיסה ניכרת של מידע.