تمرین بهینه سازی سری اول

غلامرضا دار ۴۰۰۱۳۱۰۱۸

سوال ۱

a) باید اثبات کنیم $N(A^TA) = N(A)$ $\quad N$: nullspace

$x \in N(A)$ فرض

$$\text{I} \quad Ax = 0 \xrightarrow{\times A^T} A^TAx = 0 \rightarrow x \in N(A^TA) \rightarrow N(A) \subseteq N(A^TA)$$

$x \in N(A^TA)$ فرض

$$\text{II} \quad A^TAx = 0 \rightarrow x^TA^TAx = 0 \rightarrow (Ax)^T(Ax) = 0 \rightarrow Ax = 0 \rightarrow x \in N(A)$$

$$\rightarrow N(A^TA) \subseteq N(A)$$

$$\text{I}, \text{II} \Rightarrow N(A^TA) = N(A) \rightarrow \dim(N(A^TA)) = \dim(N(A))$$

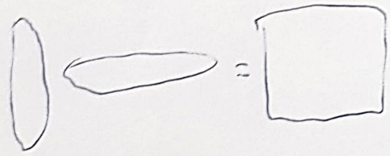
$$\rightarrow \text{Rank}(A^TA) = \text{Rank}(A) \checkmark$$

$* \; \text{Rank}(B) = \dim(N(B))$

b)

$$\underset{(n\times1)}{v} \; \underset{(1\times m)}{w^T} = \underset{(n\times m)}{A}$$

$$\begin{bmatrix} v_1 w_1^T & v_1 w_2^T & \cdots \\ v_2 w_1^T & v_2 w_2^T & \\ v_3 w_1^T & v_3 w_2^T & \\ \vdots & \vdots & \end{bmatrix}$$



همانطور که دیده می‌شود همه سطرها ضریب همدیگر و همه ستون‌ها نیز ضریب یکدیگر هستند که با این وجود یعنی همه سطرها و ستون‌ها وابسته خطی اند

پس

$$\text{Rank}(A) = 1$$

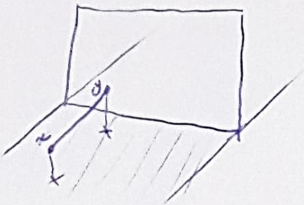
c) توابع خطی باید از نقطه $\vec{0}$ عبور کنند و در واقع هیچ‌گونه Translation ی ندارند اما توابع Affine باشند توابع خطی ای هستند که Translation هم دارند.

Linear : Ax

Affine : $Ax + \vec{b}$

سوال 2

اگر به نقاط مجموعه را بررسی کنیم می‌بینیم که همه نقاط در یک سمت یک ابرصفحه قرار می‌گیرند و در واقع مجموعه داده شده یک half space است. می‌دانیم پاره‌خط متصل کننده هر دو نقطه در یک half space در همان half space باقی می‌ماند.



سوال 3

روش 1 :

$S_1 + S_2$ is Convex $\Rightarrow (x + x , y_1 + y_2)$

$(2x , y_1 + y_2)$

$f(\alpha, \beta) = (\frac{\alpha}{2}, \beta) \longrightarrow f(2x, y_1 + y_2) = (x, y_1 + y_2)$ is Convex

$f = Ax \qquad A = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{\alpha}{2} \\ \beta \end{bmatrix}$

f یک تابع Affine است.

سوال 4

نشان دهید که یک مجموعه محدب است اگر و تنها اگر اشتراک آن با هر خطی محدب باشد.

I) ابتدا اثبات کنیم اگر یک مجموعه محدب باشد، اشتراک آن با هر خطی محدب است.

C مجموعه محدب

L یک خط

اشتراک دارند C و L

$$\begin{cases} x, y \in C \\ x, y \in L \end{cases} \Rightarrow \theta x + (1-\theta) y \in C \quad (\text{تعریف محدب بودن } C)$$

همچنین $\theta x + (1-\theta) y$ روی خطی است که x و y را به هم متصل می‌کند یعنی L. پس این پاره خط (اشتراک C و L) نیز محدب است.

