תרגול מספר 6 : אלגוריתם גרדיאנט ופרספטרון

תקציר התיאוריה

<u>אופטימיזציה ללא אילוצים</u>

נרצה למצוא את הערך x עבורו הפונקציה $f(x)$ מינימאלית. כלומר, יש למצוא x^* המקיים $f\left(x^*\right) \le f(x) \quad \forall x \in \Omega$.

נתונה פונקציית מחיר, $f\left(\boldsymbol{x}\right): \mathbb{R}^n \to \mathbb{R}$, התלויה בווקטור פרמטרים $\boldsymbol{x} = \left(x_1, x_2, \ldots, x_n\right)^T$.

נרצה למצוא $\boldsymbol{x}^*$ המביא למינימום את פונקציית המחיר, $f\left(\boldsymbol{x}^*\right) \le f\left(\boldsymbol{x}\right) \quad \forall \boldsymbol{x} \in X$, כאשר אין אילוצים על $\boldsymbol{x}$.

תנאים לאופטימליות

היעד : $\mathbf{w}^* = \arg\min_{w \in \mathbb{R}^n} f\left(\mathbf{w}\right)$.

בהנחה ש- f גזירה, תנאי הכרחי לנקודת מינימום מקומית הוא : $\nabla f(\mathbf{w}^*) = 0$

אם בנוסף f גזירה פעמיים, תנאי מספיק לנקודת מינימום מקומית מבודדת הוא :

$$\mathbf{H}(\mathbf{w}^*) > 0 \text{ and } \nabla f(\mathbf{w}^*) = 0$$

משמעות אי-השוויון היא שההסיאן הוא מטריצה חיובית מוגדרת (שכל הערכים העצמיים שלה חיוביים).

תזכורת: אופרטורים דיפרנציאליים וקטוריים (וקטור = וקטור עמודה)

וקטור הגרדיאנט :
$$\mathbf{g} \triangleq \nabla f\left(\mathbf{w}\right) \equiv \left[\frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2}, \ldots, \frac{\partial f}{\partial w_n}\right]^T$$

מטריצת ההסיאן (ריבועית סימטרית) :
$$\mathbf{H} \triangleq \nabla^2 f\left(\mathbf{w}\right) \equiv \begin{bmatrix} \frac{\partial^2 f}{\partial w_1^2} & \frac{\partial^2 f}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 f}{\partial w_1 \partial w_n} \\ \frac{\partial^2 f}{\partial w_2 \partial w_1} & \frac{\partial^2 f}{\partial w_2^2} & \cdots & \frac{\partial^2 f}{\partial w_2 \partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial w_n \partial w_1} & \frac{\partial^2 f}{\partial w_n \partial w_2} & \cdots & \frac{\partial^2 f}{\partial w_n^2} \end{bmatrix}$$

גרדיאנט של מכפלה:
$$\nabla\left(\mathbf{f}(\mathbf{w})^T \cdot \mathbf{g}(\mathbf{w})\right) = \left(\nabla \mathbf{f}^T\right)\mathbf{g} + \left(\nabla \mathbf{g}^T\right)\mathbf{f}$$

קירוב טיילור מסדר שני :
$$f(\mathbf{x}_0 + \Delta\mathbf{x}) \cong f(\mathbf{x}_0) + \mathbf{g}(\mathbf{x}_0)^T \Delta\mathbf{x} + \tfrac{1}{2}\Delta\mathbf{x}^T \mathbf{H}(\mathbf{x}_0)\Delta\mathbf{x}$$

אלגוריתם הגרדיאנט לאופטימיזציה נטולת אילוצים:

מטרה: במקרים רבים לא ניתן לחשב את $\mathbf{w}^*$ באופן ישיר ולכן משתמשים <u>בטכניקות פתרון איטרטיביות</u>. כלומר, מתחילים מניחוש $\mathbf{w}(0)=\mathbf{w}_0$ ומייצרים סדרה $\mathbf{w}(t+1)=\mathbf{w}(t)+\Delta\mathbf{w}(t)$, כך שיתקיים $f(\mathbf{w}(t+1))\le f(\mathbf{w}(t))$, "בתקווה" שהסדרה תתכנס למינימום.

קלט: רמת דיוק $thresh$ ונקודת מוצא $\mathbf{w}_0$.

שלבי האלגוריתם:

1. אתחל את נקודת המוצא של האלגוריתם $\mathbf{w}_0$.
2. עדכן את הפרמטרים בכיוון הנגדי לגרדיאנט בנקודה: $\mathbf{w}_{t+1}=\mathbf{w}_t+\Delta\mathbf{w}_t$

$$\Delta\mathbf{w}_t=-\eta\mathbf{g}(\mathbf{w}_t)$$

3. בדיקת התכנסות:
   - אם $|\mathbf{w}_{t+1}-\mathbf{w}_t|\le thresh$
     סיים אלג' ותן נקודה $\mathbf{w}_{t+1}$ כפתרון.
   - אחרת – חזור לשלב 2.

