هوش مصنوعی

نیم‌سال اول ۰۱-۰۰

مدرس: دکتر محمدحسین رهبان

دانشکده مهندسی کامپیوتر

تمرین دوم ـ بخش دوم

محمدجواد هزاره — شماره دانشجویی: ۹۸۱۰۱۰۷۴

سوال ۱

در این روش متغیر t را از یک شروع کرده و یکی یکی زیاد می‌کنیم و تا زمانی که دما یا همان متغیر T به صفر برسد الگوریتم را ادامه می‌دهیم. بنابراین مراحل اجرای الگوریتم به صورت زیر خواهد بود:

۱. در این مرحله $t = 1$ و برای وضعیت کنونی داریم $value(S_{current}) = -1$. برای T نیز داریم $T = 2$. برای انتخاب تصادفی یکی از همسایه‌های وضعیت کنونی نیز با توجه به صورت سوال خواهیم داشت $value(S_{new}) = -2$. بنابراین $\Delta E = -2 + 1 = -1$ و از آنجایی که اختلاف ارزش وضعیت‌ها منفی‌ست، به صورت احتمالاتی این وضعیت را انتخاب خواهیم کرد. احتمال انتخاب شدن وضعیت جدید برابر $e^{\frac{\Delta E}{T}} = e^{-\frac{1}{2}} \approx 0.6$ خواهد بود که چون بیش‌تر از 0.5 است این وضعیت انتخاب خواهد شد. بنابراین استیت جدید ما وضعیتی خواهد بود که ارزش آن -2 است.

۲. در این مرحله $t = 2$ و با توجه به زمان‌بند داده شده داریم $T = 1$. با توجه به مرحله قبل، برای ارزش وضعیت کنونی داریم $value(S_{current} = -2)$. به طور مشابه برای وضغیت جدید انتخابی نیز داریم $value(S_{new} = -2)$. پس اختلاف ارزش‌ها برابر $\Delta E = 0$ خواهد بود. در این حالت چون $\Delta E > 0$ نیست، انتخاب به صورت احتمالاتی صورت می‌گیرد. احتمال انتخاب شدن وضعیت جدید برابر $e^{\frac{0}{T}} = 1$ خواهد بود و درنتیجه وضعیت جدید انتخاب می‌شود. بنابراین در پایان این مرحله برای وضعیت کنونی داریم $value(S_{current} = -2)$.

۳. در این مرحله $t = 3$ و در نتیجه با توجه به زمان‌بند داده شده $T = 0$ خواهد بود. در نتیجه در این مرحله الگوریتم به پایان می‌رسد و وضعیتی با ارزش -2 خروجی داده می‌شود.

سوال ۲

آ) برای این منظور باید تابع داده شده محدب باشد. برای بررسی محدب بودن تابع باید ماتریس Hessian آن را بررسی کنیم. برای این ماتریس داریم:

$$H_f = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 12x_1^2 & 0 \\ 0 & 4 \end{pmatrix}$$

این ماتریس در تمام نقاط (x_1, x_2) مثبت نیمه معین است، چرا که دترمینان‌های مربعی آن در این نقاط نامنفی است. بنابراین تابع محدب بوده و با اجرای الگوریتم gradient descent و با انتخاب مناسب α به نقطه کمینه سراسری همگرا می‌شویم.

ب) برای گرادیان f داریم:

$$\nabla f = (4x_1^3 + 1, 4x_2)$$

پس مراحل اجرای الگوریتم با $\alpha = 0.0001$ به صورت زیر خواهد بود:

۱. مقدار گرادیان در نقطه $x_0 = (-1, 0)$ برابر $(-3, 0)$ خواهد بود، بنابراین برای مقدار جدید x داریم:

$$x_1 = x_0 - \alpha \nabla f = (-1, 0) - 0.0001(-3, 0) = (-0.9997, 0)$$

۲. در این مرحله گرادیان در نقطه x_1 تقریبا برابر $(-2.9964, 0)$ خواهد بود، پس برای x جدید داریم:

$$x_2 = x_1 - \alpha \nabla f = (-0.9997, 0) - 0.0001(-2.9964, 0) = (-0.9994, 0)$$

۳. در این مرحله نیز گرادیان در نقطه x_2 تقریبا برابر $(-2.9928, 0)$ خواهد بود که درنتیجه برای x_3 داریم:

$$x_3 = x_2 - \alpha \nabla f = (-0.9994, 0) - 0.0001(-2.9928, 0) = (-0.9991, 0)$$

با توجه به گرادیان f می‌دانیم این تابع در نقطه‌ای با مختصات تقریبی $(-0.63, 0)$ مقدار کمینه را به خود می‌گیرد. پس همانطور که دیده می‌شود اجرای الگوریتم با $\alpha = 0.0001$ بسیار کند به سمت نقطه کمینه حرکت می‌کند.

پ) مراحل اجرای الگوریتم با $\alpha = 1$ به صورت زیر خواهد بود:

۱. در نقطه شروع گرادیان $(-3,0)$ است، پس داریم:

$$x_1 = x_0 - \alpha \nabla f = (-1,0) - (-3,0) = (2,0)$$

۲. در نقطه x_1 گرادیان برابر $(33,0)$ خواهد بود، پس:

$$x_2 = x_1 - \alpha \nabla f = (2,0) - (33,0) = (-31,0)$$

۳. در x_2 نیز گرادیان برابر $(-119163,0)$ خواهد بود :) که در نتیجه داریم:

$$x_3 = x_2 - \alpha \nabla f = (-31,0) - (-119163,0) = (119132,0)$$

همانطور که دیده می‌شود با $\alpha = 1$ در هر مرحله بیش‌تر از نقطه کمینه فاصله می‌گیریم و درنتیجه با انتخاب این مقدار برای α الگوریتم به مقدار کمینه همگرا نخواهد شد.

ت) با توجه به قسمت (ب) و (پ) نتیجه می‌گیریم که اگر α را خیلی کوچک انتخاب کنیم با سرعتی بسیار کند به نقطه کمینه همگرا خواهیم شد و اگر α را خیلی بزرگ انتخاب کنیم به طور کلی همگرایی به نقطه کمینه را از دست خواهیم داد. بنابراین انتخاب α چیزی میان این دو مقدار و حدود 0.1 مقدار مناسبی خواهد بود.