이상구 | 성균관대학교

9주차. 경사하강법과 최소제곱문제의 해

* 경사하강법(gradient descent method) : 주어진 함수의 최솟값을 구하는 대표적인 수치 최적화 방법 중 하나이다.

[경사하강법] 알고리즘
[단계 1] 초기 근사해 x_1, 허용오차(tolerance) $0 \le \epsilon \ll 1$, 학습률(learning rate) η (eta, 에타)를 준다. $k := 1$이라 한다.
[단계 2] $g_k = f'(x_k)$를 계산한다. 만일 $\lvert g_k \rvert \le \epsilon$ 이면, 알고리즘을 멈춘다.
[단계 3] $x_{k+1} = x_k - \eta g_k$, $k := k+1$이라 두고 [단계 2]로 이동한다.

* 학습률(learning rate) : 경사하강법을 사용할 때 시작 점에서 다음 점으로 이동할 때의 보폭의 크기를 의미한다.

* 다변수 함수 : x, y를 독립적으로 변화하는 두 변수라 하고 z를 제 3의 변수라 하자. x, y의 값이 각각 정해지면 여기에 대응하여 z의 값이 정해질 때 z를 두 변수 x와 y의 함수라 하고 이것을 $z = f(x, y)$로 표시한다. 같은 방법으로 더 많은 변수의 함수도 정의할 수 있다. 이와 같이 이변수 이상의 함수를 일반적으로 다변수함수라 한다.

* 그래디언트(gradient) : 2변수 함수 $f(x, y)$의 그래디언트는 다음과 같이 정의된다.

$$\text{grad } f(x, y) = \nabla f(x, y) = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix}$$

여기서 $\dfrac{\partial f}{\partial x}$는 f를 변수 x에 관하여 편미분한다는 뜻으로 x를 제외한 다른 변수는 모두 상수로 취급하여 미분하는 것과 같다. $\dfrac{\partial f}{\partial y}$도 마찬가지로 이해할 수 있다.

같은 방법으로 독립변수가 3개 이상인 다변수함수 f에 대하여도 f의 그래디언트를 정의할 수 있다. 예를 들어, n변수 함수 $f(x_1, x_2, \ldots, x_n)$에 대하여 그래디언트는 다음과 같다.

$$\text{grad } f(x_1, x_2, \ldots, x_n) = \nabla f(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$