

جبر خطی

دانشکده مهندسی کامپیوتر

حمیدرضا ربیعی، مریم رمضانی
بهار ۱۴۰۴



فضای نرم، مشتق ماتریس و بردار، کمترین مربعات

تمرین ششم

تاریخ انتشار: ۴ خرداد ۱۴۰۴

۱. پرسش‌های خود درمورد این تمرین را در سامانه کوئرا مطرح کنید.

۲. سیاست ارسال با تاخیر: شما در مجموع در طول نیم‌سال می‌توانید از ۱۶ روز تاخیر استفاده کنید. همچنین هر تمرین تئوری و عملی را می‌توانید تا حداکثر ۳ روز با تاخیر تحویل دهید. این مقدار برای تمارین تئوری و عملی به صورت جداگانه حساب می‌شود. تاخیرها با مقیاس ساعت محاسبه شده و به بالا گرد می‌شوند.

۳. سیاست مشارکت دانشجویان در حل کردن تمارین: دانشجویان می‌توانند در حل تمارین برای رفع ابهام و یا به دست آوردن ایده‌ی کلی با یکدیگر مشورت و همفکری کنند. این کار مورد تایید و تشویق تیم ارائه‌ی درس می‌باشد؛ چرا که هم‌فکری و کار گروهی می‌تواند موجب تقویت یادگیری شود. اما به دست آوردن جزئیات راه‌حل و نگارش پاسخ باید تماماً توسط خود دانشجو انجام شود. حتماً در انتهای پاسخ‌های ارسالی خود نام افرادی که با آن‌ها همفکری کردید را ذکر کنید.

سوالات تئوری

پرسش ۱ ما دو نرم $\|\cdot\|_a$ و $\|\cdot\|_b$ را معادل یکدیگر تعریف می‌کنیم اگر و تنها اگر ثابت‌های c_1 و c_2 وجود داشته باشند به طوری که:

$$\forall v : c_1 \|v\|_a < \|v\|_b < c_2 \|v\|_a$$

ثابت کنید تمام نرم‌های طبیعی بردار معادل یکدیگرند.

پرسش ۲ برای فضای ماتریس‌های $n \times n$ بررسی کنید کدام یک از تعاریف نرم‌های زیر درست‌اند. (منظور از $|A|$ در اینجا، ماتریسی است که تمام درایه‌های آن قدرمطلق درایه‌ی متناظر با ماتریس A است.)

(الف) $\|A\| = \det(|A|)$

(ب) $\|A\| = \text{tr}(|A|)$

(ج)

$$\|A\| = \max \left(\frac{\|Av\|_p}{\|v\|_p} \right)$$

پرسش ۳ نشان دهید اگر $v \in \mathbb{R}^n$ و $v \neq 0$ داشته باشیم $E \in \mathbb{R}^{n \times n}$ ، آن گاه عبارت زیر برقرار خواهد بود:

$$\left\| E \left(I - \frac{vv^T}{v^T v} \right) \right\|_F^2 = \|E\|_F^2 - \frac{\|Ev\|_2^2}{v^T v}$$

تعریف نرم فروبنیوس: $\|A\|_F^2 = \text{trace}[AA^T]$

راهنمایی: خواص orthogonal projection را برای $P := \frac{vv^T}{v^T v}$ بررسی کنید.

پرسش ۴ رگرسیون خطی وظیفه یافتن بهترین برازش خطی بین برجسب‌های $\mathbf{y} \in \mathbb{R}^n$ و ویژگی‌های $\mathbf{X} \in \mathbb{R}^{m \times n}$ را دارد. به طور مشخص، ما بردار مناسب θ را تعیین می‌کنیم به طوری که $\mathbf{y} = \mathbf{X}\theta$. یکی از بهترین انتخاب‌های ممکن برای θ برداری است که خطای میانگین مربعات را کمینه کند. این مقدار به صورت زیر تعریف می‌شود:

$$J(\theta) = \frac{1}{N} (\mathbf{X}\theta - \mathbf{y})^T (\mathbf{X}\theta - \mathbf{y}).$$

مقدار θ را طوری بیابید که خطای میانگین مربعات را کمینه کند.

