אופטימיזציה

תרגיל 1 – אלגוריתם גרדיאנט

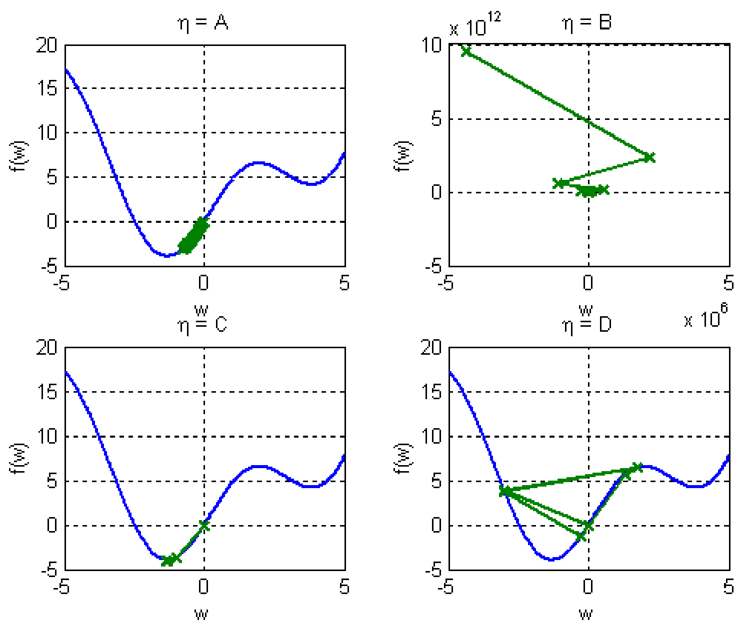
נתונה פונקצית המחיר: $f(w) = \frac{1}{2}w^2 + 5\sin(w)$

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.



פתרון:

א) תנאי הכרחי: $\frac{\partial f(w)}{\partial w} = 0$. מתקבל: $\frac{\partial f(w)}{\partial w} = w + 5\cos(w) = 0$

ב) אלגוריתם גראדיינט:

a. אתחול: w_0 כלשהוא.

b. עדכון: $w_{t+1} = w_t + \Delta w_t$, $\Delta w_t = -\eta(w_t + 5\cos(w_t))$,

$$w_{t+1} = w_t - \eta(w_t + 5\cos(w_t))$$

c. תנאי עצירה: $|\Delta w_t| < \varepsilon$

ג) שני צעדי לימוד: $w_0 = 0$, $\eta = 0.2$.

צעד ראשון:

$$\Delta w_0 = -\eta(w_0 + 5\cos(w_0)) = -0.2 \cdot 5 = -1$$

$$w_1 = w_0 + \Delta w_0 = 0 - 1 = -1$$

צעד שני:

$$\Delta w_1 = -\eta(w_1 + 5\cos(w_1)) = -0.2(-1 + 5\cos(-1)) = -0.2 \cdot 1.7015 = -0.3403$$

$$w_2 = w_1 + \Delta w_1 = -1 - 0.3403 = -1.3403$$

פתרון: $w^* = -1.30644$

ד) צעד לימוד קטן – התכנסות מובטחת אך איטית.
צעד לימוד גדול מידי – התבדרות
צעד לימוד גדול – תנודות.
שיוך הגרפים ל- η:

$$D \Leftrightarrow \eta = 0.6 \;\;, C \Leftrightarrow \eta = 0.2 \;\;, B \Leftrightarrow \eta = 3 \;\;, A \Leftrightarrow \eta = 0.01$$

תרגיל 2 – אופטימיזציה של גודל הצעד:

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

פתרון:

קרוב טיילור מסדר שני של $f(\mathbf{w})$ סביב $f(\mathbf{w}_t)$:

$$f(\mathbf{w}) \simeq f(\mathbf{w}_t) + \nabla f^T(\mathbf{w} - \mathbf{w}_t) + \frac{1}{2}(\mathbf{w} - \mathbf{w}_t)^T \mathbf{H}(\mathbf{w} - \mathbf{w}_t) \;,$$

כאשר: $\nabla f \triangleq \nabla f(\mathbf{w}_t)$ ו- $\mathbf{H}$ מטריצת ההסיאן של f.

$$f(\mathbf{w}_{t+1}) \simeq f(\mathbf{w}_t) + \nabla f^T(\mathbf{w}_{t+1} - \mathbf{w}_t) + \frac{1}{2}(\mathbf{w}_{t+1} - \mathbf{w}_t)^T \mathbf{H}(\mathbf{w}_{t+1} - \mathbf{w}_t)$$

