Gradient Descent Problem with Solution

Course: Introduction to Data Mining
Professor: Dr. Tahaei
Author: Parinaz Kanan

December 25, 2024

Problem 1. Consider the function

$$f(x,y) = \left(\frac{3}{4}x - \frac{3}{2}\right)^2 + (y-2)^2 + \frac{1}{4}xy$$

with gradient

$$\nabla f(x,y) = \left(\frac{9}{8}x - \frac{9}{4} + \frac{1}{4}y, 2y - 4 + \frac{1}{4}x\right)$$

We run gradient descent for 5 iterations within initial position (5, 4), while varying the learning rate in the range $\gamma = \{0.01, 0.1, 0.2, 0.3, 0.5, 0.75\}$.

Solution. we solve the problem for learning rates 0.1 and 0.01:

Learning Rate: 0.01:

Iteration 1:

$x = 5, y = 4$

$\nabla f(x,y) = \left(\frac{9}{8}5 - \frac{9}{4} + \frac{1}{4}4, 2 \cdot 4 - 4 + \frac{1}{4}5\right) = (2.2500, 5.5250)$

$x = x - 0.01 \cdot \nabla f(x,y)[0] = 5 - 0.01 \cdot 2.2500 = 4.9775$

$y = y - 0.01 \cdot \nabla f(x,y)[1] = 4 - 0.01 \cdot 5.5250 = 3.9475$

$f(x,y) = \left(\frac{3}{4}4.9775 - \frac{3}{2}\right)^2 + (3.9475 - 2)^2 + \frac{1}{4}4.9775 \cdot 3.9475 = 13.6918$

Iteration 2:

$x = 4.9775, y = 3.9475$

$\nabla f(x,y) = \left(\frac{9}{8}4.9775 - \frac{9}{4} + \frac{1}{4}3.9475, 2 \cdot 3.9475 - 4 + \frac{1}{4}4.9775\right) = (2.2331, 5.1394)$

$x = x - 0.01 \cdot \nabla f(x,y)[0] = 4.9625 - 0.01 \cdot 2.2331 = 4.9552$

$y = y - 0.01 \cdot \nabla f(x,y)[1] = 3.9475 - 0.01 \cdot 5.1394 = 3.8961$

$f(x,y) = \left(\frac{3}{4}4.9552 - \frac{3}{2}\right)^2 + (3.8961 - 2)^2 + \frac{1}{4}4.9552 \cdot 3.8961 = 13.3340$

Iteration 3:

$x = 4.9552, y = 3.8961$

$\nabla f(x,y) = \left(\frac{9}{8}4.9552 - \frac{9}{4} + \frac{1}{4}3.8961, 2 \cdot 3.8961 - 4 + \frac{1}{4}4.9552\right) = (2.2164, 5.0130)$

$x = x - 0.01 \cdot \nabla f(x,y)[0] = 4.9552 - 0.01 \cdot 2.2164 = 4.9330$

$y = y - 0.01 \cdot \nabla f(x,y)[1] = 3.8961 - 0.01 \cdot 5.0130 = 3.8458$

$f(x,y) = \left(\frac{3}{4}4.9330 - \frac{3}{2}\right)^2 + (3.8458 - 2)^2 + \frac{1}{4}4.9330 \cdot 3.8458 = 12.9887$

Iteration 4:

$x = 4.9330, y = 3.8458$

$\nabla f(x,y) = \left(\frac{9}{8}4.88980 - \frac{9}{4} + \frac{1}{4}3.84991, 2 \cdot 3.84991 - 4 + \frac{1}{4}4.88980\right) = (2.1998, 4.9248)$

$x = x - 0.01 \cdot \nabla f(x,y)[0] = 4.9330 - 0.01 \cdot 2.1998 = 4.9110$

$y = y - 0.01 \cdot \nabla f(x,y)[1] = 3.8458 - 0.01 \cdot 4.9248 = 3.7965$
$f(x,y) = \left(\frac{3}{4}4.9110 - \frac{3}{2}\right)^2 + (3.7965 - 2)^2 + \frac{1}{4}4.9110 \cdot 3.7965 = 12.6554$

