دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

تمرین سری سه

درس بهینه‌سازی

فرهاد دلیرانی

۹۶۱۳۱۱۲۵

dalirani@aut.ac.ir

dalirani.1373@gmail.com

فهرست

ابزارهای استفاده شده

زبان برنامه نویسی: -

محیط توسعه: -

سیستم عامل: -

1) براساس الگوریتم گرادیان مزدوج ، برای دو مسئلهٔ

$$\begin{cases} Ax = b & ① \\ Ay = r_0 & ② \end{cases}$$

در تمام گام ها $r_{①k} = r_{②k}$ است همین طور $x_k = y_k + x_0$

در قدم 0

$$r_{①0} = Ax_0 - b$$

$$r_{②0} = Ay_0 - r_0 = 0 - r_0 = -(b - Ax_0) = Ax_0 - b$$

$$p_{①0} = -r_{①0} = b - Ax_0$$

$$p_{②0} = -r_{②0} = b - Ax_0$$

$$\left.\begin{array}{l} r_{①0} = r_{②0} \\ p_{①0} = p_{②0} \end{array}\right| \rightarrow \alpha_{①0} = \alpha_{②0}$$

$$\left.\begin{array}{l} x_{1} = x_0 + \alpha_{①0} p_{①0} \\ y_{1} = y_0 + \alpha_{②0} p_{②0} \end{array}\right| \rightarrow y_1 - x_1 = -x_0 \rightarrow x_1 = y_1 + x_0$$

$$\left.\begin{array}{l} r_{①1} = Ax_1 - b \\ r_{②1} = Ay_1 - r_0 \end{array}\right| \rightarrow r_{①1} - r_{②1} = Ax_1 - b - Ay_1 + r_0$$

$$= Ax_0 + r_0 - b$$

$$= Ax_0 + b - Ax_0 - b = 0$$

$$\rightarrow \left.\begin{array}{l} \boxed{r_{①1} = r_{②1}} \\ p_{①0} = p_{②0} \end{array}\right| \rightarrow \beta_{①1} = \beta_{②1} \rightarrow \boxed{p_{①1} = p_{②1}}$$

در پایان قدم ۱ مشاهده کردیم که r_1 و P_1 برای هر دو معادله یکسان است. در ابتدا $x_0 = y_0 + x_0$ برقرار بود بعد از پایان قدم ۱ مشاهده شد که $x_1 = y_1 + x_0$ برقرار است.

اگر قدم های بعدی را مانند آنچه که در قدم ۱ انجام دادیم دوباره انجام دهیم به نتیجه گیری یکسانی می رسیم و مشاهده می کنیم که در هر قدم r_k و P_k برای هر دو معادله یکسان می شود و رابطه ی زیر بین x_k و y_k برقرار است:

$$x_k = y_k + x_0$$

در نتیجه در روش گرادیان مزدوج ۱ برای سیستم $Ax=b$ شروع از نقطه ی x_0 معادل شروع از نقطه ی $y_0=0$ برای سیستم $Ay=r_0$ است. که در آن $r_0 = b - Ax_0$.

۲) برای این کار عبارت $\frac{1}{N}\sum_{i=1}^{N}\left(d_i - d(x_i,y_i)\right)^2$ را کمینه می‌کنیم. هرچه MSE کمتر شود نشان دهنده‌ی این است که به تخمین بهتری رسیده‌ایم.

در نتیجه برای پیدا کردن P مسئله‌ی زیر را حل می‌کنیم

$$\text{minimize} \quad \frac{1}{N}\sum_{i=1}^{N}\left(d_i - d(x_i,y_i)\right)^2$$

P درون این تابع است.

این یک مسئله‌ی convex است، زیرا

$$\frac{1}{N}\sum_{i=1}^{N}\left(d_i - d(x_i,y_i)\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left[d_i^2 - 2d_i\, d(x_i,y_i) + d^2(x_i,y_i)\right]$$

$$= \underbrace{\frac{1}{N}\sum_{i=1}^{N}}_{\text{مجموع جمع چند تابع محدب با ضرایب مثبت تابعی محدب است.}}\left[\underbrace{d_i^2}_{\text{constant}} \underbrace{- 2d_i\left((x_i-y_i)^T P (x_i-y_i)\right)^{\frac{1}{2}}}_{\text{ترکیب یک تابع خطی برحسب P با یک تابع مقعر (جذر) یک تابع مقعر است که ضرب یک عدد منفی در آن موجب محدب شدن آن می‌شود.}} + \underbrace{(x_i-y_i)^T P (x_i-y_i)}_{\text{یک تابع خطی برحسب P است.}}\right]$$

در نتیجه کل عبارت convex است.