پرسش ۵ نشان دهید تابع زیر محدب است:

$$f(x) = \log \left(\sum_{k=1}^n \exp(x_k) \right)$$

راهنمایی: برای این که نشان دهید $f(x)$ محدب است باید نشان دهید به ازای همه مقادیر v ، $v^T \nabla^2 f(x) v \geq 0$.

پرسش ۶ جواب کمترین مربعات دستگاه معادلات زیر را بیابید.

$$\begin{aligned}x - z &= -1 \\ 2x - y + 3z &= 1 \\ y - 3z &= 0 \\ -5x + 2y + z &= 3\end{aligned}$$

پرسش ۷ اگر برخی از درایه های بردار $Ax - b$ از دیگر درایه ها مهم تر باشند، میتوان وزن $w_i > 0$ را به هر درایه متناظر کرد. با تشکیل ماتریس قطری W که به شکل زیر تعریف می شود، می توان جواب تقریب کمترین مربعات مسئله جدید وزن دار را با پیدا کردن $\|W(Ax - b)\|_2$ پیدا کرد.

$$W = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{bmatrix}$$

به رویکرد فوق تقریب کمترین مربعات وزن دار گفته می شود.

(آ) نشان دهید که ماتریس وزن W متقارن معین مثبت است.

(ب) با الهام گرفتن از همان شیوه ای که معادلات نرمال $A^T Ax = A^T b$ را به دست آوردیم، معادلات نرمال تقریب کمترین مربعات وزن دار را به صورت $A^T W A x = A^T W b$ تشکیل دهید.

(ج) فرض کنید $W^{1/2} = \text{diag}(\sqrt{w_1} + \sqrt{w_2} + \cdots + \sqrt{w_n})$ باشد، به طوری که $(W^{1/2})^2 = W$. معادلات نرمال قسمت ب را به نحوی بازنویسی کنید که تبدیل به معادلات نرمال معمولی شوند با این تفاوت که به جای ماتریس A از ماتریس مقیاس شده $W^{1/2} A$ استفاده شده است.

پرسش ۸ در برخی کاربردها، لازم است نقاط داده را با چیزی غیر از یک خط مستقیم برازش داد. معادله ماتریسی $X\beta = y$ را در نظر بگیرید، آمارشناسان معمولاً یک بردار باقیمانده (ϵ) معرفی می کنند که به صورت زیر تعریف می شود:

$$\epsilon = y - X\beta$$

و می نویسند:

$$y = X\beta + \epsilon$$

هر معادله ای از این فرم به عنوان یک مدل خطی شناخته می شود. زمانی که X و y تعیین شدند، هدف کمینه سازی طول ϵ است که معادل یافتن جواب حداقل مربعات برای $X\beta = y$ است. در هر مورد، جواب حداقل مربعات $\hat{\beta}$ جوابی برای معادلات نرمال زیر است:

$$X^T X \beta = X^T y$$

حال فرض کنید:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n), \bar{y} = \frac{1}{n}(y_1 + y_2 + \cdots + y_n)$$

نشان دهید که خط حداقل مربعات برای داده های $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ باید از نقطه $(\bar{x}, \bar{y})$ عبور کند. به عبارت دیگر، نشان دهید که $\bar{y}$ و در معادله خطی زیر صدق می کنند:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

پاسخ ۱ ابتدا ثابت می کنیم تمام نرم های طبیعی با نرم بی نهایت معادلند و پس از آن به سادگی اثبات می شود که تمام نرم ها با یکدیگر معادلند. برای اینکار فرض می کنیم فضای برداری R^n است و ثابت می کنیم نرم p با نرم ∞ معادل است. واضح است نرم p همواره بزرگتر از نرم بی نهایت است. همچنین اگر تمام مولفه های بردار برابر x باشند، داریم