הסבר שלב העדכון:

אלגוריתם הגרדיאנט מתבסס על קירוב מסדר ראשון:

$$f(\mathbf{w}_{t+1})=f(\mathbf{w}_t+\Delta\mathbf{w}_t)\cong f(\mathbf{w}_t)+\mathbf{g}(\mathbf{w}_t)^T\Delta\mathbf{w}_t$$

עבור η חיובי וקטן:

$$f(\mathbf{w}_{t+1})-f(\mathbf{w}_t)\cong-\eta\|\mathbf{g}(\mathbf{w}_t)\|^2\le 0$$

האלגוריתם מתכנס לנקודה סטציונרית (קיצון/אוכף), בה מתקיים $\nabla f(\mathbf{w}^*)=0$.

עבור η חיובי קטן: הקירוב תקף ומובטח שהמחיר קטן. אבל - קצב הירידה קטן.

עבור η חיובי גדול: הקירוב אינו תקף. יתכנו תנודות ותיתכן התבדרות.

<u>בעיות</u>: התבדרות, התכנסות למינימום מקומי.

תרגיל 1 – אלגוריתם גרדיאנט

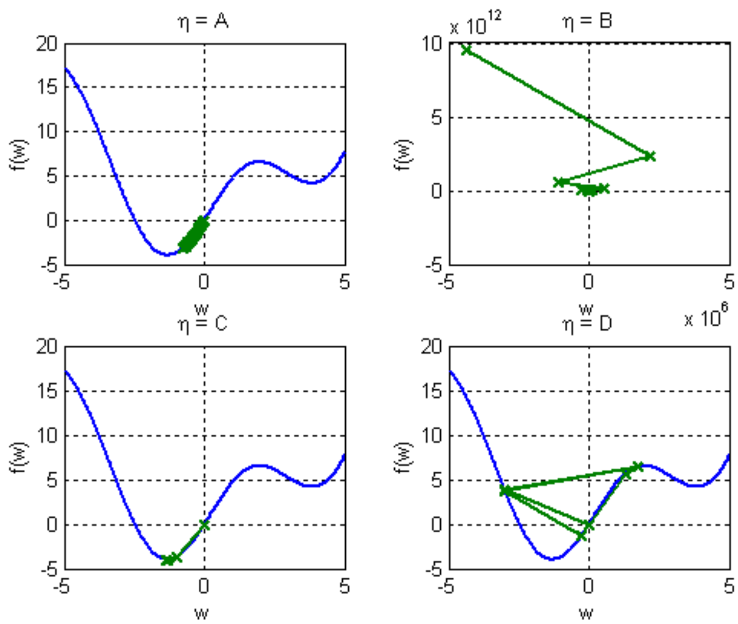
נתונה פונקציית המחיר: $f\left(w\right)=\frac{1}{2}w^2+5\sin\left(w\right)$

א. מהו תנאי הכרחי לנקודת מינימום?

ב. רשום את אלגוריתם הגרדיאנט לבעיה זו.

ג. הדגם שני צעדי לימוד, עבור הניחוש ההתחלתי הוא $w_0=0$, וצעד הלימוד $\eta=0.2$.

ד. הגרפים הבאים מציגים עשר איטרציות של אלגוריתם גרדיאנט עבור ארבעה ערכים שונים של צעד לימוד: $\eta\in\{0.01,0.2,0.6,3\}$. התאם בין צעד הלימוד לגרף.



תרגיל 2 – אופטימיזציה של גודל הצעד:

מצא קרוב לגודל הצעד $\eta(t)$ ה"אופטימלי" של אלגוריתם הגרדיאנט ע"י שימוש בקירוב טיילור מסדר שני של פונקצית המחיר.

תקציר התאוריה- פרספטרון בודד

פרספטרון הוא רכיב בעל d כניסות, x_i, ויציאה בודדת, y המחושבת באופן הבא:

$$y = \varphi\left(\sum_{i=1}^{d} w_i \cdot x_i + b\right) = \varphi\left(\mathbf{w}^T\mathbf{x} + b\right).$$

אופי פונקציית האקטיבציה, φ, קובע את סוג הפרספטרון.

נהוג לסמן: $v = \mathbf{w}^T\mathbf{x} + b$. כמו כן נהוג להתייחס לרכיב ההיסט (bias) ככניסה נוספת, $x_0 = 1$,

בעלת משקל $w_0 = b$: $\sum_{i=1}^{d} x_i \cdot w_i + b = \sum_{i=0}^{d} x_i \cdot w_i$

פונקציות אקטיבציה שכיחות הן:

פרספרטון ליניארי- $\varphi(v) = v$

פונקצית אקטיבציה לוגיסטית- $\varphi(v) = \dfrac{1}{1+\exp(-v)}$

פונקציית אקטיבציה טנגנס היפרבולי- $\varphi(v) = \tanh(v) = \dfrac{e^{2x}-1}{e^{2x}+1}$

שתי הפונקציות האחרונות הן פונקציה סיגמואידית (רוויה).

