“Optimization”

Shervin Halat

98131018

Homework #3

.1

a) برابر نرم ℓ_1

$$\min \|x\mathbf{1}-b\|_1 = \min \sum_{i=1}^{n} |x-b_i|$$

لذا حاصل برابر میانه b_i ها می‌باشد.

b) برابر نرم ℓ_2

$$\min \|x\mathbf{1}-b\|_2 = \min \sum_{i=1}^{n} |x-b_i|^2$$

$$= \min \underbrace{\sum_{i=1}^{n} (x-b_i)^2}_{f(x)} \Rightarrow \nabla f(x^*) = 0 \Rightarrow nx = \sum_{i=1}^{n} b_i \Rightarrow x^* = \frac{\sum b_i}{n}$$

لذا حاصل برابر میانگین b_i ها می‌باشد.

c)

$$\min \|x\mathbf{1}-b\|_\infty = \min \max |x-b_i|$$

لذا حاصل برابر $\frac{\max b_i - \min b_i}{2}$ می‌باشد.

.2

a) مسئله بدون قید روی x و r می‌باشد. ابتدا Lagrangian Dual Function را بدست می‌آوریم:

تابع Lagrange:

$$L(x,r,\nu)=\sum_{i=1}^{m}\phi(r_i)+\nu^T(Ax-b-r)$$

دوگان تابع Lagrange_Dual:

$$g(\nu)=\inf_{x,r} L(x,r,\nu)$$

از تابع $L(x,r,\nu)$ می‌دانیم که در صورتی روی x bounded below است که $A^T\nu=0$ لذا برای $g(\nu)$ داریم:

$$g(\nu)=\begin{cases}-b^T\nu+\sum_{i=1}^{m}\inf_{r_i}\left(\phi(r_i)-\nu_i r_i\right) & A^T\nu=0\\ -\infty & \text{otherwise}\end{cases}$$

با تبدیل inf به sup داریم:

$$\inf_{r_i}\left(\phi(r_i)-\nu_i r_i\right)=-\sup_{r_i}\left(\nu_i r_i-\phi(r_i)\right)=-\phi^*(\nu_i)$$

$$\Longrightarrow\quad \boxed{\begin{aligned}&\text{Dual Problem: } \max\ -b^T\nu-\sum_{i=1}^{m}\phi^*(\nu_i)\\ &\text{s.t. } A^T\nu=0\end{aligned}}$$

b) تابع Deadzone_linear به فرم زیر است:

$$\phi(x)=\begin{cases}0 & |x|<1\\ |x|-1 & |x|\geq 1\end{cases}\ \Rightarrow\ \phi^*(z)=\begin{cases}|z| & |z|\leq 1\\ \infty & |z|>1\end{cases}$$

: Conjugate of Deadzone_linear

پس مسئله Dual به فرم روبرو تغییر می‌کند:

(g(ν) به شرطی bounded above است که $\|\nu\|_\infty\leq 1$ باشد)

$$\boxed{\begin{aligned}&\max\ -b^T\nu-\|\nu\|_1\\ &\text{s.t. } A^T\nu=0,\ \|\nu\|_\infty\leq 1\end{aligned}}$$

.3

برای یک متغیر تصادفی نرمال تابع CDF متناظر آن به قرار زیر می‌باشد:

$$\Phi_X(x) = P(X \le x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{\frac{-u^2}{2}\right\} du$$

$$\Longrightarrow P(y=1) = P(v \le -(a^T u + b)) = \Phi(-a^T u - b)$$

$$P(y=0) = 1 - P(v \le -(a^T u + b)) = 1 - \Phi(-a^T u - b)$$

likelihood function:

$$l(a,b) = \prod_{y_i=1} \Phi(-a^T u_i - b) \prod_{y_i=0} (1 - \Phi(-a^T u_i - b))$$

log-likelihood function:

$$l(a,b) = \sum_{y_i=1} \log \Phi(-a^T u_i - b) + \sum_{y_i=0} \log(1 - \Phi(-a^T u_i - b))$$

$\Leftarrow$ مسئله بهینه‌سازی محدب به یکی از دو صورت مقابل است:

$$\min \; -l(a,b) \quad \equiv \quad \max \; l(a,b)$$

.4

Likelihood function:

$$L(a, S) = \prod_{i=1}^{N} P(y_i)$$

log-likelihood function:

$$\mathcal{L}(a, S) = \sum_{i=1}^{N} \log P(y_i) = \sum_{i=1}^{N} \left[-\log\left((2\pi)^{n/2} (\det S)^{1/2}\right) - \frac{1}{2} (y_i - a)^T S^{-1} (y_i - a) \right]$$

$$= -\frac{Nn}{2} \log(2\pi) - \frac{N}{2} \log(\det S) - \frac{1}{2} \sum_{i=1}^{N} (y_i - a)^T S^{-1} (y_i - a)$$

$$= \frac{N}{2} \left[-n \log(2\pi) - \log \det S - tr(S^{-1} C) - (a - \mu)^T S^{-1} (a - \mu) \right]$$

به طوری برای μ (میانگین نمونه‌ها) و C (کوواریانس نمونه‌ها) داریم:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} y_i \quad , \quad C = \frac{1}{N} \sum_{i=1}^{N} (y_i - \mu)(y_i - \mu)^T$$

معکوس ماتریس کوواریانس شرطی (R_{ij}^{-1}) به کمک ماتریس $S_{2\times2}^{-1}$ بدست می‌آیند به طوری که i و j به ترتیب سطر و ستون می‌باشند.

