۹۸۱۰۳۸۹۷ محمدابوالفضل زادیان

Subject:
Date

به نام خدا

تمرین دوم، بخش دوم

۱) مرحله اول

(a) $Value(S_1) = -h(S_1) = -1$

(b) $Value(S_2) = -h(S_2) = -2$

$\Longrightarrow \Delta E_{12} = Value(S_2) - Value(S_1) = -2+1 = -1$

$\longleftarrow$ چون $\Delta E \leq 0$ پس با احتمال $e^{\Delta E / T}$ به S_2 می‌رویم.

$$e^{\frac{\Delta E}{T}} = e^{\frac{\Delta E_{12}}{schedule(1)}} = e^{\frac{-1}{2}} = ۰٫۶$$

$\Longleftarrow$ طبق صورت سوال چون acceptance Prob از ۰٫۵ بیشتر است، به S_2 می‌رویم.

مرحله دوم)

(a) $Value(S_2) = -h(S_2) = -2$

(b) $Value(S_3) = -h(S_3) = -2$

$\Longrightarrow \Delta E_{23} = Value(S_3) - Value(S_2) = -2+2 = 0$

$\longleftarrow$ چون $\Delta E \leq 0$، پس با احتمال $e^{\Delta E / T}$ به S_3 می‌رویم.

$$e^{\frac{\Delta E}{T}} = e^{\frac{0}{1}} = e^{0} = 1 \Longrightarrow$$ به S_3 می‌رویم

مرحله سوم) چون T در این مرحله ۰ می‌شود، جستجو پایان می‌یابد.

نتیجه نهایی: $S_1 \xrightarrow{T=2,\ e^{\frac{\Delta E}{T}}=0.6} S_2 \xrightarrow{T=1,\ e^{\frac{\Delta E}{T}}=1} S_3$

Subject:
Date

۲) آ) می‌دانیم روش gradient descent برای توابع Convex به صورت قطعی به مینیمم (در حالت کلی Optimum) سراسری همگرا می‌شوند.

← پس اگر ثابت کنیم تابع f، یک تابع convex است خواهیم داشت که با استفاده از gradient descent و با مقدار مناسب α می‌توان به مقدار مینیمم سراسری همگرا شد.

برای اثبات همگرا بودن f، با توجه به ماتریس Hessian آن، اگر مقادیر

$$\frac{\partial^2 f}{\partial x_1^2}, \quad \frac{\partial^2 f}{\partial x_2^2}, \quad \frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 \geq 0$$

باشند، آنگاه f، convex است. داریم:

$$f(x_1, x_2) = x_1^4 + 2x_2^2 + x_1$$

1) $\frac{\partial^2 f}{\partial x_1^2} = 12x_1^2 \geq 0$

2) $\frac{\partial^2 f}{\partial x_2^2} = 4 \geq 0$

3) $\frac{\partial^2 f}{\partial x_1^2} \times \frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 = 48x_1^2 - 0 \geq 0$

⇐ تابع f، Convex است و جواب (آ)، بله است.

ب) ابتدا ∇f را محاسبه می‌کنیم:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1^3 + 1 \\ 4x_2 \end{bmatrix}$$

۹۸۱۰۳۸۹۷ محمد ابول نژاد اردبیل

Subject:
Date

ادامه سوال ۲) ب) - مرحله ۱)

$x_0 = (-1, 0)$

$$x \leftarrow \begin{bmatrix} -1 \\ 0 \end{bmatrix} - 10^{-4} \times \begin{bmatrix} 4\times -1+1 \\ 4\times 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} - \begin{bmatrix} -3\times 10^{-4} \\ 0 \end{bmatrix}$$

$$\Rightarrow x \leftarrow \begin{bmatrix} -0.9997 \\ 0 \end{bmatrix}$$

- مرحله ۲)

$$x \leftarrow \begin{bmatrix} -0.9997 \\ 0 \end{bmatrix} - 10^{-4} \times \begin{bmatrix} 4\times(-0.9997)+1 \\ 4\times 0 \end{bmatrix}$$

$$\Rightarrow x \leftarrow \begin{bmatrix} -0.99940012 \\ 0 \end{bmatrix}$$

- مرحله ۳)

$$x \leftarrow \begin{bmatrix} -0.9994 \\ 0 \end{bmatrix} - 10^{-4} \begin{bmatrix} 4\times(-0.9994)+1 \\ 4\times 0 \end{bmatrix}$$

$$\Rightarrow x \leftarrow \begin{bmatrix} -0.9991 \\ 0 \end{bmatrix}$$

ج) - مرحله ۱)

$$x \leftarrow \begin{bmatrix} -1 \\ 0 \end{bmatrix} - 1\times \begin{bmatrix} 4\times -1+1 \\ 4\times 0 \end{bmatrix} = \begin{bmatrix} -1+3 \\ 0-0 \end{bmatrix}$$

$$\Rightarrow x \leftarrow \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

- مرحله ۲)

$$x \leftarrow \begin{bmatrix} 2 \\ 0 \end{bmatrix} - 1\times \begin{bmatrix} 4\times 2+1 \\ 4\times 0 \end{bmatrix} = \begin{bmatrix} 2-9 \\ 0-0 \end{bmatrix}$$

$$\Rightarrow x \leftarrow \begin{bmatrix} -7 \\ 0 \end{bmatrix}$$

Subject:
Date

ادامه سوال ۲) ج) — مرحله ۳)

$$x \leftarrow \begin{bmatrix} -7 \\ 0 \end{bmatrix} - 1 \times \begin{bmatrix} 4 \times -7 + 1 \\ 4 \times 0 \end{bmatrix}$$

$$\Rightarrow x \leftarrow \begin{bmatrix} -7+27 \\ 0 \end{bmatrix} = \begin{bmatrix} 20 \\ 0 \end{bmatrix}$$

د) چون تابع convex است، اگر بخواهیم با روش least squared، مینیمم واقعی تابع f را بدست آوریم تا با جواب های "ب" و "ج" مقایسه کنیم، خواهیم داشت:

$$global\text{-}min\,(f(x_1, x_2)) \cong -0.47 \quad (x_1 = -0.62\,,\ x_2 = 0)$$

← در قسمت "ب"، جواب ها به کندی کم می‌شدند و این به دلیل step rate کوچک بود. پس α بهینه باید از این مقدار بیشتر باشد.

← در قسمت "ج" نیز، جواب ها به سرعت تغییر کردند که این نشانه‌ای از step_rate بزرگ بود، پس α بهینه باید از مقدار کوچکتر باشد.

← اگر $\alpha = 0.05$ قرار دهیم، مقادیر اولیه طی سه مرحله به شکل زیر تغییر می‌کند:

$$x_1: \ -1 \xrightarrow{\text{مرحله ۱}} -0.85 \xrightarrow{\text{مرحله ۲}} -0.73 \xrightarrow{\text{مرحله ۳}} \boxed{-0.634}$$

$$x_2: \ 0 \xrightarrow{\text{مرحله ۱}} 0 \xrightarrow{\text{مرحله ۲}} 0 \xrightarrow{\text{مرحله ۳}} \boxed{0}$$

پس با α های نسبتا کوچک (طبق قضیه دال الاید) مقدار f به صورت مناسبی به سمت optimum کاهش پیدا می‌کند که α پیشنهادی برابر 0.05 بود.

PAPCO