Iteration 5:
$x = 4.9110, y = 3.7965$
$\nabla f(x,y) = \left(\frac{9}{8}4.9110 - \frac{9}{4} + \frac{1}{4}3.7965, 2 \cdot 3.7965 - 4 + \frac{1}{4}4.9110\right) = (2.1833, 4.8208)$
$x = x - 0.01 \cdot \nabla f(x,y)[0] = 4.9110 - 0.01 \cdot 2.1833 = 4.8892$
$y = y - 0.01 \cdot \nabla f(x,y)[1] = 3.7965 - 0.01 \cdot 4.8208 = 3.7483$
$f(x,y) = \left(\frac{3}{4}4.8892 - \frac{3}{2}\right)^2 + (3.7483 - 2)^2 + \frac{1}{4}4.8892 \cdot 3.7483 = 12.3336$

Similarly, for learning rate of 0.1 we have:
Iteration 1:
$x = 5, y = 4$
$\nabla f(x,y) = \left(\frac{9}{8} \cdot 5 - \frac{9}{4} + \frac{1}{4} \cdot 4, 2 \cdot 4 - 4 + \frac{1}{4} \cdot 5\right) = (2.25, 5.25)$
$x = x - 0.1 \cdot \nabla f(x,y)[0] = 5 - 0.1 \cdot 2.25 = 4.775$
$y = y - 0.1 \cdot \nabla f(x,y)[1] = 4 - 0.1 \cdot 5.25 = 3.475$
$f(x,y) = \left(\frac{3}{4} \cdot 4.775 - \frac{3}{2}\right)^2 + (3.475 - 2)^2 + \frac{1}{4} \cdot 4.775 \cdot 3.475 = 10.6555$

Iteration 2:
$x = 4.775, y = 3.475$
$\nabla f(x,y) = \left(\frac{9}{8} \cdot 4.775 - \frac{9}{4} + \frac{1}{4} \cdot 3.475, 2 \cdot 3.475 - 4 + \frac{1}{4} \cdot 4.775\right) = (2.0813, 4.1438)$
$x = x - 0.1 \cdot \nabla f(x,y)[0] = 4.5669$
$y = y - 0.1 \cdot \nabla f(x,y)[1] = 3.0606$
$f(x,y) = 8.3255$

Iteration 3:
$x = 4.5669, y = 3.0606$
$\nabla f(x,y) = \left(\frac{9}{8} \cdot 4.5669 - \frac{9}{4} + \frac{1}{4} \cdot 3.0606, 2 \cdot 3.0606 - 4 + \frac{1}{4} \cdot 4.5669\right) = (1.9252, 3.2630)$
$x = x - 0.1 \cdot \nabla f(x,y)[0] = 4.3744$
$y = y - 0.1 \cdot \nabla f(x,y)[1] = 2.7343$
$f(x,y) = 6.7006$

Iteration 4:
$x = 4.3744, y = 2.7343$
$\nabla f(x,y) = \left(\frac{9}{8} \cdot 4.3744 - \frac{9}{4} + \frac{1}{4} \cdot 2.7343, 2 \cdot 2.7343 - 4 + \frac{1}{4} \cdot 4.3744\right) = (1.7808, 2.5622)$
$x = x - 0.1 \cdot \nabla f(x,y)[0] = 4.1963$
$y = y - 0.1 \cdot \nabla f(x,y)[1] = 2.4781$
$f(x,y) = 5.5416$

Iteration 5:
$x = 4.1963, y = 2.4781$
$\nabla f(x,y) = \left(\frac{9}{8} \cdot 4.1963 - \frac{9}{4} + \frac{1}{4} \cdot 2.4781, 2 \cdot 2.4781 - 4 + \frac{1}{4} \cdot 4.1963\right) = (1.6472, 2.0053)$
$x = x - 0.1 \cdot \nabla f(x,y)[0] = 4.0316$
$y = y - 0.1 \cdot \nabla f(x,y)[1] = 2.2776$
$f(x,y) = 4.6942$

Iteration	Learning Rate = 0.01	Learning Rate = 0.1
1	(2.2500, 5.5250)	(2.25, 5.25)
2	(2.2331, 5.1394)	(2.0813, 4.1438)
3	(2.1975, 4.7522)	(1.9252, 3.2630)
4	(2.1571, 4.3653)	(1.7808, 2.5622)
5	(2.1110, 3.9781)	(1.6472, 2.0053)

compared to learning rate of 0.01, 0.1 seemingly has a better performance.
0.01 is considered too small of a learning rate for this problem and makes the learning process slow. □