

باسمه تعالی



پروژه‌ی درس بهینه‌سازی محدب I

## مقدمه‌ای بر بهینه‌سازی تصادفی

استاد درس

دکتر محمدحسین یاسائی میبدی

دانشکده‌ی مهندسی برق  
دانشگاه صنعتی شریف

خرداد ۱۴۰۲

آخرین مهلت تحویل:  
۹ تیر ۱۴۰۲

## فهرست مطالب

۲	۱ مقدمه
۳	۲ هادی اگر تویی که کسی گم نمی شود!
۴	۳ از گوشه‌ای برون آی ای کوکبِ هدایت!
۶	۴ مرا زین قید ممکن نیست جستن!
۸	۵ مرا امید وصال تو زنده می دارد!
۱۰	۶ نبود خیر در آن خانه که عصمت نبود!
۱۱	۷ نصیحتی کُنت بشنو و بهانه مگیر!

## ۱ مقدمه

در این پروژه می‌خواهیم مسائل بهینه‌سازی تصادفی را بررسی کنیم. در مسائل بهینه‌سازی تصادفی، بخشی از تابع هدف مسئله یک متغیر تصادفی است. در نتیجه تابع هدف را می‌توان به صورت  $F(\mathbf{x}, \mathbf{Z})$  نوشت که  $\mathbf{Z} \in \mathbb{R}^m$  برداری تصادفی است. فضای نمونه‌ی این بردار تصادفی را  $\mathcal{Z}$  و تابع چگالی/جرم احتمال آن را  $p_{\mathbf{Z}}(\mathbf{z})$  می‌نامیم. همچنین فرض می‌کنیم تابع  $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  نسبت به ورودی اول خود (یعنی  $\mathbf{x}$ ) محدب است. حال تعریف می‌کنیم:

$$f(\mathbf{x}) = \mathbb{E}_{\mathbf{Z}} [F(\mathbf{x}, \mathbf{Z})] = \int_{\mathbf{z} \in \mathcal{Z}} F(\mathbf{x}, \mathbf{z}) p_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z}. \quad (1)$$

هدف مسئله‌ی بهینه‌سازی محدب تصادفی حل مسئله‌ی زیر است:

$$\begin{aligned} &\text{Minimize} && f(\mathbf{x}) = \mathbb{E}_{\mathbf{Z}} [F(\mathbf{x}, \mathbf{Z})] \\ &\text{subject to:} && \mathbf{x} \in \mathcal{C}. \end{aligned} \quad (2)$$

که  $\mathcal{C}$  یک مجموعه‌ی محدب است.



### ۳ از گوشه‌ای برون آی ای کوکب هدایت!<sup>۲</sup>

حال یک قدم در راستای سخت کردن مسئله برمی داریم. در حالتی که تابع  $f$  در همه‌ی نقاط مشتق پذیر نباشد، باید از مفهوم زیرگرادیان استفاده کنیم. در درس با مفهوم زیرگرادیان آشنا شده ایم.

**تعریف ۱-۳.** تابع  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  را در نظر بگیرید، بردار  $\mathbf{v}$  را یک زیرگرادیان برای تابع  $f$  در نقطه‌ی  $\mathbf{x} \in \mathbb{R}^n$  می نامیم، اگر به ازای هر  $\mathbf{y} \in \mathbb{R}^n$  داشته باشیم:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{y} - \mathbf{x}, \mathbf{v} \rangle \quad (۳)$$

مجموعه‌ی همه‌ی زیرگرادیان‌های  $f$  در نقطه‌ی  $\mathbf{x}$  را با  $\partial f(\mathbf{x})$  نشان می دهیم.

تعمیم الگوریتم ۱ به حالتی که به جای گرادیان، به زیرگرادیان دسترسی داشته باشیم، سراسر است:

---

#### Algorithm 2 Sub-Gradient Descent

---

**parameters:** Set of Scalars  $\{\eta_t\}_{t=1}^T$  where for all  $t \in [T]$ ,  $\eta_t > 0$ , integer  $T > 0$ , Initial Point  $\mathbf{x}_0$

**Initialize**  $\mathbf{x}^{(1)} = \mathbf{x}_0$

**for**  $t = 1, 2, \dots, T$  **do**

    choose  $\mathbf{v}^{(t)}$  such that  $\mathbf{v}^{(t)} \in \partial f(\mathbf{x}^{(t)})$   
    update  $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta_t \mathbf{v}^{(t)}$