$$|v|_p = \sqrt[p]{n \cdot x^p}$$

و نتیجه می دهد که

$$|v|_p = \sqrt[p]{n} \cdot x = \sqrt[p]{n} \cdot |v|_\infty$$

و این بیشینه ی مقدار نرم p به ازای یک نرم بی نهایت خاص است. بنابراین همواره داریم:

$$|v|_\infty \leq |v|_p \leq \sqrt[p]{n} |v|_\infty$$

پس تمام نرم های طبیعی با نرم بی نهایت معادلند. به سادگی می توان اثبات کرد که اگر نرم a, b معادل یکدیگر و نرم b, c معادل یکدیگر باشند، آنگاه نرم a, c معادلند و می توان نتیجه بگیریم که تمام نرم های طبیعی معادل یکدیگرند.

پاسخ ۲ (الف) این تعریف نرم غلط است. می‌دانیم طبق تعریف نرم، نرم یک بردار تنها زمانی صفر است که خود آن برابر صفر شود. اما در تعریف دترمینان ماتریس‌هایی وجود دارند که صفر نیستند ولی دترمینان آنها صفر اند (و تمام درایه‌هایشان مثبت است). برای مثال، ماتریسی که تمام درایه‌های آن برابر یک است دترمینان صفر دارد ولی برابر صفر نیست.

(ب) این تعریف نرم غلط است. می‌دانیم طبق تعریف نرم، نرم یک بردار تنها زمانی صفر است که خود آن برابر صفر شود. اما دوباره ماتریس‌هایی وجود دارند که تمام درایه‌های قطر اصلی آنها صفرند ولی خودشان برابر صفر نیستند.

(ج) این تعریف نرم صحیح است. مشخص است که این تعریف نرم همواره نامنفی است. همچنین اگر نرم فوق صفر شود، یعنی به ازای تمامی بردارها ضرب ماتریس در آنها صفر شده و نتیجه می‌دهد که ماتریس صفر است. (می‌توانیم برای پی بردن به این موضوع، ماتریس را در بردارهای یکه ضرب کنیم و نتیجه بگیریم که تمام ستون‌های آن صفر است.) همچنین داریم:

$$|aA| = \max \left(\frac{|aAv|_p}{|v|_p} \right) = |a| \max \left(\frac{|Av|_p}{|v|_p} \right) = |a| \|A\|$$

حال کافی است نامساوی مثلث را اثبات کنیم. برای سادگی تعریف نرم را کمی عوض می‌کنیم.

$$\|A\| = \max (|Av|_p), \quad |v|_p = 1$$

مشخص است که این تعریف با تعریف قبلی سازگار است. حال، فرض می‌کنیم بردار بیشینه برای $(A+B)v|_p$ در u است. یعنی:

$$\|A+B\| = |(A+B)u|_p$$

بنابر نامساوی مثلث برای بردارها داریم:

$$|(A+B)u|_p \leq |Au|_p + |Bu|_p$$

از طرفی داریم:

$$|Au|_p \leq \|A\|, \quad |Bu|_p \leq \|B\|$$

بنابراین:

$$\|A+B\| \leq \|A\| + \|B\|$$

پاسخ ۳ پاسخ فرض کنید داشته باشیم:

$$P := \frac{vv^T}{v^T v}$$

حال با استفاده از تعریف نرم فروبنیوس عبارت داده شده را ساده می‌کنیم:

$$\|E - EP\|_F^2 = \text{trace}[(E - EP)(E - EP)^T] = \text{trace}(EE^T - EP^T E^T - EPE^T + EPP^T E^T)$$

با توجه به تعریف P می‌دانیم که یک orthogonal projection است که نتیجه می‌دهد:

$$P^T = P, \quad P^2 = PP^T = P$$

$$\begin{aligned} \|E - EP\|_F^2 &= \text{trace}(EE^T - EPE^T - EPE^T + EPE^T) \\ &= \text{trace}(EE^T - EPE^T) = \text{trace}(EE^T - EP(EP)^T) \end{aligned}$$

$$= \text{trace}(EE^T) - \text{trace}[EP(EP)^T] = \|E\|_F^2 - \|EP\|_F^2$$

حال کافی است نشان دهیم: $\|EP\|_F^2 = \frac{\|Ev\|_2^2}{v^T v}$.

