אופטימיזציה קמורה ושימושים (67506) | פרויקט

אלי קורן (תואר מוסמך), אהוד דהן (תואר מוסמך), נריה כהן (תואר ראשון).

5 יוני, 2022

הגדרת הבעיה

נתונים $x \in \mathbb{R}^{n\times1}$ $y \in \mathbb{R}^{m\times1}$ $H \in \mathbb{R}^{m\times n}$

$$x^* = \underset{x\in\mathbb{R}^n}{\arg\min}\ \|y - Hx\|_2$$
$$s.t\ \vec{1}^T x = 1$$
$$x \geq 0$$

נשים לב שאפשר לסדר מחדש על ידי העלאה בריבוע, חילוץ איברים והשמטת מי שלא משפיע על המינימום כך שהבעיה שקולה ל:

$$\Leftrightarrow \quad \begin{aligned} x^* = \underset{x\in\mathbb{R}^n}{\arg\min}\ 2y^T Hx + x^T H^T Hx \\ s.t\ \vec{1}^T x - 1 = 0 \\ -Ix \leq \vec{0} \end{aligned}$$

הבעיה הדואלית

החלטנו לתקוף את הבעיה על ידי מציאת פתרון לביעיה הדואלית. מכיוון שזו בעיה קמורה והאילוצים אפיניים ידוע לנו שהחסם יהיה הדוק. כעת פיתחנו את נוסחת הלגרנגיאן והגדרנו את הפונקציה $g(\lambda, \mu)$. כפי שלמדנו בכיתה, פתחנו באופן אנליטי את הx האופטימילי למינימיזציה של נוסחת הלגרנגיאן וכך הגדרנו את g. כעת נרצה למצוא את הλ, μ הממקסמות את g תוך שאנו מוודאים ש$\lambda \geq 0$. לבעית האופטימיזציה הזאת ניגשנו בעזרת ניוטון מלווה בגרדיאנט דסנט כאשר באילוץ $\lambda \geq 0$ טיפלנו בעזרת שיטת Log Barrier.

הנוסחאות והפיתוחים מצורפים להלן.

כיוון ששיטה זו דורשת חישוב של המטריצה H^{-1} וכאשר המימדים שלה גדולים מאוד היא מאוד קשה מבחינה נומרית החלטנו לטפל במצבים האלו בעזרת גישה לבעיה הפרימאלית (עם שימוש בLog Barrier על האילוצים המקוריים והרצה של GD בלבד).

כעת נסביר את הבעיה הפרימאלית ולאחר מכן את הדואלית.

Log Barrier method

על מנת לטפל באילוצים החלטנו להשתמש במתודת Log Barrier. על מנת לטפל באי השיוויון $-Ix \leq \vec{0}$ נגדיר

$$\Phi\left(x\right)=-\sum_{i=1}^{n}\log\left(-f_{i}\left(x\right)\right)=-\log\left(\prod_{i=1}^{n}-x_{i}\right)$$

ועכשיו נוכל להעביר את המגבלות לתוך פונקציית המטרה:

$$\begin{aligned}&x^{*}=\underset{x\in\mathbb{R}^{n}}{\arg\min}\,x^{T}H^{T}Hx-2y^{T}Hx-\log\left(\prod_{i=1}^{n}-x_{i}\right)\cdot\frac{1}{t}\\&s.t\ \vec{1}^{T}x-1=0\end{aligned}$$

כאשר השמטנו את $y^{T}y$ מהביטוי כיוון שהוא קבוע ולא משפיע על המינימום. עם אילוץ השיוויון החלטנו להתמודד בשלב הבא.

GD+Newton

נשתמש עכשיו בניוטון בשביל להתחיל להתקדם בכיוון הנכון אך כיוון שידוע שניוטון זורק את החיפוש באופן אגרסיבי לאחר מספר איטרציות נעבור לגרידאנט דסנט בשביל לעדכן את וקטור החיפוש בצורה יותר ממוקדת. לטובת השימוש בהם נחשב את הנגזרות וכדי להימנע מחלוקה באפס כאשר יש פעולת חילוק הוספנו אפסילון:

$$\begin{aligned}g\left(x\right)&=\frac{df}{dx}=2H^{T}Hx-2y^{T}H-\frac{1}{tx+\epsilon}\\h\left(x\right)&=\frac{d^{2}f}{dx}=2H^{T}H+diag\left(\frac{1}{tx^{2}+\epsilon}\right)\end{aligned}$$

ולכן בכל עדכון נשתמש במשוואות הבאות:

$$\begin{aligned}Newton:&\qquad x_{n}=x_{n-1}-h\left(x\right)^{-1}g\left(x\right)\\GD:&\qquad x_{n}=x_{n-1}-\eta g\left(x\right)\end{aligned}$$

הערות: מצאנו ש$\eta=1e-8$ עובד לנו בצורה הטובה ביותר. בשביל להתמודד עם אילוץ השיוויון עדכנו בכל איטרציה רק נקודות x_{i}שהנגזרת בנקודה שלהם השאירה אותם חיוביים כלומר

$$x_{n}^{i}=\begin{cases}x_{n-1}^{i}-\eta g\left(x\right) & g\left(x_{i}\right)>0\\x_{n-1}^{i} & else\end{cases}$$

מימוש

למדנו היטב על בשרנו כי עד כמה שהתיאוריה עבדה יפה מאוד, בפועל יש כל מיני בעיות שנובעות מהמעבר למחשב. להלן נפרט את הדרכים בהם התמודדנו עם המעבר. פירוט הדרכים הוא לפי סדר הופעתם בקוד.