همچنین اگر $(S^{-1})_{ij} = 0 \Leftarrow$ دو متغیر i و j مستقل شرطی هستند.

$\Longrightarrow$ مسئله ML به صورت مقابل در می‌آید:

$$\max \ \mathcal{L}(S, a)$$

$$\text{s.t.} \ (S^{-1})_{i,j} = 0 \ , (i, j) \in \mathcal{N}$$

.5

* پاسخ های گرافیکی و کد مربوط به این سؤال در کنار حل بخش پیاده سازی قرار داده شده است! *

a) با توجه به وجود تنها یک نقطه feasible ، یعنی $\boxed{x^*=(1,0)}$

مقدار p^* متناظر برابر است با:

$$p^*(x_1^{*2}+x_2^{*2}) = 1+0 = \boxed{1 = p^*}$$

b)

$$L(x,\lambda) = x_1^2+x_2^2+\lambda_1\left[(x_1-1)^2+(x_2-1)^2-1\right]+\lambda_2\left[(x_1-1)^2+(x_2+1)^2-1\right]$$

شروط KKT مسئله به صورت زیر می‌باشند:

1) $f_1(x)\leq 0 \,,\; f_2(x)\leq 0$ 2) $\lambda_1\geq 0 \,,\; \lambda_2\geq 0$

3) $\lambda_1 f_1(x)=0 \,,\; \lambda_2 f_2(x)=0$ 4) $\begin{cases} 2x_1+2\lambda_1(x_1-1)+2\lambda_2(x_1-1)=0 \\ 2x_2+2\lambda_1(x_2-1)+2\lambda_2(x_2+1)=0 \end{cases}$

حال با جایگذاری x^* در شروط بالا، (شروط 2 و 4) روابط زیر را برای λ_1^* و λ_2^* داریم:

2) $\lambda_1^*, \lambda_2^* \geq 0$ 4) $\begin{cases} -\lambda_1^*+\lambda_2^*=0 \\ 2=0 \; \times \end{cases}$

با توجه به اینکه به دستگاهی رسیدیم که جواب ندارد پس چنین ضرایب لاگرانژین λ_1^* و λ_2^* وجود ندارند

c) برای تابع لاگرانژین داریم:

$$L(x_1,x_2,\lambda_1,\lambda_2) = (1+\lambda_1+\lambda_2)x_1^2+(1+\lambda_1+\lambda_2)x_2^2-2(\lambda_1+\lambda_2)x_1-2(\lambda_1-\lambda_2)x_2+\lambda_1+\lambda_2$$

رابطه بالا درصورتی محدب است که ضرایب x_1^2 و x_2^2 مثبت باشند یعنی:

$$1+\lambda_1+\lambda_2\geq 0$$

که دراینصورت $L(x,\lambda)$ مینیمم دارد.

حال به کمک گرادیان و با فرض $1+\lambda_1+\lambda_2>0$ نقاط مینیمم به کمک گرادیان صفر برابر است با:

$$x_1=\frac{\lambda_1+\lambda_2}{1+\lambda_1+\lambda_2}\quad,\quad x_2=\frac{\lambda_1-\lambda_2}{1+\lambda_1+\lambda_2}$$

$$\Longrightarrow \inf_x L(x,\lambda)=g(\lambda_1,\lambda_2)=\begin{cases}\dfrac{-(\lambda_1+\lambda_2)^2+(\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2}+\lambda_1+\lambda_2 & 1+\lambda_1+\lambda_2\geq 0\\ -\infty & \text{O/W}\end{cases}$$

لذا Dual Problem متناظر به قرار زیر است:

$$\max \quad \frac{-(\lambda_1+\lambda_2)^2+(\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2}+\lambda_1+\lambda_2$$

$$\text{s.t.}\quad \lambda_1\geq 0$$

$$\lambda_2\geq 0$$

$$1+\lambda_1+\lambda_2\geq 0 \quad (\text{بدیهی})$$

.6 ابتدا به تساوی روبرو $\|Ax-b\|_2^2=(Ax-b)^T(Ax-b)=x^TA^TAx+\dots$

می دانیم که مسئله ی Primal یک مسئله ی Convex بوده و چون قید نامساوی نداریم پس Strong Duality برقرار است لذا با شروط KKT می توان قسمت های B و C را حل نمود یعنی Primal & Dual solutions .