$$\begin{aligned} \|EP\|_F^2 &= \text{trace}[(EP)^T EP] = \text{trace}(P^T E^T EP) \\ &= \text{trace} \left(\frac{v v^T E^T E v v^T}{(v^T v)^2} \right) \\ &= \frac{v^T E^T E v \cdot v^T v}{(v^T v)(v^T v)} = \frac{\|Ev\|_2^2}{v^T v} \end{aligned}$$

پاسخ ۴ با باز کردن عبارت J داریم:

$$\begin{aligned} J(\theta) &= \frac{1}{N}(\mathbf{X}\theta - \mathbf{y})^T(\mathbf{X}\theta - \mathbf{y}) \\ &= \frac{1}{N}(\theta^T \mathbf{X}^T - \mathbf{y}^T)(\mathbf{X}\theta - \mathbf{y}) \\ &= \frac{1}{N}(\theta^T \mathbf{X}^T \mathbf{X}\theta - \theta^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\theta + \mathbf{y}^T \mathbf{y}) \end{aligned}$$

می‌دانیم که ماتریس $\mathbf{X}^T \mathbf{X}$ متقارن است پس داریم:

$$\frac{d(\theta^T \mathbf{X}^T \mathbf{X}\theta)}{d\theta} = 2\mathbf{X}^T \mathbf{X}\theta$$

همچنین مشتق‌های زیر را به دست می‌آوریم.

$$\frac{d(\theta^T \mathbf{X}^T \mathbf{y})}{d\theta} = \mathbf{X}^T \mathbf{y} \quad \& \quad \frac{d(\mathbf{y}^T \mathbf{X}\theta)}{d\theta} = \mathbf{X}^T \mathbf{y} \quad \& \quad \frac{d(\mathbf{y}^T \mathbf{y})}{d\theta} = 0$$

$$\begin{aligned} \Rightarrow \frac{dJ(\theta)}{d\theta} &= \frac{1}{N} \left(\frac{d(\theta^T \mathbf{X}^T \mathbf{X}\theta)}{d\theta} - \frac{d(\theta^T \mathbf{X}^T \mathbf{y})}{d\theta} - \frac{d(\mathbf{y}^T \mathbf{X}\theta)}{d\theta} + \frac{d(\mathbf{y}^T \mathbf{y})}{d\theta} \right) \\ &= \frac{1}{N} (2\mathbf{X}^T \mathbf{X}\theta - \mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{y} + 0) \\ &= \frac{2}{N} (\mathbf{X}^T \mathbf{X}\theta - \mathbf{X}^T \mathbf{y}). \end{aligned}$$

با قرار دادن این عبارت مساوی با 0 و فرض وارون پذیری $\mathbf{X}^T \mathbf{X}$ داریم:

$$\begin{aligned} \mathbf{X}^T \mathbf{X}\theta - \mathbf{X}^T \mathbf{y} &= 0 \\ \Rightarrow \mathbf{X}^T \mathbf{X}\theta &= \mathbf{X}^T \mathbf{y} \\ \Rightarrow \theta &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \end{aligned}$$

پاسخ ۵ ابتدا مشتق $f(x)$ نسبت به x_i را به دست می‌آوریم:

$$\frac{\partial f(x)}{\partial x_i} = \frac{\exp(x_i)}{\sum_{k=1}^n \exp(x_k)}$$

حال مشتق دوم را برای دو حالت به دست می‌آوریم:

$$\frac{\partial^2 f(x)}{\partial x_i^2} = \frac{\partial \left(\frac{\exp(x_i)}{\sum_{k=1}^n \exp(x_k)} \right)}{\partial x_i} = \frac{\exp(x_i)}{\sum_{k=1}^n \exp(x_k)} - \frac{(\exp(x_i))^2}{(\sum_{k=1}^n \exp(x_k))^2}$$