(۳

y is a random variable. $y \in \{0,1\}$

$$y = \begin{cases} 1 & a^T u + b + v \leq 0 \\ 0 & a^T u + b + v > 0 \end{cases}$$

u: $u \in R^n$ is a vector of explanatory

v: v is a zero mean unit variance gaussian variable. $N(0,1)$

Integral in Normal Gaussian Distribution:

$$F(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

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

$$= \int_{-\infty}^{-a^T u - b} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \xlongequal{\text{even Function}} \int_{a^T u + b}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2}{2}} dx = F(a^T u + b)$$

$$P(y=0) = P(a^T u + b + v > 0) = P(-a^T u - b < v) =$$

$$= \int_{-a^T u - b}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2}{2}} dx = F(-a^T u - b)$$

Log-Likelihood:

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

ادامه ۳)

تابع $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ یک

تابع log-concave است. یعنی در رابطه

زیر صدق می‌کند

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

$$e^{-\frac{(\theta x+(1-\theta)y)^2}{2}} \geq \left(e^{-\frac{x^2}{2}}\right)^{\theta}\left(e^{-\frac{y^2}{2}}\right)^{1-\theta} \longleftrightarrow$$

$$e^{-\frac{(\theta x+(1-\theta)y)^2}{2}} \geq \left(e^{-\frac{x^2}{2}}\right)^{\theta}\left(e^{-\frac{y^2}{2}}\right)^{(1-\theta)} \longleftrightarrow$$

$$-(\theta x+(1-\theta)y)^2 \geq -\theta x^2-(1-\theta)y^2 \longleftrightarrow$$

$$(\theta x+(1-\theta)y)^2 \leq \theta x^2+(1-\theta)y^2 \longleftrightarrow$$

$$\theta^2x^2+(1-\theta)^2y^2+2\theta(1-\theta)xy-\theta x^2-(1-\theta)y^2 \leq 0 \longleftrightarrow$$

$$(\theta^2-\theta)x^2+((1-\theta)^2-(1-\theta))y^2+2\theta(1-\theta)xy \leq 0 \longleftrightarrow$$

$$\theta(\theta-1)x^2+(1-\theta)((1-\theta)-1)y^2+2\theta(1-\theta)xy \leq 0 \longleftrightarrow$$

$$+\theta(\theta-1)x^2+(1-\theta)(-\theta)y^2+(2\theta)(1-\theta)xy \leq 0 \longleftrightarrow$$

$$(-\theta)x^2+(-\theta)y^2+(2\theta)xy \leq 0 \longleftrightarrow$$

$$-x^2-y^2+2xy \leq 0 \longleftrightarrow -(x-y)^2 \leq 0 \quad \checkmark$$

نشان داده شد که $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ یک تابع log-concave است. توابع log-concave تحت marginalization خاصیت خود را حفظ می‌کنند. در نتیجه $F(x)=\int^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$ یک تابع log-concave است. اگر تابع تابع log-concave با log ترکیب شوند تابع حاصل concave خواهند بود در نتیجه

$\log F(x)$ یک تابع concave است

ادامه‌ی ۳)

در نتیجه ترکیب affine تابع $\log F(a)$ هم یک تابع

concave است. پس $\log F(a^T u_i + b)$ و $\log F(-a^T u_i - b)$

هر دو concave هستند و در نهایت نتیجه می‌شود که تابع

$L(a,b)$ یک تابع concave است.