הערה: עבור פרספטרון ליניארי יש לדרוש $b = 0$, על מנת שהפלט יהיה ליניארי בקלט.

אלגוריתם לימוד מבוסס Gradient Descent:

אלגוריתם לימוד מעדכן את המשקלים (וההיסט), w_i על מנת למצוא מיפוי רצוי בין הכניסה ליציאה. המטרה מנוסחת לרוב ע"י מינימיזציה של פונקצית מחיר מתאימה.

אלגוריתם גרדיאנט מתאים לשמש כאלגוריתם לימוד של פרספטרון בעל **פונקצית אקטיבציה רציפה ומונוטונית**.

האלגוריתם ממזער את פונקציית המחיר: $f(\mathbf{w}) = \sum_{k=1}^{n} e^2_k = \sum_{k=1}^{n}\left(d_k - y_k\right)^2$

$d_k \in R$ הוא הפלט הרצוי ו- $y_k \in R$ הוא פלט הרשת עבור הקלט $\mathbf{x}_k \in \Re^d$.

המשקלים מקבלים ערך התחלתי אקראי.

כלל העדכון: $\mathbf{w}_{t+1} = \mathbf{w}_t + \Delta\mathbf{w}_t$, כאשר:

גרסת Batch: $\Delta\mathbf{w}_t = \eta\sum_{k=1}^{n}\left(d_k - y_k\right)\cdot\varphi'\left(v_k\right)\cdot\mathbf{x}_k$

עדכון סדרתי: $\Delta\mathbf{w}_t = \eta\left(d_t - y_t\right)\cdot\varphi'\left(v_t\right)\cdot\mathbf{x}_t$

פרספטרון דיסקרטי (ספרתי)

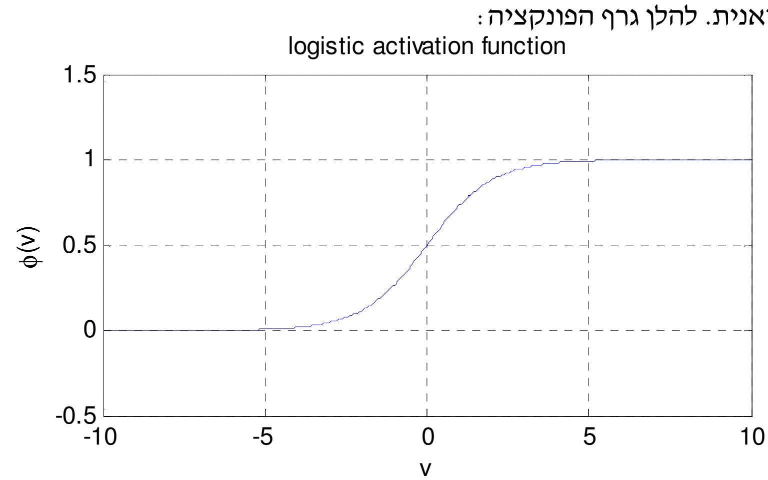
יש שני סוגים של פרספטרון דיסקרטי.

סוג הפרספטרון	בוליאני	סימן
ערכי קלט/פלט	$x_i \in R$ $y \in \{0,1\}$	$x_i \in R$ $y \in \{-1,0,1\}$
פונקצית אקטיבציה	$\varphi(v) = \mathrm{I}(v \geq 0) = \begin{cases} 1 & v \geq 0 \\ 0 & v < 0 \end{cases}$	$(v) = sign(v) = \begin{cases} 1 & v > 0 \\ 0 & v = 0 \\ -1 & v < 0 \end{cases}$

לא ניתן להשתמש באלגוריתם גרדיאנט ללימוד פרספטרון בינארי בגלל פונקצית האקטיבציה הלא רציפה.

תרגיל 3 – פרספטרון לא ליניארי

פונקצית אקטיבציה לוגיסטית ($\varphi(v) = \dfrac{1}{1+\exp(-v)}$) מהווה גרסה חלקה וגזירה של פונקצית אקטיבציה בוליאנית. להלן גרף הפונקציה:



א. רשום את נוסחת עדכון המשקלים (הסדרתי) עבור פרספטרון בעל פונקצית אקטיבציה זו.

ב. נתון פרספטרון בעל 2 כניסות, x_1, x_2 המקבלות ערכים מהקבוצה $x_i \in \{100, 200\}$. הפלט הרצוי מקבל את הערכים $d \in \{0,1\}$. השתמש באתחול משקלים $w_1 = w_2 = 1$ ומקדם לימוד $\eta = 0.1$.

1. מצא חסם לגודל גורם התיקון של המשקלים, Δw_i באיטרציה הראשונה של עדכון סדרתי.
2. מה תוכל להסיק מחסם זה אודות קצב הלימוד.
3. מהו הגורם לבעיה? הצע פתרון.