II) اگر مجموعه C اشتراک آن با هر خطی محدب باشد آنگاه C محدب است.

$$\begin{cases} x, y \in C \\ x, y \in L \end{cases}$$

مقصود داریم اثبات کنیم که $\theta x + (1-\theta) y \in C$:

می‌دانیم که x و y در $C \cap L$ هستند. طبق فرض سوال $C \cap L$ محدب است. بنابراین طبق تعریف مجموعه محدب $\theta x + (1-\theta) y \in C \cap L$ و در نتیجه هم در L و هم در C است. $\leftarrow \theta x + (1-\theta) y \in C$ ✓

ادامه: پس می‌دانیم یک مجموعه محدب است اگر و تنها اگر اشتراک آن با یک خط دلخواه محدب باشد. یک خط دلخواه در نظر می‌گیریم $\{x + tv\}$

$$(x+tv)^T A (x+tv) + b^T (x+tv) + c = \alpha t^2 + \beta t + \gamma$$

ضرب می‌کنیم و مرتب سازی $\leftarrow$

$$\alpha = v^T A v, \quad \beta = b^T v + 2x^T A v, \quad \gamma = c + b^T x + x^T A x$$

اشتراک C با خط $\{x+tv\}$ برابر مجموعه زیر است

$$\{x + tv \mid \alpha t^2 + \beta t + \gamma \leq 0\}$$

برای اینکه این مجموعه محدب باشد باید داشته باشیم $\alpha \geq 0$ پس $v^T A v \geq 0$ که طبق تعریف ماتریس $\leftarrow$ Positive Semi definite

$$\boxed{A \succeq 0}$$

a) I ابتدا اثبات می‌کنیم اگر f شبه محدب باشد، آنگاه $f(\theta x+(1-\theta)y) \le \max\{f(x), f(y)\}$ باشد.

هر x و y در α sublevelset

تعریف می‌کنیم $\alpha = \max\{f(x), f(y)\}$

طبق تعریف توابع شبه محدب، x و y هر دو متعلق به α-sublevel set هستند. هم چنین تمام نقاط بین x و y در این مجموعه محدب نیز متعلق به α-sublevelset هستند. ← $\theta x+(1-\theta)y \in \alpha\text{-sublevelset}$

بنابراین، این نقاط هم زیر Threshold α قرار می‌گیرند

$$f(\theta x+(1-\theta)y) \le \alpha$$
$$f(\theta x+(1-\theta)y) \le \max\{f(x), f(y)\}$$

II حالا اثبات می‌کنیم که اگر $f(\theta x+(1-\theta)y) \le \max\{f(x), f(y)\}$ باشند، آنگاه f شبه محدب است.

هر x و y در α-sublevelset

$$f(x) \le \alpha$$
$$f(y) \le \alpha$$

فرض سوال: $f(\theta x+(1-\theta)y) \le \max\{f(x), f(y)\} \le \alpha$

پس: $\theta x+(1-\theta)y \in \alpha\text{ sublevel set}$

چون x و y را مشخص نکردیم و برای هر x و y ای این رابطه برقرار است، پس f شبه محدب است.

طبق I و II اثبات شد.

b) اگر g یک تابع Affine باشد باید اثبات کنیم اگر f محدب باشد fog نیز محدب است.

باید اثبات کنیم ← $f\circ g(\theta x+(1-\theta)y) \le \theta f\circ g(x) + (1-\theta) f\circ g(y)$ ← به ازای هر x, y

طبق Affine بودن g ⟸ $g(\theta x+(1-\theta)y) = \theta g(x) + (1-\theta) g(y)$

$$f\circ g(\theta x+(1-\theta)y) = f(g(\theta x+(1-\theta)y))$$

$$= f(\theta g(x) + (1-\theta) g(y))$$ ← طبق Aff بودن g

$$\le \theta f(g(x)) + (1-\theta) f(g(y))$$ ← طبق محدب بودن f

8 a)

دو نقطه فرضی $(\bar{x},\bar{y})$, (x,y)

باید اثبات کنیم

$$f(\theta(x,y)+(1-\theta)(\bar{x},\bar{y})) \geq \theta f((x,y)) + (1-\theta) f((\bar{x},\bar{y}))$$

جاگذاری تابع $\sqrt{xy}$

$$\sqrt{(\theta x + (1-\theta)\bar{x})(\theta y + (1-\theta)\bar{y})} \geq \theta\sqrt{xy} + (1-\theta)\sqrt{\bar{x}\bar{y}}$$

توان 2 می‌رسانیم

$$(\theta x + (1-\theta)\bar{x})(\theta y + (1-\theta)\bar{y}) \geq (\theta\sqrt{xy} + (1-\theta)\sqrt{\bar{x}\bar{y}})^2 \checkmark$$

که طبق نامساوی کوشی شوارتز درست است.