(1) לוג על ערכים נמוכים - בהגדרת פונקציית המטרה $f\left(x\right)$ מופיע הביטוי לוג שמגיע ממתודת הLog Barrier . כאשר ערכי האיקס נהיים קטנים מאוד זה מוביל לזריקת שגיאה. לכן בחרנו לבדוק האם $\sum_{i=1}^{n}x_i>0$ ורק אם כן לחשב את לוג הביטוי כפי שמופיע בנוסחה.

(2) חלוקה באפס - בהגדרת הנגזרת $g\left(x\right)$ וכן בהגדרת ההסיאן $h\left(x\right)$ מופיעה חלוקה בx ובx^2 בהתאמה. גם שם הוספנו אפסילון על מנת להימנע מחלוקה באפס או במספרים נמוכים מאוד.

(3) הגדרנו ביטויים קבועים בתחילת הסקופ וכך נמנענו מלחשב אותם פעמיים.

(4) חישבנו את $h\left(x\right)^{-1}=h\left(x\right)\cdot I$ כי להפוך מטריצה זה כבד.

תנאי עצירה

במימוש Log barrier יצרנו שתי לולאות מקוננות. הלולאה החיצונית מעדכנת את t והלולאה הפנימית אחראית להתקדמות של GD או של Newton.

(1) העצירה של האיטרציות על t. התחלנו עם t קטן על מנת לאפשר בהתחלה צעד יחסית גדול. וככל שהתקדמו האיטרציות הגדלנו את t על מנת להעניש יותר התרחקות. ההגדלה של

$$t_n=1+\frac{1}{13\cdot\sqrt{0.01}\cdot t_{n-1}}$$

תנאי העצירה היה כאשר $\left\|h\left(x\right)^{-1}g\left(x\right)\right\|<1e-6$ מתוך הנחה שאם זה המצב אז העדכון שנעשה כבר קטן כל כך שלא ישנה הרבה ואפשר לעצור.

(2) בלולאה הפנימית הגדרנו את הריצה להתרחש כל עוד מספר האיטרציות קטן מ10.

(3) החלטנו להשתמש ב5 צעדים באלגוריתם ניוטון לפני שנבצע גרדיאנט דיסנט. ראינו כי לאחר 5 צעדים הדיוק המתקבל כבר גדול וצעד נוסף כבר לא יקדם אותנו הרבה. ולהיפך - אנו אפילו עלולים להתקבע במערבולת ללא סוף.

קבועים

השתמשנו בקבועים הבאים:

(1) ϵ : מספר קטן שהוספנו לכל מכנה.

(2) $\eta=1e-08$: גודל הצעד של גרידאנט דסנט.

דואליות חזקה

שמנו שהבעיה הפרימלאית קמורה ולכן נוכל לקבל דואליות חזקה כך שנוכל לקבל ממש את האיקס האופטימלי ולא רק חסם תחתון שלו כפי שהוכחנו בכיתה.

על מנת להתמודד עם האילוצים הגדרנו את הלגרנגיאן באופן הבא:

$$L\left(x,\lambda,\mu\right)=2y^THx+x^TH^THx-\lambda_ix_i+\mu\vec{1}^Tx-\mu$$

נעיר שהשתמשנו בפיתוח של פונקציית המטרה כפי שעשינו בחלק הראשון של העבודה.

נוכל אם כן למצוא את x^* באמצעות השוואה ל-0:

$$\frac{dL}{dx}x^* = 0 \quad \Leftrightarrow \quad x^* = \left(H^T H\right)^{-1}\left(y^T H + \frac{\lambda_i}{2} - \frac{\mu}{2}\vec{1}^T\right)$$