נציב את $\mathbf{w}_{t+1} - \mathbf{w}_t = -\eta(t)\nabla f$ במשוואה:

$$f(\mathbf{w}_{t+1}) \simeq f(\mathbf{w}_t) + \nabla f^T(-\eta(t)\nabla f) + \frac{1}{2}(-\eta(t)\nabla f)^T \mathbf{H}(-\eta(t)\nabla f)$$

$$f(\mathbf{w}_{t+1}) \simeq f(\mathbf{w}_t) - \eta(t)\|\nabla f\|^2 + \eta^2(t)\frac{1}{2}\nabla f^T \mathbf{H} \nabla f$$

נגזור לפי $\eta(k)$ ונשווה ל- 0 כדי למזער את $f(\mathbf{w}_{t+1})$:

$$\frac{\partial f(\mathbf{w}_{t+1})}{\partial \eta(t)} = -\|\nabla f\|^2 + \eta(t)\nabla f^T \mathbf{H} \nabla f = 0 \qquad \Rightarrow \eta(t) = \frac{\|\nabla f\|^2}{\nabla f^T \mathbf{H} \nabla f}$$

פרספטרון בודד

תרגיל 3 – פרספטרון לא ליניארי

פונקצית אקטיבציה לוגיסטית מהווה גרסה חלקה וגזירה של פונקצית אקטיבציה בוליאנית. צייר את פונקצית האקטיבציה $\varphi(v)=\dfrac{1}{1+\exp(-v)}$.

רשום את נוסחת עדכון המשקלים עבור פרספטרון בעל פונקצית אקטיבציה זו.

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.
2. מה תוכל להסיק מחסם זה אודות קצב הלימוד.
3. מהו הגורם לבעיה. הצע פתרון.

פתרון:

א. כלל עדכון (עדכון סדרתי): $\Delta \mathbf{w}_t = -\eta\left(d_t-\varphi\left(v_t\right)\right)\varphi'\left(v_t\right)\mathbf{x}_t$, כאשר:

$$v_t=\sum_{i=1}^{d} w_{it}x_{it}+b_t=\mathbf{w}_t{}^T\mathbf{x}_t\ ,\ \varphi(v)=\frac{1}{1+\exp(-v)}\ ,\ \varphi'(v)=\frac{\exp(-v)}{\left(1+\exp(-v)\right)^2}$$

באופן סקלרי: $\Delta w_{it}=-\eta\left(d_t-\varphi\left(v_t\right)\right)\varphi'\left(v_t\right)x_{it}$, $\Delta b_t=-\eta\left(d_t-\varphi\left(v_t\right)\right)\varphi'\left(v_t\right)$

ב.

1) חסם עליון: $\left|\Delta w_t\right| \le \eta \max_{\mathbf{x}}\left\{\left(d_t-\varphi\left(v_t\right)\right)\right\}\max_{\mathbf{x}}\left\{\varphi'\left(v_t\right)\right\}\max_{\mathbf{x}}\left\{\mathbf{x}_t\right\}$ עבור $w_1=w_2=1$, $x_i\in\{100,200\}$

$$0\le\varphi\left(v_t\right)\le 1\ ,\ d_t\in\{0,1\}\ \Rightarrow \max\left\{\left(d_t-\varphi\left(v_t\right)\right)\right\}\le 1$$

$$\max_{\mathbf{x}}\left\{\varphi'\left(v_t\right)\right\}=\max_{\mathbf{x}}\left\{\frac{\exp\left(-\mathbf{w}^T\mathbf{x}\right)}{\left(1+\exp\left(-\mathbf{w}^T\mathbf{x}\right)\right)^2}\right\}=\frac{\exp\left(-(100+100)\right)}{\left(1+\exp\left(-(100+100)\right)\right)^2}$$
$$=\frac{\exp(-200)}{\left(1+\exp(-200)\right)^2}\simeq 1.38\cdot 10^{-87}$$

$$\left|\Delta w_t\right|\le 0.1\cdot 1\cdot 1.38\cdot 10^{-87}\begin{pmatrix}200\\200\end{pmatrix}\le\begin{pmatrix}10^{-86}\\10^{-86}\end{pmatrix}$$

2) קצב הלימוד - מתקבל צעד עדכון אפסי – מבחינה מעשית זה 0, ולכן לא יהיה עדכון.

3) הגורם לבעיה הוא ש- v נמצא בתחום הרוויה של $\varphi\left(v_t\right)$, והנגזרת שם אפסית. כדי להתגבר על הבעיה יש לנרמל את הקלט או את המשקולות. דוגמא לנרמול: $\tilde{x}_i=\dfrac{x_i-150}{50}$. נירמול זה חשוב

בהקשר של תנאי ההתחלה – קיימים תנאי התחלה ש״תוקעים״ את התהליך. יש להקפיד על כך שכל הפרספטרונים יהיו בתחום הלינארי בתחילת התהליך.