מודלים לא ליניאריים בחקר ביצועים - תרגיל בית 5

26 באפריל 2021

הנחיות להגשה:

- יש להגיש PDF עם שם קובץ HW##_ID1_ID2.
- יש לצרף ל־PDF את כל הקודים יחד עם הפלטים. בנוסף, יש להגיש קובץ ZIP המכיל את כל הקודים.

שאלה 1

פונקציית מטרה שימושית בלמידת מכונה מונחית היא $l\colon \mathbb{R}^{n+1} \to \mathbb{R}$ המוגדרת על־ידי

$$l(\mathbf{w}, v) = \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + e^{b_i\left(\mathbf{w}^T\mathbf{a}_i + v\right)}\right),$$

כאשר $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N \in \mathbb{R}^n$ עבור $N \in \mathbb{N}$ הם וקטורים כלשהם, $b_1, b_2, \ldots, b_N \in \{-1, 1\}$ קבועים ו־ $\mathbf{w} \in \mathbb{R}^n, v \in \mathbb{R}$ משתנים. הסבירו במילים מדוע l גזירה ברציפות והוכיחו שקבוע ליפשיץ של הגרדיאנט של l הוא $\frac{1}{4N}\sum_{i=1}^{N}\left(\|\mathbf{a}_i\|^2 + 1\right)$.

הדרכה: אל תחשבו את הקבוע ישירות, אלא השתמשו בטענות ובדוגמה שראיתם בתרגול.

שאלה 2

תהי $\{\mathbf{x}_k\}_{k\geq 0}$ סדרה הנוצרת ע"י שיטת הגרדיאנט עם חיפוש קווי מדויק למזעור הפונקציה הריבועית $f(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{x}$ עבור מטריצה $\mathbf{Q} \in \mathbb{R}^{n\times n} \succ 0$. הוכיחו כי לכל $k \in \mathbb{N}$ מתקיים

$$f(\mathbf{x}_{k+1}) \leq \left(\frac{\lambda_{\max}(\mathbf{Q}) - \lambda_{\min}(\mathbf{Q})}{\lambda_{\max}(\mathbf{Q}) + \lambda_{\min}(\mathbf{Q})}\right)^2 f\left(\mathbf{x}^k\right).$$

הכוונה: בתשובתכם השתמשו באי־שוויון הבא (ללא הוכחה): אם $\mathbf{Q} \succ 0$ אז לכל $\mathbf{y} \in \mathbb{R}^n$ מתקיים

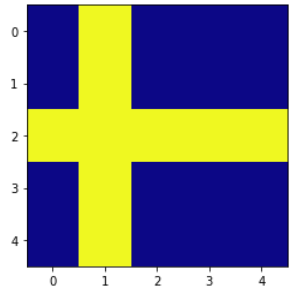
$$\frac{\left(\mathbf{y}^T\mathbf{y}\right)^2}{\left(\mathbf{y}^T\mathbf{A}\mathbf{y}\right)\left(\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}\right)} \geq \frac{4\lambda_{\max}(\mathbf{Q})\,\lambda_{\min}(\mathbf{Q})}{\left(\lambda_{\max}(\mathbf{Q}) + \lambda_{\min}(\mathbf{Q})\right)^2}.$$

כמו כן, ניתן להניח כי $\mathbf{x}_k^T\mathbf{Q}\mathbf{x}_k > 0$ לכל k.

שאלה 3

שאלה זו עוסקת בשחזור תמונה מטושטשת (Image Deblurring). תמונה היא מטריצה שבה כל רכיב מייצג פיקסל עם צבע לפי שיטת צבעים מסוימת (gray, plasma, etc.). למשל, הפקודות הבאות יוצרות את דגל שוודיה:

```
A=[[1,2,1,1,1],[1,2,1,1,1],[2,2,2,2,2],[1,2,1,1,1],[1,2,1,1,1]]
A = np.array(A)
plt.imshow(A, cmap='plasma')
plt.show()
```



ניתן לייצג טשטוש תמונה כהעתקה לינארית עליה (למשל, צילום תוך כדי נסיעה). לצורך כך נתייחס לתמונה בתצורתה הווקטורית, המתקבלת ע"י שרשור עמודותיה לווקטור עמודה אחד. מכאן, פעולת הטשטוש של תמונה מקורית $\mathbf{x}$ ניתנת להצגה כ־$\mathbf{b}=\mathbf{Ax}$. המטריצה $\mathbf{A}$ נקראת מטריצת הטשטוש והווקטור $\mathbf{b}$ הוא התמונה המטושטשת. אנו מעוניינים למצוא את התמונה המקורית $\mathbf{x}$ (ללא הטשטוש), ולכן נפתור את בעיית הריבועים הפחותים הבאה:

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left\{f\left(\mathbf{x}\right)=\left\|\mathbf{Ax}-\mathbf{b}\right\|^2\right\}.$$

הורידו את הקובץ blur.ipynb מאתר הקורס. אין צורך להבין כיצד פועלת פונקציה זו. הפקודות הבאות יוצרות תמונה ואת התמונה המטושטשת:

```
A,b,x = blur(256, 5, 1)
plt.figure(figsize=(6,6))
plt.imshow(x.reshape(256,256), cmap='gray')
plt.show()
plt.figure(figsize=(6,6))
plt.imshow(b.reshape(256,256), cmap='gray')
plt.show()
```

א. הפעילו את שיטת הגרדיאנט עם חיפוש קווי מדויק על־מנת לשחזר את התמונה הלא מטושטשת $\mathbf{x}$. התמונה ההתחלתית היא תמונת האפסים. צרפו את הפלט המתקבל לאחר 1, 10, 100 ו־1000 איטרציות. (ניתן להשתמש בפונקציות אותן כתבתן בתרגיל הבית הקודם - רק שימו לב לשנות את תנאי העצירה בהתאם). אין צורך לבדוק את נכונות הקלט. הניחו ש־$\mathbf{A}$ בעלת עמודות בת"ל.

ב. ראיתם בכיתה שקבוע ליפשיץ של הגרדיאנט של f הוא $2\lambda_{\max}\left(\mathbf{A}^T\mathbf{A}\right)$. הוכיחו שזהו קבוע ליפשיץ ההדוק (קטן) ביותר.

הכוונה: הראו שיש וקטורים מסוימים עבורם האי־שוויון שקיים בהגדרת קבוע ליפשיץ הוא שוויון.

חיזרו על הסעיף הקודם עם גודל צעד קבוע $\frac{1}{L}$, כאשר L הוא קבוע ליפשיץ ההדוק ביותר של הפונקציה ∇f. השתמשו בפונקציה eigs בשביל לחשב את L (בזמן ההרצה, חשבו את L והציבו את ערכו במשתנה, והימנעו מלחשב את L בכל איטרציה מחדש). איזו משתי הגישות מהירה יותר? איזו משתי הגישות נותנת תוצאה איכותית יותר?

ג. כאשר פונקציית המטרה קמורה ובעלת גרדיאנט ליפשיצי, ניתן להאיץ את שיטת הגרדיאנט. ידוע שפונקציית המטרה בבעיה זו היא קמורה (נראה בהמשך הקורס את המשמעות של כך). אחת הדרכים להאצה נקראת

FISTA, והיא פותחה על־ידי פרופ' אמיר בק. השיטה נראית כך:

set $\mathbf{y}^1 := \mathbf{x}^0$ and $t^1 = 1$

for $k = 1, 2, \ldots$ do:

$$\mathbf{x}^k = \mathbf{y}^k - \frac{1}{L}\nabla f\left(\mathbf{y}^k\right)$$

$$t^{k+1} = \frac{1+\sqrt{1+4\left(t^k\right)^2}}{2}$$

$$\mathbf{y}^{k+1} = \mathbf{x}^k + \frac{t^k - 1}{t^{k+1}}\left(\mathbf{x}^k - \mathbf{x}^{k-1}\right)$$

end for

קצב ההתכנסות התיאורטי של שיטה זו הוא $O\left(\frac{1}{k^2}\right)$, בעוד שזה של שיטת הגרדיאנט הוא $O\left(\frac{1}{k}\right)$.

כיתבו פונקציית python המופעלת באופן הבא:

```
x ,fs ,gs ,ts = fista(f, gf, L, x0, eps)
```

עבור f פונקציה בעלת גרדיאנט ליפשיצי עם קבוע L. תנאי העצירה הוא $\left\|\nabla f\left(\mathbf{x}^k\right)\right\| < \epsilon$ ומשמעות שאר המשתנים האחרים זהה למה שמוגדר בשאלה 1 בתרגיל הבית הקודם.

ד. חיזרו על סעיף א' עבור שיטת FISTA. איזו משלוש השיטות נותנת תוצאה איכותית יותר?