**end**

**return**  $\mathbf{x}^{(T+1)}$

---

پرسشی که وجود دارد، آنست که آیا همواره حرکت در خلاف جهت زیرگرادیان تابع را کاهش می دهد یا خیر؟

**پرسش تئوری ۳.** فرض کنید که  $\mathbf{0} \notin \partial f(\mathbf{x}^{(t)})$  و تعریف کنید  $\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \{f(\mathbf{x})\}$ . نشان دهید به ازای هر بردار  $\mathbf{v}^{(t)} \in \partial f(\mathbf{x}^{(t)})$  که در مورد آن داشته باشیم  $\mathbf{v}^{(t)} \in \partial f(\mathbf{x}^{(t)})$ ، می توان یک  $\eta_t > 0$  یافت که به ازای آن داشته باشیم:

$$\|\mathbf{x}^{(t+1)} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}^{(t)} - \eta_t \mathbf{v}^{(t)} - \mathbf{x}^*\|_2^2 < \|\mathbf{x}^{(t)} - \mathbf{x}^*\|_2^2.$$


---

حالا آماده هستیم که همگرایی الگوریتم ۲ را بررسی کنیم. ابتدا دو فرض به مسئله اضافه می کنیم:

**فرض ۱-۳.** مقدار بهینه‌ی تابع  $f$  در بی نهایت رخ نمی دهد.  
 به تعبیر دیگر اگر تعریف کنیم:  $\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \{f(\mathbf{x})\}$ ، خواهیم داشت:  $\|\mathbf{x}^*\|_2 \leq B$  که  $B < +\infty$ .

**فرض ۲-۳.** تابع  $f$ ، یک تابع  $\rho$ -لیپشیتز است. یعنی به ازای هر  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ :

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq \rho \|\mathbf{x} - \mathbf{y}\|_2.$$


---

**پرسش تئوری ۴.** برای هر تابع  $\rho$ -لیپشیتز مانند  $f$  نشان دهید که درباره‌ی هر بردار زیرگرادیان مانند  $\mathbf{v}$  داریم:

$$\|\mathbf{v}\|_2 \leq \rho.$$

**پرسش تئوری ۵.** تعریف می کنیم:  $\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \{f(\mathbf{x})\}$ ، با توجه به الگوریتم ۲ نشان دهید که به ازای هر  $\eta_t$  داریم:

$$\frac{1}{2} \|\mathbf{x}^{(t+1)} - \mathbf{x}^*\|_2^2 \leq \frac{1}{2} \|\mathbf{x}^{(t)} - \mathbf{x}^*\|_2^2 - \eta_t (f(\mathbf{x}^{(t)}) - f(\mathbf{x}^*)) + \frac{\eta_t^2}{2} \|\mathbf{v}^{(t)}\|_2^2.$$

---

<sup>۲</sup> در این شب سیاهم گم گشت راه مقصود/ از گوشه‌ای برون آی ای کوکب هدایت  
 از هر طرف که رفتم جز وحشتم بَیْفَزود/ زِ نهار از این بیابان وین راه بی نهایت [حافظ]

پرسش تئوری ۰۶. فرض کنید  $\{\eta_t\}_{t=1}^T$  هر دنباله‌ی دلخواهی از اعداد نامنفی باشد. همچنین فرض کنید که مفروضات ۱-۳ و ۲-۳ برقرار باشند. نشان دهید اگر در الگوریتم ۲ قرار دهیم:  $\mathbf{x}_0 = \mathbf{0}$ ، آن‌گاه:

$$\sum_{t=1}^T \eta_t (f(\mathbf{x}^{(t)}) - f(\mathbf{x}^*)) \leq \frac{1}{2} B^2 + \frac{1}{2} \rho^2 \sum_{t=1}^T \eta_t^2.$$

پرسش تئوری ۰۷. تعریف می‌کنیم:

$$\bar{\mathbf{x}}_T = \frac{\sum_{t=1}^T \eta_t \mathbf{x}^{(t)}}{\sum_{t=1}^T \eta_t}.$$

نشان دهید:

$$f(\bar{\mathbf{x}}_T) - f(\mathbf{x}^*) \leq \frac{B^2 + \rho^2 \sum_{t=1}^T \eta_t^2}{2 \sum_{t=1}^T \eta_t}.$$

پرسش تئوری ۰۸. قرار دهید  $\eta_t = \eta$ . کرانی که در پرسش تئوری ۰۷ اثبات کردید را بازنویسی کنید و مقدار بهینه‌ی  $\eta$  را برای به دست آوردن بهترین کران محاسبه کنید. در نهایت بهترین کران را گزارش کنید. پاسخ نهایی شما باید برحسب  $B, \rho, T$  باشد.