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial \left(\frac{\exp(x_i)}{\sum_{k=1}^n \exp(x_k)} \right)}{\partial x_j} = - \frac{\exp(x_i) \exp(x_j)}{(\sum_{k=1}^n \exp(x_k))^2}$$

بردار z را برداری تعریف می‌کنیم که مولفه i ام آن $\exp(x_i)$ باشد. در نتیجه عبارات بالا به شکل زیر ساده می‌شوند:

$$\frac{\partial^2 f(x)}{\partial x_i^2} = \frac{z_i}{\sum_{k=1}^n z_k} - \frac{(z_i)^2}{(\sum_{k=1}^n z_k)^2} = \frac{z_i}{1^T z} - \frac{(z_i)^2}{(1^T z)^2}$$

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = - \frac{z_i z_j}{(\sum_{k=1}^n z_k)^2} = - \frac{z_i z_j}{(1^T z)^2}$$

حال با کنار هم قرار دادن دو عبارت بالا خواهیم داشت:

$$\nabla^2 f(x) = \frac{\text{diag}(z)}{1^T z} - \frac{z z^T}{(1^T z)^2}$$

کافی است نشان دهیم $v^T \nabla^2 f(x) v \geq 0$

$$\begin{aligned} v^T \nabla^2 f(x) v &= v^T \frac{\text{diag}(z)}{\sqrt{z}} v - v^T \frac{z z^T}{(\sqrt{z})^2} v \\ &= \frac{\sum_{i=1}^n v_i^2 z_i}{\sqrt{z}} - \frac{\|z^T v\|_2^2}{(\sqrt{z})^2} \\ &= \frac{(\sum_{i=1}^n v_i^2 z_i) (\sum_{i=1}^n z_i) - \sum_{i=1}^n v_i^2 z_i^2}{(\sqrt{z})^2} \end{aligned}$$

طبق نامساوی کشی-شوارتز می‌دانیم:

$$\left(\sum_{i=1}^n v_i^2 z_i \right) \left(\sum_{i=1}^n z_i \right) \geq \sum_{i=1}^n v_i^2 z_i^2$$

از آن جا که مخرج نیز مثبت است خواهیم داشت:

$$\frac{(\sum_{i=1}^n v_i^2 z_i) (\sum_{i=1}^n z_i) - \sum_{i=1}^n v_i^2 z_i^2}{(\sqrt{z})^2} \geq 0$$

پس نشان دادیم به ازای هر v دلخواه $v^T \nabla^2 f(x) v \geq 0$ که نتیجه می‌دهد $f(x)$ محدب است.

پاسخ ۶ بررسی تعداد جواب ها

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & -1 & 3 \\ 0 & 1 & -3 \\ -5 & 2 & 1 \end{bmatrix} \begin{bmatrix} s \\ t \\ m \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 3 \end{bmatrix}$$

$\text{Rank}(A) = 3 \neq \text{Rank}([A|b]) = 4$ پس می‌توانیم نتیجه بگیریم دستگاه سازگار نیست.

$$B = AA^T = \begin{bmatrix} 1 & 2 & 0 & -5 \\ 0 & -1 & 1 & 2 \\ -1 & 3 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 2 & -1 & 3 \\ 0 & 1 & -3 \\ -5 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 30 & -12 & 0 \\ -12 & 6 & -4 \\ 0 & -4 & 20 \end{bmatrix}$$

$$c = A^T b = \begin{bmatrix} 1 & 2 & 0 & -5 \\ 0 & -1 & 1 & 2 \\ -1 & 3 & -3 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} -14 \\ 5 \\ 7 \end{bmatrix}$$

$$BX^* = c \Rightarrow \begin{bmatrix} 30 & -12 & 0 \\ -12 & 6 & -4 \\ 0 & -4 & 20 \end{bmatrix} X^* = \begin{bmatrix} -14 \\ 5 \\ 7 \end{bmatrix}$$