בנוסף, צרפו תרשים אחד עם ארבעה גרפים עבור ערכי פונקציית המטרה והזמן המצטבר ל־1000 איטרציות עבור שיטת צעד קבוע ועבור שיטת FISTA. השתמשו בסולם לוגריתמי בציר ה־y. איזו משתי השיטות סיימה 1000 איטרציות מהר יותר? איזו משתי השיטות זקוקה לפחות איטרציות בשביל להגיע לערך פונקציית מטרה נמוך יותר?

שאלה 4

בבעיית איכון מיקום של עצם מסוים (למשל, מקום נפילת רקטה על־סמך גורמי תצפית שונים) נתונים m מיקומים של חיישנים $\mathcal{A} \equiv \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\} \subset \mathbb{R}^n$ ומרחקים מוערכים $d_1, d_2, \ldots, d_m \geq 0$ בין החיישנים לבין העצם. כלומר, אם נסמן את מיקום העצם ב־$\mathbf{x} \in \mathbb{R}^n$, אז אנו יודעים כי $\|\mathbf{x} - \mathbf{a}_i\| \approx d_i$ לכל $1, 2, \ldots, m$. לכן, דרך להעריך את מיקום העצם $\mathbf{x} \in \mathbb{R}^n$ היא לפתור את בעיית האופטימיזציה הבאה:

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left\{f(\mathbf{x}) \equiv \sum_{i=1}^{m}\left(\|\mathbf{x} - \mathbf{a}_i\| - d_i\right)^2\right\}.$$

א. הסבירו (אין צורך בהוכחה מתמטית) מדוע פתרון בעיית אופטימיזציה זו אכן נותן הערכה עבור מיקומו האמיתי של העצם.

ב. הראו שתנאי האופטימליות $\nabla f(\mathbf{x}) = \mathbf{0}_n$ עבור $\mathbf{x} \notin \mathcal{A}$ שקול לתנאי

$$\mathbf{x} = \frac{1}{m}\left(\sum_{i=1}^{m}\mathbf{a}_i + \sum_{i=1}^{m}d_i\frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}\right).$$

ג. הראו ששיטת נקודה השבת הבאה

$$\mathbf{x}^{k+1} = \frac{1}{m}\left(\sum_{i=1}^{m}\mathbf{a}_i + \sum_{i=1}^{m}d_i\frac{\mathbf{x}^k - \mathbf{a}_i}{\|\mathbf{x}^k - \mathbf{a}_i\|}\right),$$

היא שיטת גרדיאנט, תחת ההנחה ש־$\mathbf{x}^k \notin \mathcal{A}$ לכל k. מהו גודל הצעד?

שאלה 5

א. תהי $\mathbf{A} \in \mathbb{R}^n$ הפיכה ויהי $\mathbf{b} \in \mathbb{R}^n$. יהי $\Delta \in \mathbb{R}^n$ כך ש־$\|\Delta\| \approx 0$. נסמן $\Delta\mathbf{b} \equiv \mathbf{b} + \Delta$. נסמן ב־$\mathbf{x}$ וב־$\Delta\mathbf{x}$ את הפתרונות של מערכות המשוואות $\mathbf{Ay} = \mathbf{b}$ ו־$\mathbf{Ay} = \Delta\mathbf{b}$ בהתאמה. הוכיחו כי

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \kappa\left(\mathbf{A}\right) \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}.$$

מהי המשמעות של אי־שוויון זה?

ב. נתונות הפונקציות $f\colon \mathbb{R}^n \to \mathbb{R}$ ו־$g\colon \mathbb{R}^n \to \mathbb{R}$ המוגדרות על־ידי $f\left(\mathbf{x}\right) = \frac{1}{2}\mathbf{x}^T\mathbf{F}\mathbf{x}$ ו־$g\left(\mathbf{x}\right) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x}$ עבור

$$\mathbf{F} = \begin{pmatrix} 2 & 2 & 3 \\ 2 & 5 & 5 \\ 3 & 5 & 7 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 10 & 8 & 9 \\ 8 & 13 & 11 \\ 9 & 11 & 15 \end{pmatrix}.$$

נניח והפעלנו את שיטת הגרדיאנט למציאת ממזער של f ושל g עם גודל צעד קבוע מיטבי. נניח שתנאי העצירה הוא $\|\nabla f\left(\mathbf{x}\right)\| \leq \varepsilon$. איזו משתי הבעיות צפויה להתכנס מהר יותר עבור נקודת ההתחלה $(1000, 1000, 1000)$? הסבירו את תשובתכם.