پرسش تئوری ۰۹. به طور شهودی توجیه کنید که چرا کران نهایی برحسب  $f(\bar{\mathbf{x}}_T)$  به دست آمد، نه برحسب  $f(\mathbf{x}^{(T)})$  یا  $f(\mathbf{x}^{(T+1)})$ .

پرسش تئوری ۰۱۰. تعریف می‌کنیم:  $t_T^* = \arg \min_{t \in [T]} f(\mathbf{x}^{(t)})$ . نشان دهید:

$$f(\mathbf{x}^{(t_T^*)}) - f(\mathbf{x}^*) \leq \frac{B^2 + \rho^2 \sum_{t=1}^T \eta_t^2}{2 \sum_{t=1}^T \eta_t}.$$

پرسش شبیه‌سازی ۰۱. قرار دهید  $m = 100, n = 10$ . بردارهای تصادفی  $\{\mathbf{Z}_i\}_{i=1}^m$  را به صورت مستقل و هم‌توزیع از توزیع  $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  بسازید. همچنین متغیرهای تصادفی  $\{B_i\}_{i=1}^m$  را به صورت مستقل و هم‌توزیع از توزیع  $\mathcal{N}(0, 1)$  بسازید. این متغیرها در طول شبیه‌سازی تغییر نمی‌کنند. تابع هدف زیر را تشکیل دهید:

$$f(\mathbf{x}) = \frac{1}{m} \sum_{i=1}^m |\mathbf{Z}_i^\top \mathbf{x} - B_i|.$$

زیرگرایان این تابع را به صورت تئوری محاسبه کنید. سپس الگوریتم ۲ را با  $\mathbf{x}_0 = \mathbf{0}, T = 4000$  و  $\eta_t = \eta$  اجرا کنید و کمیت  $f(\mathbf{x}^{(t)}) - f(\mathbf{x}^*)$  را در طول اجرای الگوریتم محاسبه و نمودار آن را رسم کنید.

پرسش شبیه‌سازی ۰۲. به ازای  $\{\mathbf{Z}_i\}_{i=1}^m, \{B_i\}_{i=1}^m$  ای که در پرسش شبیه‌سازی ۰۱ ساختید و پارامترهای داده‌شده در آن پرسش، الگوریتم ۲ را اجرا کنید و کمیت  $f(\mathbf{x}^{(t^*)}) - f(\mathbf{x}^*)$  را در طول اجرای الگوریتم محاسبه و نمودار آن را رسم کنید.

## ۴. چرا زین قید ممکن نیست جستن!

حال فرض کنید بخواهیم یک مسئله‌ی بهینه‌سازی محدب و دارای قید مانند (۲) را با کمک الگوریتم ۲ حل کنیم. موقتاً فرض می‌کنیم تابع  $f$  را در اختیار داریم و خاصیت تصادفی مسئله را فراموش می‌کنیم. وقتی می‌خواهیم یک مسئله‌ی بهینه‌سازی مقید را حل کنیم، با پیمودن گام موجود در الگوریتم ۲، یعنی  $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta_t \mathbf{v}^{(t)}$  ممکن است از مجموعه‌ی  $\mathcal{C}$  خارج شویم! در نتیجه بعد از پیمودن هر گام، باید نقطه‌ی به دست آمده را روی مجموعه‌ی  $\mathcal{C}$  تصویر کنیم. تابع زیر را به این منظور در نظر می‌گیریم:

$$\Pi_{\mathcal{C}}(\mathbf{x}) = \arg \min_{\mathbf{y} \in \mathcal{C}} \{\|\mathbf{y} - \mathbf{x}\|_2\}. \quad (۴)$$

پرسش تئوری ۱۱. فرض کنید  $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}$  که  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ . فرم بسته‌ی تابع  $\Pi_{\mathcal{C}}(\mathbf{x})$  را به ازای هر  $\mathbf{x} \in \mathbb{R}^n$  بیابید.