در نتیجه داریم

maximum $L(a,b)$

یا

minium $-L(a,b)$

(4

الف)

$$\text{minimize} \quad \sum_{i=1}^{m} \phi(r_i)$$

$$\text{S.T} \quad r = Ax - b$$

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

$$g(\nu) = \inf_{r} \left(\sum_{i=1}^{m} \phi(r_i) + \cancel{\nu^T Ax} - \nu^T b - \nu^T r \right)$$

برای جلوگیری از $-\infty$ شدن $g(\nu)$ باید ضریب x برابر با صفر باشد.

$$= \inf_{r} \left(\sum_{i=1}^{m} \phi(r_i) - \nu^T r \right) - \nu^T b$$

روز مباهله پیامبر اسلام صلی الله علیه و آله (۱۰ هـ ق)

۱۷ | ۲۵ ذی‌الحجه ۱۴۳۶ | 9 October 2015 | جمعه | مهر

$$= -\sup_{r} \left(\nu^T r - \sum_{i=1}^{m} \phi(r_i) \right) - \nu^T b$$

$$g(\nu) = -f_0^*(\nu) - \nu^T b$$

$$g(\nu) = \begin{cases} -f_0^*(\nu) - \nu^T b & \nu^T A = 0 \\ -\infty & \text{Else} \end{cases}$$

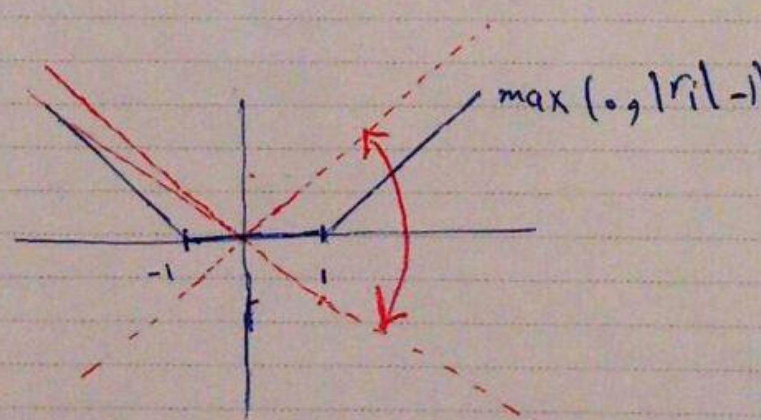
$$\left.\begin{array}{ll} \text{maximize} & g(\nu) = -f_0^*(\nu) - \nu^T b \\ \text{subject to:} & \nu^T A = 0 \end{array}\right\}$$

$$f_0(r) = \sum_{i=1}^{m} \phi(r_0) = \sum_{i=1}^{m} \max(0, |r_i|-1) \quad (ب$$

$$f_0^*(\nu) = \sup_r \left(\nu^T r - \sum_{i=1}^{m} \phi(r_i) \right)$$

$$= \sup_r \left(\sum_{i=1}^{m} \left[\nu_i r_i - \phi(r_i) \right] \right) =$$

$$= \sup_r \left(\sum_{i=1}^{m} \left[\nu_i r_i - \max(0, |r_i|-1) \right] \right)$$



$$f^*(\nu) = \begin{cases} +\infty & \|\nu\|_\infty > 1 \\ \|\nu\|_1 & \|\nu\|_\infty \leq 1 \end{cases}$$

$$\rightarrow g(\nu) = \begin{cases} -\|\nu\|_1 - \nu^T b & \|\nu\|_\infty \leq 1 \text{ and } \nu^T A = \emptyset \\ -\infty & \text{Else} \end{cases}$$

$$\begin{aligned} \text{maximum} \quad & g(\nu) = -\|\nu\|_1 - \nu^T b \\ \text{Subject to} \quad & \|\nu\|_\infty \leq 1 \\ & \nu^T A = \emptyset \end{aligned}$$

(5

$$\min \quad 0.5\|y-y_0\|_2 + \sum_{i=1}^{m} \|x_i\|_2$$
$$\text{St} \quad x_i = C_i y + d_i$$

$$L(x,y,\nu) = 0.5\|y-y_0\|_2 + \sum_{i=1}^{m}\|x_i\|_2 + \sum_{i=1}^{m}\nu_i^T(x_i - C_i y - d_i)$$

$$\rightarrow L(x,y,\nu) = 0.5\|y-y_0\|_2 + \sum_{i=1}^{m}\nu_i^T(-C_i y) + \sum_{i=1}^{m}\|x_i\|_2 + \sum_{i=1}^{m}\nu_i^T x_i + \sum_{i=1}^{m}\nu_i^T(-d_i)$$