با استفاده از یکی از روش‌های حل خواهیم داشت:

$$X^* = \begin{bmatrix} \frac{1}{4} \\ \frac{2}{3} \\ \frac{3}{4} \end{bmatrix}$$

$$r^* = b - AX^* = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 3 \end{bmatrix} - \begin{bmatrix} \frac{1}{4} - \frac{2}{3} + \frac{9}{4} \\ \frac{2}{3} - \frac{2}{3} + \frac{9}{4} \\ \frac{2}{3} - \frac{9}{4} \\ -\frac{5}{4} + \frac{4}{3} + \frac{3}{4} \end{bmatrix} = \begin{bmatrix} -\frac{7}{12} \\ \frac{1}{12} \\ \frac{1}{4} \\ -\frac{1}{12} \end{bmatrix}$$

که نشان دهنده این است که همان‌طور که در اول آورده شد دستگاه اصلی ناسازگار است در حالی که جواب کمترین مربعات را دارد.

$$\|r^*\|_2 = \left\| \begin{bmatrix} -\frac{7}{12} & \frac{1}{12} & \frac{1}{4} & -\frac{1}{12} \end{bmatrix}^T \right\|_2 = 0.6455$$

پاسخ ۷ (آ) چون W قطری است، متقارن هم می‌باشد. یعنی $W = W^T$.

برای اثبات معین مثبت بودن W باید نشان دهیم برای هر بردار دلخواه غیر صفر X داریم که $X^T W X > 0$. با فرض اینکه داشته باشیم $X = [x_1 \ x_2 \ \dots \ x_n]^T$ ، آنگاه:

$$X^T W X = [x_1 \ x_2 \ \dots \ x_n] \begin{bmatrix} w_1 & 0 & 0 & \dots & 0 \\ 0 & w_2 & 0 & \dots & 0 \\ 0 & 0 & w_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & w_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^n w_i x_i^2 > 0$$

چرا که $w_i > 0$ به ازای هر $1 \leq i \leq n$ و همچنین وجود دارد حداقل یک j ($1 \leq j \leq n$) که $x_j \neq 0$. پس W متقارن معین مثبت است.

(ب) هدف حل مساله $\min_{X \in \mathbb{R}^n} \|W(b - AX)\|^2$ می باشد.
با در نظر گرفتن نرم-۲ و تعریف $f(X) = \|W(b - AX)\|^2$ خواهیم داشت:

$$\begin{aligned} f(X) &= \|W(b - AX)\|^2 = (W(b - AX))^T (W(b - AX)) \\ &= ((b - AX)^T W^T) (W(b - AX)) \\ &= (b^T - X^T A^T) W^T (W(b - AX)) \\ &= (b^T W^T - X^T A^T W^T) (Wb - WAX) \\ &= b^T W^T Wb - b^T W^T WAX - X^T A^T W^T Wb + X^T A^T W^T WAX \end{aligned}$$

با توجه به اینکه $b^T W^T WAX$ و $X^T A^T W^T Wb$ هر دو اسکالر هستند

$$X^T A^T W^T Wb = (X^T A^T W^T Wb)^T = b^T W^T WAX$$

و لذا خواهیم داشت :

$$f(X) = b^T W^T Wb - 2b^T W^T WAX + X^T A^T W^T WAX$$

با تعریف $b' := Wb$ و $A' := WA$ خواهیم داشت که $f(X) = b'^T b' - 2b'^T A'X + X^T A'^T A'X$ و طبق قضیه ای داریم :

$$\nabla f(X) = 2A'^T A'X - 2A'^T b'$$

پس اگر X^* جواب $\min_{X \in \mathbb{R}^n} \|W(b - AX)\|^2$ باشد، باید گرادین f در X^* صفر بوده و لذا :

$$\nabla f(X^*) = 2A'^T A'X^* - 2A'^T b' = 2A^T W^T WAX^* - 2A^T W^T Wb = 0$$

چون W متقارن است، $W = W^T$ و در نتیجه معادلات نرمال تقریب کمترین مربعات وزن دار به صورت $A^T W^T A X^* = A^T W^T b$ به دست می آید.
در ادامه فضا را تغییر داده و اگر $\langle x, y \rangle_S = x^T S y$ را در نظر بگیریم، فضای جدید را مجهز به نرم به دست آمده از این ضرب داخلی کرده و مساله کمینه سازی را با استفاده از این نرم حل می کنیم. توجه داریم که در تمرین پنجم سری اول ثابت کردیم که اگر S متقارن معین مثبت باشد، $\langle x, y \rangle_S$ یک ضرب داخلی معتبر است و چون W متقارن معین مثبت است، $\langle x, y \rangle_W$ معتبر است. پس هدف حل مساله $\min_{X \in \mathbb{R}^n} \|AX - b\|^2$ با این نرم است.