b) تغییر متغیر می‌دهیم ← $\begin{cases} u_i = e^{x_i} \\ v_i = e^{y_i} \end{cases}$ ←

$$f(\theta x + (1-\theta)y) = \log\left(\sum_{i=1}^{n} e^{\theta x_i + (1-\theta)y_i}\right) = \log\left(\sum_{i=1}^{n} u_i^{\theta} v_i^{(1-\theta)}\right)$$

با استفاده از نامساوی هولدر ←

$$\log\left(\sum_{i=1}^{n} u_i^{\theta} v_i^{(1-\theta)}\right) \leq \log\left(\left(\sum_{i=1}^{n} u_i^{\theta\frac{1}{\theta}}\right)^{\theta}\left(\sum_{i=1}^{n} v_i^{(1-\theta)\frac{1}{(1-\theta)}}\right)^{(1-\theta)}\right)$$

$$\underbrace{\log\left(\sum_{i=1}^{n} u_i^{\theta} v_i^{(1-\theta)}\right)}_{f(\theta x + (1-\theta)y)} \leq \underbrace{\theta\log\sum_{i=1}^{n} u_i}_{\theta f(x)} + \underbrace{(1-\theta)\log\sum_{i=1}^{n} v_i}_{(1-\theta)f(y)}$$

بنابراین یک تابع محدب است.

$\boxed{\text{سوال } 8}$

c)

می‌خواهیم اثبات کنیم :

$$f(\theta x + (1-\theta) y) \geq \theta f(x) + (1-\theta) f(y)$$

چون $-\sqrt{x}$ محدب است اگر $\sqrt{x}$ مقعر باشد. چون $\sqrt{x}$ ساده‌تر است از این سمت استفاده می‌کنیم.

$$\sqrt{\theta x + (1-\theta) y} \geq \theta \sqrt{x} + (1-\theta)\sqrt{y}$$

$$\text{توان } 2 \rightarrow \theta x + (1-\theta) y \geq \theta^2 x + (1-\theta)^2 y + 2\theta(1-\theta)\sqrt{xy}$$

$$\Updownarrow$$

$$\theta x - \theta^2 x + (1-\theta) y - (1-\theta)^2 y \geq 2\theta(1-\theta)\sqrt{xy}$$

$$\theta(1-\theta) x + (1-\theta) y(\cancel{1}-\cancel{1}+\theta) \geq 2\theta(1-\theta)\sqrt{xy}$$

$$x + y \geq 2\sqrt{xy}$$

$$x + y - 2\sqrt{xy} \geq 0$$

$$(\sqrt{x} - \sqrt{y})^2 \geq 0 \checkmark$$

اثبات شد که $\sqrt{x}$ مقعر و در نتیجه $-\sqrt{x}$ محدب است.

$\boxed{\text{سوال } 9}$

$$J(x) = \|Y - AX\|_2^2 = (Y-AX)^T(Y-AX) = Y^TY - Y^TAX - X^TA^TY + X^TA^TAX$$

$$= Y^TY - 2Y^TAX + X^TA^TAX$$

$$\underset{(1\times m)(m\times n)(n\times 1)}{X^TA^TY} = (Y^TAX)^T = \underset{(1\times n)(n\times m)(m\times 1)}{Y^TAX}$$

$$\frac{\partial J(x)}{\partial x} = \frac{\partial}{\partial x}(Y^TY) - 2\frac{\partial}{\partial x}(Y^TAX) + \frac{\partial}{\partial x} X^TA^TAX = 0$$

$$0 - 2A^TY + 2A^TAX = 0 \rightarrow \cancel{2}A^TAX = \cancel{2}A^TY$$

$$\boxed{\rightarrow X = (A^TA)^{-1}A^TY}$$

سوال ۱۰

$$J(x) = (Y-AX)^T(Y-AX) + \lambda x^T x = Y^T Y - 2Y^T A^T X + X^T A^T A X + \lambda x^T x$$

$$\frac{\partial}{\partial x} J(x) = \frac{\partial}{\partial x} Y^T Y - 2\frac{\partial}{\partial x} Y^T A^T X + \frac{\partial}{\partial x} X^T A^T A X + \frac{\partial}{\partial x} \lambda x^T x = 0$$

$$= 0 - 2A^T Y + 2A^T A X + 2\lambda X = 0$$

$$\Rightarrow \cancel{2}A^T A X + \cancel{2}\lambda X = \cancel{2}A^T Y \Rightarrow \cancel{2}(A^T A + \lambda I)X = A^T Y$$

$$\Rightarrow \boxed{X = (A^T A + \lambda I)^{-1} A^T Y}$$