پرسش تئوری ۱۲. فرض کنید  $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$  که  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ . فرم بسته‌ی تابع  $\Pi_{\mathcal{C}}(\mathbf{x})$  را به ازای هر  $\mathbf{x} \in \mathbb{R}^n$  بیابید.

پرسش تئوری ۱۳. فرض کنید  $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 \leq b\}$  که  $b \in \mathbb{R}^{\geq 0}$ . فرم بسته‌ی تابع  $\Pi_{\mathcal{C}}(\mathbf{x})$  را به ازای هر  $\mathbf{x} \in \mathbb{R}^n$  بیابید.

با این ویرایش، الگوریتم کاهش زیرگرادیان در حالت مقید به این صورت درمی‌آید:

---

### Algorithm 3 Constrained Sub-Gradient Descent

---

**parameters:** Set of Scalars  $\{\eta_t\}_{t=1}^T$  where for all  $t \in [T]$ ,  $\eta_t > 0$ , integer  $T > 0$ , Initial Point  $\mathbf{x}_0$

**Initialize**  $\mathbf{x}^{(1)} = \mathbf{x}_0$

**for**  $t = 1, 2, \dots, T$  **do**

    choose  $\mathbf{v}^{(t)}$  such that  $\mathbf{v}^{(t)} \in \partial f(\mathbf{x}^{(t)})$   
    update  $\mathbf{x}^{(t+1)} = \Pi_{\mathcal{C}}(\mathbf{x}^{(t)} - \eta_t \mathbf{v}^{(t)})$

**end**

**return**  $\bar{\mathbf{x}}_T = \frac{\sum_{t=1}^T \eta_t \mathbf{x}^{(t)}}{\sum_{t=1}^T \eta_t}$

---

پرسش تئوری ۱۴. نشان دهید رابطه‌ی به‌روز کردن الگوریتم ۳ یعنی  $\mathbf{x}^{(t+1)} = \Pi_{\mathcal{C}}(\mathbf{x}^{(t)} - \eta_t \mathbf{v}^{(t)})$  معادل با رابطه‌ی زیر است:

$$\mathbf{x}^{(t+1)} = \arg \min_{\mathbf{x} \in \mathcal{C}} \left\{ f(\mathbf{x}^{(t)}) + \langle \mathbf{v}^{(t)}, \mathbf{x} - \mathbf{x}^{(t)} \rangle + \frac{1}{2\eta_t} \|\mathbf{x} - \mathbf{x}^{(t)}\|_2^2 \right\}.$$

تعبیر شهودی این رابطه چیست؟

در حالت مقید نیاز به دو فرض جدید برای تحلیل تئوری الگوریتم ۳ داریم:

فرض ۴-۱. مقدار بهینه‌ی تابع  $f$  در بی‌نهایت رخ نمی‌دهد و مجموعه‌ی  $\mathcal{C} \subset \mathbb{R}^n$  یک مجموعه‌ی محدب و فشرده است. به تعبیر دیگر اگر تعریف کنیم:  $\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{C}} \{f(\mathbf{x})\}$ ، به ازای هر  $\mathbf{x} \in \mathcal{C}$  خواهیم داشت:  $\|\mathbf{x} - \mathbf{x}^*\|_2 \leq B < +\infty$ .

فرض ۴-۲. تابع  $f$ ، یک تابع  $\rho$ -لیپشیتز است. یعنی به ازای هر  $\mathbf{x}, \mathbf{y} \in \mathcal{C}$  داریم:

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq \rho \|\mathbf{x} - \mathbf{y}\|_2.$$

---

<sup>۳</sup> همی گویم بگریم در غمت زار / دگر گویم بخندی برگردن  
 گر آزادم کنی ورنده خوانی / مرا زین قید ممکن نیست جستن  
 گرم دشمن شوی و دوست گیری / نخواهم دستت از دامن گسستن [سعدی]

