تمرین شماره سوم درس یادگیری ماشین

پریا پاسه‌ورز

شماره دانشجویی: 810101393

سوال اول)

الف) فرم کلی بهینه‌سازی الگوریتم‌های line search را بیان کنید.

در مسائل line search، هدف این است که مقدار بهینه α > 0 را پیدا کنیم که تابع f را در جهت $p^{(t)}$ مینیمم کند. فرم کلی آن به شکل زیر است:

$$\min_{\alpha>0} f\left(x^t + \alpha p^t\right)$$

x^t: نقطه شروع است.

p^t : جهت جستجو است.

Step size: α

ب) در الگوریتم line search تعریف جهت نزول چیست؟

تعریف جهت نزول، جهتی است که در آن مقدار تابع f، کاهشی می‌شود، به عبارت دیگر:

$$\nabla f(x^t)^T p^t < 0$$

پ) برای به دست آوردن step size در الگوریتم‌های line search از چه روش‌هایی استفاده می‌شود؟

1. Golden Section search

این روش برای زمانی کاربرد دارد که می‌دانیم تابع ما در یک بازه خاص، فقط یک مینینمم دارد. مزیتش این است که از مشتق تابع استفاده نمی‌کند و برای زمان‌هایی که به مشتق دسترسی نداریم، مناسب است.

1. ابتدا بازه‌ای که می‌خواهیم مینیمم را در آن پیدا کنیم تعریف می‌کنیم: $[\alpha_{\text{low}}, \alpha_{\text{high}}]$
2. دو نقطه جدید زیر از این روابط محاسبه کنید:

$$\alpha_1 = \alpha_{\text{low}} + (1-\phi)(\alpha_{\text{high}} - \alpha_{\text{low}})$$
$$\alpha_2 = \alpha_{\text{low}} + \phi(\alpha_{\text{high}} - \alpha_{\text{low}})$$

منظور از ϕ در اینجا، عدد نسبت طلایی است.

3. حال مقدار $f(x^t + \alpha p^t)$ را در این دو نقطه محاسبه کنید.
4. مقادیر به دست آمده را با مقادیر دو نقطه قبلی مقایسه کنید و بازه را کوچک کنید.
5. آنقدر این مراحل را تکرار کنید تا بازه به اندازه کافی کوچک شود، سپس نقطه وسط بازه را به عنوان α در نظر بگیرید.

2. Newton’s method

این روش معمولا از روش‌های دیگر سریع‌تر است، ولی احتیاج به مشتق مرتبه اول ودوم تابع f دارد.

1. یک حدس اولیه برای مقدار α در نظر بگیرید.
2. مشتق اول و دوم تابع f را حساب کنید:

$$f'(\alpha) = \frac{d}{d\alpha} f(x^t + \alpha p^t)$$
$$f''(\alpha) = \frac{d^2}{d\alpha^2} f(x^t + \alpha p^t)$$

3. حال مقدار α را با توجه به مشتق مرتبه اول و دوم آپدیت کنید:

$$\alpha_{k+1} = \alpha_k - \frac{f'(\alpha_k)}{f''(\alpha_k)}$$

4. آنقدر این کار را تکرار کنید که $|f''(\alpha_k)|$ بسیار کوچک و نزدیک به صفر شود.

3. Gradient descent (Backtracking with line search)

این روش مقدار α را با توجه به اینکه تابع f چقدر سریع کاهشی می‌شود تعیین می‌کند.

1. یک حدس اولیه برای مقدار α در نظر بگیرید.
2. چک کنید که آیا نامساوی زیر برقرار است یا نه؟

$$f(x^t + \alpha p^t) \leq f(x^t) + c\alpha \nabla f(x^t)^T p^t$$

c در اینجا از پارامترهای مسئله و یک مقدار ثابت بین 0 و 1 است.

3. اگر این نامساوی برقرار نبود، مقدار α را با کمک پارامتر β آپدیت کنید. β در اینجا مقداری ثابت بین صفر و یک است:

$$\alpha = \beta\alpha$$

4. آنقدر این کار را تکرار کنید تا نامساوی بالا برقرار شود.

ت) تابع $f(x_1, x_2) = \left(x_1 + x_2^2\right)^2$ را در نظر بگیرید. در نقطه $x^T = (0, 1)$ جهت جستجوی $p^T = (-1, -1)$ را در نظر می‌گیریم. نشان دهید که p یک جهت نزول است و تمامی مقادیر مینیمم کننده مسئله زیر را در هر تکرار t بیابید.

طبق قسمت ب، برای نشان دادن اینکه p یک جهت نزول است باید درستی رابطه زیر را برای x , p داده شده محاسبه کنیم:

$$\nabla f(x^t)^T p^t < 0$$

اول گرادیان تابع f را حساب می‌کنیم:

$$\frac{\partial f}{\partial x_1} = 2(x_1 + x_2^2)$$

$$\frac{\partial f}{\partial x_2} = 2(x_1 + x_2^2) \cdot 2x_2 = 4x_2(x_1 + x_2^2)$$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2(\mathrm{x}_1 + \mathrm{x}_2^2) \\ 4x_2(x_1 + x_2^2) \end{bmatrix}$$

سپس مقدار گرادیان را در نقطه $x^T = (0,1)$ محاسبه می‌کنیم:

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2 \times (0+1) \\ 4 \times 1(0+1) \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

حال:

$$\nabla f(x^t)^T p^t = \begin{bmatrix} 2 \\ 4 \end{bmatrix} . [-1 \quad -1\,] = -6$$

از آنجایی که مقدار این عبارت کمتر از صفر شد، پس تابع در این جهت نزولی است.

برای محاسبه مقادیر مینیمم کننده:

ابتدا مقدار تابع f را برای نقاط داده شده محاسبه می‌کنیم:

$$x^t + \alpha p^t = (0,1) + \alpha(-1,-1) = (-\alpha, 1-\alpha)$$

$$f(-\alpha, 1-\alpha) = (-\alpha + (1-\alpha)^2)^2$$

$$g(\alpha) = (1 - 3\alpha + \alpha^2)^2$$

برای پیدا کردن نقاط مینیمم، باید مشتق این تابع را محاسبه کرده و برابر صفر قرار دهیم:

$$g'(\alpha) = 2(1 - 3\alpha + \alpha^2) \cdot (-3 + 2\alpha) = 0$$

$$-3 + 2\alpha = 0$$

$$\alpha = \frac{3}{2}$$

$$1 - 3\alpha + \alpha^2 = 0$$

$$\alpha = \frac{3 \pm \sqrt{9-4}}{2} = \frac{3 \pm \sqrt{5}}{2}$$