با تعریف $g(X) = \|AX - b\|^2$ خواهیم داشت :

$$\begin{aligned} g(X) &= \|AX - b\|^2 \\ &= \langle AX - b, AX - b \rangle_W \\ &= (AX - b)^T W (AX - b) \\ &= X^T A^T WAX - X^T A^T Wb - b^T WAX + b^T Wb \\ &= X^T A^T WAX - 2b^T WAX + b^T Wb \end{aligned}$$

با تعریف $b'' := W^{\frac{1}{2}} b$ و $A'' := W^{\frac{1}{2}} A$ خواهیم داشت :

$$g(X) = X^T A''^T A'' X - 2b''^T A'' X + b''^T b''$$

و طبق قضیه ای داریم :

$$\begin{aligned} \nabla g(X) &= 2A''^T A'' X - 2A''^T b'' \\ \implies A''^T A'' X^* &= A''^T b'' \\ \implies A^T WAX^* &= A^T Wb \end{aligned}$$

(ج) توجه داریم که $W^{\frac{1}{2}}$ نیز متقارن است، پس :

$$\begin{aligned} A^T WAX^* = A^T Wb &\implies A^T (W^{\frac{1}{2}})^T W^{\frac{1}{2}} A X^* = A^T (W^{\frac{1}{2}})^T W^{\frac{1}{2}} b \\ &\implies (W^{\frac{1}{2}})^T W^{\frac{1}{2}} A X^* = (W^{\frac{1}{2}})^T W^{\frac{1}{2}} b \end{aligned}$$

با تعریف $b''' := W^{\frac{1}{2}} b$ و $A''' := W^{\frac{1}{2}} A$ خواهیم داشت :

$$A'''^T A''' X^* = A'''^T b'''$$

که به فرم معادلات نرمال معمولی است.

پاسخ ۸ در بحث مدل های خطی با بردار w و ماتریس زیر سر و کار داریم :

$$y_{n \times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X_{n \times 2} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta_{2 \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \varepsilon_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

همانطور که در صورت سوال گفته شده، جواب حداقل مربعات $\hat{\beta}$ جوابی برای معادلات نرمال $X^T X \beta = X^T y$ است. لذا $\hat{\beta}$ در این معادلات نرمال صدق می‌کند.

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

که از آن به دو معادله زیر می‌رسیم:

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum y_i \\ \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i \end{cases}$$

کافی است تنها معادله اول را در نظر گرفته و دو طرف را بر n تقسیم کنیم.

$$\hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}$$

که همان مطلوب سوال است. با اینکه صورت سوال معادلات نرمال را به طور مستقیم در اختیار گذاشته بود، دانشجویان می‌توانستند با کمینه کردن $\sum \varepsilon_i = \sum (y_i - \hat{y}_i)^2$ نسبت به تنها یک متغیر $\hat{\beta}_0$ به معادله مطلوب برسند، که $\hat{y}_i$ مقدار برازش شده y_i بوده و به صورت $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ است. خواهیم داشت: (در آمار ε_i^2 را با SSE نیز نمایش می‌دهند).

$$\begin{aligned} \frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_0} &= \frac{\partial SSE}{\partial \hat{\beta}_0} = \frac{\partial}{\partial \hat{\beta}_0} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\ &= -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\ &= -2 \left(\sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n x_i \right) = 0 \end{aligned}$$

لذا خواهیم داشت:

$$\sum_{i=1}^n y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i$$

که با تقسیم طرفین بر n به معادله مطلوب می‌رسیم.

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$