پرسش تئوری ۰۱۵. فرض کنید دنباله‌ی  $\{\eta_t\}_{t=1}^T$  دنباله‌ای ناصعودی باشد و همچنین فرض کنید مفروضات ۱-۴ و ۲-۴ برقرار باشند. از روند اثبات قضایای مربوط به الگوریتم ۲ کمک بگیرید و نشان دهید که در این حالت، اگر الگوریتم ۳ را با هر  $\mathbf{x}_0 \in \mathcal{C}$  اجرا کنیم، داریم:

$$\sum_{t=1}^T (f(\mathbf{x}^{(t)}) - f(\mathbf{x}^*)) \leq \frac{1}{2} B^2 + \frac{1}{2} \rho^2 \sum_{t=1}^T \eta_t.$$

پرسش تئوری ۰۱۶. قرار دهید:  $\eta_t = \frac{\alpha}{\sqrt{t}}$  و فرض کنید مفروضات ۱-۴ و ۲-۴ برقرار باشند. نشان دهید که در این حالت، اگر الگوریتم ۳ را با هر  $\mathbf{x}_0 \in \mathcal{C}$  اجرا کنیم و تعریف کنیم  $\bar{\mathbf{x}}_T = \frac{1}{T} \sum_{t=1}^T \mathbf{x}^{(t)}$ ، داریم:

$$f(\bar{\mathbf{x}}_T) - f(\mathbf{x}^*) \leq \frac{B^2}{2\alpha\sqrt{T}} + \frac{\rho^2\alpha}{\sqrt{T}}.$$

همچنین مقدار بهینه‌ی  $\alpha$  برای محاسبه‌ی بهترین کران را بیابید.



## ۵. مرا امید وصالِ تو زنده می‌دارد!<sup>۴</sup>

حالا می‌توانیم مسئله‌ای مانند مسئله‌ی (۲) را حل کنیم. برای حل این مسئله ابتدا مفهوم زیرگرادیان تصادفی را تعریف می‌کنیم:

**تعریف ۵-۱.** تابع  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  را در نظر بگیرید. بردار تصادفی  $\mathbf{V} \in \mathbb{R}^n$  را یک زیرگرادیان تصادفی برای تابع  $f$  در نقطه‌ی  $\mathbf{x} \in \mathbb{R}^n$  می‌نامیم، اگر به ازای هر  $\mathbf{y} \in \mathbb{R}^n$  داشته باشیم  $\mathbb{E}[\mathbf{V}] \in \partial f(\mathbf{x})$ ، یا به تعبیر دیگر:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{y} - \mathbf{x}, \mathbb{E}[\mathbf{V}] \rangle. \quad (۵)$$

و در نتیجه به الگوریتم کاهش گرادیان تصادفی در حالت مقید می‌رسیم:

---

### Algorithm 4 Stochastic Gradient Descent (SGD)

---

**parameters:** Set of Scalars  $\{\eta_t\}_{t=1}^T$  where for all  $t \in [T]$ ,  $\eta_t > 0$ , integer  $T > 0$ , Initial Point  $\mathbf{x}_0$

**Initialize**  $\mathbf{x}^{(1)} = \mathbf{x}_0$

**for**  $t = 1, 2, \dots, T$  **do**

    choose  $\mathbf{V}^{(t)}$  at random from a distribution such that  $\mathbb{E}[\mathbf{V}^{(t)} | \mathbf{X}^{(t)}] \in \partial f(\mathbf{X}^{(t)})$

    update  $\mathbf{X}^{(t+1)} = \Pi_C(\mathbf{X}^{(t)} - \eta \mathbf{V}^{(t)})$

**end**

**return**  $\bar{\mathbf{X}}_T = \frac{1}{T} \sum_{t=1}^T \mathbf{X}^{(t)}$

---

پرسش تئوری ۱۷. در الگوریتم ۴ بردارهای  $\mathbf{X}^{(t)}$  با حروف بزرگ نوشته شده‌اند به این دلیل که این بردارها تصادفی هستند. منشأ این خاصیت تصادفی چیست؟

پرسش تئوری ۱۸. چرا در انتخاب بردار زیرگرادیان تصادفی امید ریاضی شرطی دیده می‌شود؟ تعبیر شهودی این عبارت چیست؟

---

برای تحلیل تئوری الگوریتم ۴ نیاز به فرض ۴-۱ داریم. همچنین باید فرض ۴-۲ را به صورت زیر تغییر دهیم:

فرض ۵-۱. درباره‌ی بردار گرادیان تصادفی می‌دانیم:

$$\mathbb{E}[\|\mathbf{V}\|_2^2] \leq \rho^2.$$

پرسش تئوری ۱۹. تعریف می‌کنیم  $\mathbf{x}^* = \arg \min_{\mathbf{x} \in C} \{f(\mathbf{x})\}$ . با توجه به الگوریتم ۴، اگر تعریف کنیم  $\xi_t = \mathbf{V}^{(t)} - \mathbb{E}[\mathbf{V}^{(t)} | \mathbf{X}^{(t)}]$ ، نشان دهید:

$$\frac{1}{T} \|\mathbf{X}^{(T+1)} - \mathbf{x}^*\|_2^2 \leq \frac{1}{T} \|\mathbf{X}^{(1)} - \mathbf{x}^*\|_2^2 - \eta_t (f(\mathbf{X}^{(t)}) - f(\mathbf{x}^*)) + \frac{\eta_t^2}{T} \|\mathbf{V}^{(t)}\|_2^2 - \eta_t \langle \xi_t, \mathbf{X}^{(t)} - \mathbf{x}^* \rangle.$$

پرسش تئوری ۲۰. نشان دهید:

$$\mathbb{E}[\langle \xi_t, \mathbf{X}^{(t)} - \mathbf{x}^* \rangle] = 0.$$

پرسش تئوری ۲۱. فرض کنید  $\{\eta_t\}_{t=1}^T$  دنباله‌ای ناصعودی باشد. نشان دهید:

$$\mathbb{E}[f(\bar{\mathbf{X}}_T)] - f(\mathbf{x}^*) \leq \frac{B^2}{2T\eta_T} + \frac{1}{2T}\rho^2 \sum_{t=1}^T \eta_t.$$

---

<sup>۴</sup> هزار دشمنم آر می‌کنند قصد هلاک/ گم تو دوستی از دشمنان ندارم باک

مرا امید وصال تو زنده می‌دارد/ و گر نه هر دم از هجرت نوست بیم هلاک

نفس نفس اگر از باد نشنوم بویش/ زمان زمان چو گل از غم گتم گریبان چاک [حافظ]

پرسش تئوری ۲۲. قرار دهید  $\eta_t = \frac{B}{\rho\sqrt{t}}$ . نشان دهید:

$$\mathbb{E} [f(\bar{X}_T)] - f(x^*) \leq \frac{3B\rho}{2\sqrt{T}}.$$

پرسش تئوری ۲۳. فرض کنید تابع  $f$ ، قویاً محدب با پارامتر  $\lambda \geq 0$  باشد، یعنی به ازای هر  $x, y \in C$  و هر بردار مانند  $v$  که  $v \in \partial f(x)$ ، داشته باشیم:

$$f(y) \geq f(x) + \langle v, y - x \rangle + \frac{\lambda}{2} \|y - x\|_2^2.$$

همچنین فرض کنید مفروضات ۱-۴ و ۱-۵ برقرار باشند. نشان دهید در این حالت با انتخاب  $\eta_t = \frac{1}{\lambda t}$  خواهیم داشت:

$$\mathbb{E} [f(\bar{X}_T)] - f(x^*) \leq \frac{B^2}{2\lambda} \frac{1 + \ln(T)}{T}.$$

حال برای حل مسئله ی (۲) می‌توانیم به ازای هر تحقق بردار تصادفی  $Z$ ، مانند  $z$  و به ازای هر  $x \in C$  بردار  $v$  را به گونه‌ای تعریف کنیم که  $v \in \partial_x F(x, Z)$  باشد.

پرسش تئوری ۲۴. نشان دهید در این حالت  $V$  یک گرادین تصادفی برای تابع  $f(x) = \mathbb{E}_Z [F(x, Z)]$  است.

پرسش شبیه‌سازی ۰۳. در پرسش شبیه‌سازی ۱ الگوریتم ۴ را اجرا کنید و کمیت  $f(X^{(t)}) - f(x^*)$  را در طول الگوریتم محاسبه و نمودار آن را رسم کنید. نتیجه را با پرسش شبیه‌سازی ۱ مقایسه کنید.

پرسش شبیه‌سازی ۰۴. حجم محاسبات انجام‌شده در هر تکرار از الگوریتم ۲ و الگوریتم ۴، هنگامی که برای بهینه‌کردن تابع هدف تعریف‌شده در پرسش شبیه‌سازی ۱ استفاده می‌شوند را تقریب بزنید. حالا نموداری که در پرسش قبل برای مقایسه‌ی کمیت  $f(X^{(t)}) - f(x^*)$  در طول دو الگوریتم ۲ و ۴ رسم کرده بودید را بر حسب حجم محاسبات بازترسیم کنید.