נסמן $H^T H = A$ ונציב את x^* בפונקציית המטרה על מנת לקבל את הבעיה הדואלית

$$
\begin{aligned}
g\left(\lambda,\mu\right) &= \max_{\lambda,\mu}\min_{x} x^{*T}Ax^{*} - 2y^{T}Hx^{*} - \vec{\lambda}^{T}x^{*} + \mu_1\vec{1}x^{*} - \mu_1 \\
&= \max_{\lambda,\mu}\min_{x} \left(H^{T}y^{T}\left(A\right)^{-1T} + \frac{1}{2}\lambda^{T}A^{-1T} - \frac{1}{2}\vec{1}\mu_1\left(A\right)^{-1T}\right)\left(A\right)\left(\left(A\right)^{-1T}y^{T}H + \frac{1}{2}\lambda - \frac{1}{2}\mu_1\left(A\right)^{-1T}\vec{1}\right) \\
&\qquad - 2y^{T}H\left(A\right)^{-1T}\left(y^{T}H + \frac{1}{2}\lambda - \frac{1}{2}\mu_1\vec{1}\right) + \\
&\qquad - \lambda^{T}\left(A\right)^{-1T}\left(y^{T}H + \frac{1}{2}\lambda - \frac{1}{2}\mu_1\vec{1}\right) + \\
&\qquad + \mu_1\vec{1}^{T}\left(A\right)^{-1T}\left(y^{T}H + \frac{1}{2}\lambda - \frac{1}{2}\mu_1\vec{1}\right) \\
&\qquad - \mu_1 \\
&= \max_{\lambda,\mu}\min_{x} y^{T}HA^{-T}H^{T}y \\
&\qquad + \frac{1}{2}y^{T}HA^{-T}\lambda \\
&\qquad - \frac{1}{2}\mu_1 y^{T}HA^{-T}A^{-1}\vec{1} \\
&\qquad + \frac{1}{2}\lambda^{T}A^{-T}H^{T}y \\
&\qquad + \frac{1}{4}\lambda^{T}A^{-T}\lambda \\
&\qquad - \frac{1}{4}\mu_1\lambda^{T}A^{-T}\vec{1} \\
&\qquad - \frac{1}{2}\mu_1\vec{1}^{T}A^{-T}H^{T}y \\
&\qquad - \frac{1}{4}\mu_1\vec{1}^{T}A^{-T}\lambda \\
&\qquad + \frac{1}{4}\mu_1^{2}\vec{1}^{T}A^{-T}\vec{1} \\
&\qquad - 2y^{T}HA^{-1}H^{T}y \\
&\qquad - y^{T}HA^{-1}\lambda \\
&\qquad + \mu_1 y^{T}HA^{-1}\vec{1} \\
&\qquad - \lambda^{T}A^{-1}H^{T}y \\
&\qquad - \frac{1}{2}\lambda^{T}A^{-1}\lambda \\
&\qquad + \frac{1}{2}\mu_1\lambda^{T}A^{-1}\vec{1} \\
&\qquad + \mu_1\vec{1}^{T}A^{-1}H^{T}y \\
&\qquad + \frac{1}{2}\mu_1\vec{1}^{T}A^{-1}\lambda \\
&\qquad - \frac{1}{2}\mu_1\vec{1}^{T}A^{-1}\vec{1} \\
&\qquad - \mu_1
\end{aligned}
$$

עתה אין צורך למנם את x ונותר רק למקסם את λ,μ וכך למצוא חסם תחתון לx האופטימלי.

GD+Newton

נחשב את הנגזרות:

$$\begin{aligned}\frac{d\left(g\left(\lambda,\mu\right)\right)}{d\lambda} =& \frac{1}{2}A^{-T}\lambda\\ &-\frac{1}{2}A^{-1}\lambda\\ &+\frac{1}{2}A^{-1}H^{T}y\\ &+\frac{1}{2}A^{-T}-A^{-1}\\ &+\frac{1}{4}\mu_1A^{-1}\vec{1}\\ &+\frac{1}{2}\mu_1A^{-1T}\vec{1}\end{aligned}$$

$$\begin{aligned}\frac{d\left(g\left(\lambda,\mu\right)\right)}{d\mu_1} &=\\ (non-const) &-\frac{1}{4}\lambda^{T}A^{-T}\vec{1}\\ &-\frac{1}{4}\vec{1}^{T}A^{-1}\lambda\\ &+\frac{1}{2}\mu_1\vec{1}^{T}A^{-T}\vec{1}\\ &+\frac{1}{2}\lambda^{T}A^{-1}\vec{1}\\ &+\frac{1}{2}\vec{1}^{T}A^{-1}\lambda\\ (const) &+y^{T}HA^{-1}\vec{1}\\ &-\frac{1}{2}\vec{1}^{T}A^{-T}H^{T}y\\ &+y^{T}HA^{-1}\vec{1}\\ &+\vec{1}^{T}A^{-1}H^{T}y\\ &-\frac{1}{2}\vec{1}^{T}A^{-1}\vec{1}\\ &-1\end{aligned}$$

נגזרות שניות:

$$\frac{d^2g}{d\lambda} = \frac{1}{2}A^{-1}H^Ty$$
$$\frac{d^2g}{d\mu} = \frac{1}{2}\vec{1}^TA^{-1}\vec{1}$$

כאשר מהלך האלגוריתם זהה בדיוק למה שעשינו למעלה.