(a

$$L(x,\nu)=\|Ax-b\|_2^2+\nu^T(Gx-h)$$

تابع لاگرانژین :

$$=\boxed{x^TA^TAx+(G^T\nu-2A^Tb)^Tx-\nu^Th}=L(x,\nu)$$

حال برای حل b و c به سراغ شروط KKT می رویم لذا با توجه به Optimality Conditions برای Equality Constrained داریم :

$$\underbrace{Gx^*=h}_{\text{I}}\quad,\quad \nabla_x L(x^*,\nu^*)=0\implies \underbrace{2A^T(Ax^*-b)+G^T\nu^*=0}_{\text{II}}$$

(b , c

از رابطه (II) : $x^*=(A^TA)^{-1}(A^Tb-(\tfrac{1}{2})G^T\nu^*)$ (III)

جایگذاری x^* در (I) :

$$\xrightarrow{\text{II},\,\text{I}}\quad G(A^TA)^{-1}A^Tb-(\tfrac{1}{2})G(A^TA)^{-1}G^T\nu^*=h$$

$$\longrightarrow\ \boxed{\nu^*=-2(G(A^TA)^{-1}G^T)^{-1}(h-G(A^TA)^{-1}A^Tb)}\ \text{(IV)}$$

حال با جایگذاری (IV) در (III) x^* نیز به طور دقیق مشخص می گردد .

.8

(a

Lagrange function:

$$L(x,\lambda,\nu) = c^T x + \lambda^T (Gx - h) + \nu^T (Ax - b)$$

$$= (c^T + \lambda^T G + \nu^T A)x - h\lambda^T - \nu^T b$$

حال برای یافتن Lagrange Dual Function می‌رویم با توجه به اینکه تابع L یک تابع Affine و Unbounded می‌باشد.

$\Rightarrow$ Dual function: $g(\lambda,\nu) = \inf_x L(x,\lambda,\nu)$

تابع دوگان:

$$\Rightarrow g(\lambda,\nu) = \begin{cases} -\lambda^T h - \nu^T b & c + G^T\lambda + A^T\nu = 0 \\ -\infty & \text{O.W.} \end{cases}$$

(b با توجه به $g(\lambda,\nu)$ بدست آمده، مسئله دوگان به قرار زیر است:

Implicit Dual Problem: $\max\ g(\lambda,\nu)$ s.t. $\lambda \geq 0$ $\Longrightarrow$ Explicit Dual Problem: $\max\ -\lambda^T h - \nu^T b$ s.t. $\lambda \geq 0$, $c + G^T\lambda + A^T\nu = 0$

9.

مطابق توضیحات سؤال، مسئله اولیه به فرم زیر می‌باشد:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

برای گام نیوتن مسئله صورت سؤال داریم:

$$\begin{bmatrix} H + A^T Q A & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \omega \end{bmatrix} = \begin{bmatrix} -g - 2A^T Q A x + 2A^T Q b \\ 0 \end{bmatrix}$$

لذا داریم: $A \Delta x = 0$ ، پس:

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \omega \end{bmatrix} = \begin{bmatrix} -g - 2A^T Q A x + 2A^T Q b \\ 0 \end{bmatrix} \quad \text{(I)}$$

$$\Longrightarrow \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \tilde{\omega} \end{bmatrix} = \begin{bmatrix} -g \\ 0 \end{bmatrix} \quad \text{(II)}$$

از مقایسه (I) و (II) داریم:

$$\underline{\tilde{\omega} = \omega + 2QAx - 2Qb}$$

لذا می‌توان نتیجه گرفت که گام نیوتن برای هر 2 مسئله مورد نظر یکسان می‌باشد.

۱۵.

از شروط KKT داریم:

$$t \nabla f_0(x^*(t)) + \nabla \phi(x^*(t)) + A^T \nu = 0 \quad (\mathrm{I})$$

همچنین می‌دانیم:

$$\frac{d f_0(x^*(t))}{d(t)} = \nabla f_0(x^*(t))^T \frac{d x^*(t)}{dt} \quad (\mathrm{II})$$

حال با مشتق گیری از (I) نسبت به t داریم:

$$\nabla f_0(x^*(t)) + \left(t \nabla^2 f_0(x^*(t)) + \nabla^2 \phi(x^*(t))\right)\frac{dx^*}{dt} = 0$$

$$\Longrightarrow \frac{dx^*}{dt} = -\left(t \nabla^2 f_0(x^*(t)) + \nabla^2 \phi(x^*(t))\right)^{-1} \nabla f_0(x^*(t)) \quad (\mathrm{III})$$

با جایگذاری (III) در (II) داریم:

$$\frac{d f_0(x^*(t))}{dt} = -\nabla f_0(x^*(t))^T \left(t \nabla^2 f_0(x^*(t)) + \nabla^2 \phi(x^*(t))\right)^{-1} \nabla f_0(x^*(t)) \quad (\mathrm{IV})$$

رابطه (IV) به وضوح منفی است. لذا داریم:

$$\boxed{\frac{d f_0(x^*(t))}{dt} < 0}$$

لذا با افزایش t، $f_0(x^*(t))$ کاهش می‌یابد.