## ۶ نبود خیر در آن خانه که عصمت نبود!

فرض کنید شما یک تحلیلگر داده در یک شرکت املاک هستید. از شما خواسته شده است که قیمت فروش خانه‌ها را براساس ویژگی‌های آن‌ها پیش‌بینی کنید. همچنین از طرف شرکت به شما تعدادی خانه با قیمت مشخص داده شده است تا مدل را براساس آن‌ها آموزش دهید. ویژگی‌هایی که از این خانه‌ها به شما داده شده است، شامل مساحت خانه برحسب متر مربع، تعداد اتاق خواب‌ها، تعداد سرویس‌های بهداشتی، عمر خانه برحسب سال و مکان خانه است. می‌خواهیم از مدل رگرسیون خطی برای مدل پیش‌بینی استفاده کنیم و برای آموزش این مدل نیز از الگوریتم SGD استفاده خواهیم کرد.

**تعریف ۱-۶.** فرض کنید مجموعه دادگانی از  $n$  مشاهده دارید که هر کدام دارای  $m$  ویژگی و یک مقدار خروجی هستند.  $\mathbf{X} \in \mathbb{R}^{n \times m}$  ماتریس ویژگی است که در آن هر سطر یک بردار ویژگی برای یک مشاهده است. همچنین  $\mathbf{y} \in \mathbb{R}^n$  بردار هدف است. در این بردار هر عنصر نماینده‌ی مقدار خروجی مشاهده‌شده است.  $\mathbf{w} \in \mathbb{R}^m$  بردار وزن است که در آن هر عنصر وزنی که به هر ویژگی نسبت می‌دهیم را تعیین می‌کند. همچنین یک پارامتر اسکالر مانند  $b$  برای مدل در نظر می‌گیریم. در این صورت مقدار پیش‌بینی شده با مدل رگرسیون خطی به صورت زیر است:

$$\hat{y} = \mathbf{X}\mathbf{w} + b\mathbf{1}.$$

که  $\mathbf{1} \in \mathbb{R}^n$  بردار تمام‌یک است.

هدف از یک مسئله رگرسیون خطی تخمین پارامترهای  $\mathbf{w}$  و  $b$  است.

**تعریف ۲-۶.** خطای میانگین مربعات یا همان  $MSE$  به صورت زیر تعریف می‌شود:

$$l(\mathbf{w}, b) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2.$$

که  $y_i$  برابر با مقدار واقعی و  $\hat{y}_i$  برابر با مقدار پیش‌بینی شده است.

**پرسش شبیه‌سازی ۵.** دادگان `houseData.csv` را فراخوانی کنید و نمودارهایی برای ارزیابی هر ویژگی آن رسم کنید. برای مثال قیمت خانه را براساس تعداد اتاق خواب‌ها، تعداد سرویس‌های بهداشتی، مکان و سن خانه رسم کنید و تحلیل خود را از دادگان و تاثیر هر ویژگی بنویسید.

**پرسش شبیه‌سازی ۶.** مجموعه دادگان را به دو دسته آموزش و آزمون تقسیم کنید. همچنین با کم کردن میانگین و نقسیم کرنم برانحراف معیار دادگان را بهنجار<sup>۵</sup> کنید.

**پرسش شبیه‌سازی ۷.** تابعی برای محاسبه میانگین مربعات خطا ( $MSE$ ) بین قیمت‌های پیش‌بینی شده و واقعی تعریف کنید.

**پرسش شبیه‌سازی ۸.** تابعی برای اجرای یک مرحله از الگوریتم SGD تعریف کنید. این تابع باید در ورودی پارامترهای فعلی که شامل  $\mathbf{x}$ ,  $b$  هستند و همچنین دسته‌ای از دادگان را بگیرد و در خروجی نیز پس از انجام الگوریتم پارامترهای به‌روزرشده را برگرداند.