$$\rightarrow L(x,y,\nu) = -\sum_{i=1}^{m}\nu_i^T C_i y + 0.5\|y-y_0\|_2 + \sum_{i=1}^{m}\left(\|x_i\|_2 + \nu_i^T x_i\right) - \sum_{i=1}^{m}\nu_i^T d_i$$

$$\rightarrow L(x,y,\nu) = -\left[\sum_{i=1}^{m}\nu_i^T C_i y - 0.5\|y-y_0\|_2\right] + \sum_{i=1}^{m}\left[\|x_i\|_2 + \nu_i^T x_i\right] - \sum_{i=1}^{m}\nu_i^T d_i$$

$$\rightarrow L(x,y,\nu) = -\left[\sum_{i=1}^{m}\nu_i^T C_i y - 0.5\|y-y_0\|_2\right] - \sum_{i=1}^{m}\left[-\nu_i^T x_i - \|x_i\|_2\right] - \sum_{i=1}^{m}\nu_i^T d_i$$

$$g(\nu) = \inf_y -\left[\sum_{i=1}^{m}\nu_i^T C_i y - 0.5\|y-y_0\|_2\right] + \inf_x -\sum_{i=1}^{m}\left[-\nu_i^T x_i - \|x_i\|_2\right] - \sum_{i=1}^{m}\nu_i^T d_i$$

$$g(\nu) = \sup_y \left[\sum_{i=1}^{m}\left(\nu_i^T C_i y - \frac{0.5}{m}\|y-y_0\|_2\right)\right] + \sup_x \sum_{i=1}^{m}\left[-\nu_i^T x_i - \|x_i\|_2\right] - \sum_{i=1}^{m}\nu_i^T d_i$$

$$\boxed{f_1(y) = \frac{0.5}{m}\|y-y_0\|_2} \qquad \boxed{f_2(x_i) = \|x_i\|}$$

$$\Rightarrow g(\nu) = \sum_{i=1}^{m} f_1^*\left((\nu_i^T C_i)^T\right) + \sum_{i=1}^{m} f_2^*(-\nu_i) - \sum_{i=1}^{m}\nu_i^T d_i$$

$$\rightarrow \boxed{g(\nu) = \sum_{i=1}^{m} f_1^*(C_i^T\nu_i) + \sum_{i=1}^{m} f_2^*(-\nu_i) - \sum_{i=1}^{m}\nu_i^T d_i}$$

$$g(\nu) = \begin{cases} -\sum_{i=1}^{m}\left(\nu_i^T d_i - y_0^T C_i^T \nu_i\right) & \|C_i^T \nu\|_2 \le 1 \text{ and } \|-\nu_i\| \le 1 \\ \infty & \text{Else} \end{cases}$$

minimize $f_0(x)$ (6

s.t. $Ax=b$

$$L(x,\nu) = f_0(x) + \nu^T(Ax-b)$$

$$g(\nu) = \inf_x L(x,\nu) \longrightarrow$$

$$g(\nu) = \inf_x \left(f_0(x) + \nu^T(Ax-b)\right)$$

$$p(x) = f(x) + \alpha\|Ax-b\|_2^2 = f(x) + \alpha(Ax-b)^T(Ax-b)$$

$$\rightarrow p(x) = f(x) + \alpha(x^TA^T - b^T)(Ax-b)$$

$$= f(x) + \alpha(x^TA^TAx - x^TA^Tb - b^TAx + b^Tb)$$

$$\rightarrow 0 = \nabla f_0(x^*) + \alpha(2A^TAx^* - 2A^Tb)$$

$$\rightarrow 0 = \nabla f_0(x^*) + 2\alpha A^T(Ax^*-b)$$

درصورتی که ν را برابر با $2\alpha(Ax^*-b)$ در نظر بگیریم آنگاه x^* تابع زیر را کمینه می کند

$$f_0(x) + \nu^T(Ax-b)$$

که ν یک dual visible مسئله‌ی داده شده است

$$g(\nu) = \inf_x \left(f_0(x) + \nu^T(Ax-b)\right) = f_0(x^*) + 2\alpha\|Ax^*-b\|_2^2$$

بنابراین حد پایین برابر می‌شود با

$$f_0(x^*) + 2\alpha\|Ax^*-b\|_2^2 \leq f_0(x)$$

when $Ax=b$