**پرسش شبیه‌سازی ۹.** در یک حلقه ترتیب داده‌های آموزش را به طور تصادفی عوض کنید. همچنین دادگان بخش آموزش را به تعدادی دسته<sup>۷</sup> تقسیم کنید و برای هر دسته الگوریتم SGD را اجرا کنید. نهایتاً با پارامترهای بدست آمده قیمت خانه‌های بخش آموزش و آزمون را پیش‌بینی کنید و با توجه به مقدار واقعی قیمت‌ها، مقدار خطا برای هر یک از آن‌ها به دست آورید. این حلقه باید چند دوره<sup>۸</sup> اجرا شود. مقدار خطا برای هر یک از دوره‌ها برای دادگان آموزش و آزمون را در یک نمودار رسم کنید.

تعداد دوره‌ها، اندازه‌ی هر دسته و طول گام الگوریتم SGD را به طور دلخواه انتخاب کنید.

<sup>۵</sup> گر مدد خواستم از پیر مغان عیب مکن / شیخ ما گفت که در صومعه همت نبود

چون طهارت نبود کعبه و بنخانه یکست / نبود خیر در آن خانه که عصمت نبود  
حافظا علم و ادب ورز که در مجلس شاه / هر که را نیست ادب لایق صحبت نبود [حافظ]

<sup>۶</sup>Normalize

<sup>۷</sup>batch

<sup>۸</sup>batch

## ۷ نصیحتی کُنت بشنو و بهانه مگیر!<sup>۹</sup>

لطفاً به نکات زیر دقت کنید:

۱. این پروژه بخشی از نمره‌ی شما در این درس را تشکیل خواهد داد.
۲. می‌توانید پروژه را در قالب گروه‌های ۲ نفره انجام دهید. فرمی برای ثبت گروه‌ها در اختیار شما قرار خواهد گرفت. دقت داشته باشید که در هنگام تحویل پروژه باید تمامی اعضای گروه به تمامی بخش‌ها مسلط باشند و در نهایت همه‌ی اعضای یک گروه نمره‌ی واحدی را دریافت خواهند کرد.
۳. عنوان بخش‌های مختلف پروژه از آثار شعرا و بزرگان ادبیات فارسی انتخاب شده است. این اشعار بی‌ربط به مفاهیمی که در هر بخش با آن‌ها برخورد می‌کنید نیستند.
۴. تمامی شبیه‌سازی‌ها باید با کمک زبان Python انجام شود. شما تنها مجاز به استفاده از کتابخانه‌های `plotly`، `random`، `scipy`، `numpy` و `matplotlib` هستید. اگر روی عنوان هر کتابخانه کلیک کنید، به راهنمای آن کتابخانه هدایت می‌شوید.
۵. مجموعه‌داده‌ی مورد استفاده در پرسش‌های شبیه‌سازی در CW بارگذاری شده است.
۶. تحویل پروژه به صورت گزارش و کدهای نوشته‌شده است. گزارش باید شامل پاسخ پرسش‌ها، تصاویر و نمودارها و نتیجه‌گیری‌های لازم باشد. توجه کنید که قسمت عمده بارم شبیه‌سازی را گزارش شما و نتیجه‌ای که از خروجی کد می‌گیرید دارد. همچنین تمیزی گزارش بسیار مهم است. کدها و گزارش را در یک فایل فشرده‌شده در سامانه‌ی درس‌افزار آپلود کنید.
۷. اگر برای پاسخ به پرسش‌ها، از منبعی (کتاب، مقاله، سایت و...) کمک گرفته‌اید، حتماً به آن ارجاع دهید.
۸. نوشتن گزارش کار با  $\text{IAT}_{\text{EX}}$  نمره‌ی امتیازی دارد.
۹. پرسش‌های شبیه‌سازی با رنگ سبز و پرسش‌های تئوری با رنگ آبی مشخص شده‌اند.
۱۰. بخش‌های تئوری گزارش که در قالب پرسش‌ها طرح شده‌اند را می‌توانید روی کاغذ بنویسید و تصویر آن‌ها را در گزارش خود بیاورید، ولی توصیه‌ی برادرانه می‌کنم که این کار را نکنید!
۱۱. در صورت مشاهده‌ی تقلب، نمره‌ی هردو فرد صفر منظور خواهد شد.

موفق باشید!

---

<sup>۹</sup> نصیحتی کُنت بشنو و بهانه مگیر/ هر آنچه ناصح مُشَفِّق بگویدَت بپذیر [حافظ]