CMO Assignment 0

Biswadeep Debnath BISWADEEPD@IISC.AC.IN
24528

Question 1

The oracle provides access to both the function value $f(x)$ and its first derivative $f'(x)$ for any real input x.
Given this, gradient descent can be used for finding the minimum of the function, as it iteratively updates x by moving in the direction opposite to the gradient to reduce the function value.

Gradient Descent

The gradient descent update rule is given by:

$$\theta := \theta - \eta \nabla J(\theta) \tag{1}$$

where η is the learning rate, and $J(\theta)$ represents the objective function to be minimized: (Wikipedia, 2025).

Gradient Descent with momentum

Gradient descent can be improved by using *momentum*. Instead of just moving the point based on the current slope, momentum also considers the previous movement's direction and speed. (Rumelhart et al., 1986; Ruder, 2016).

$$\mathbf{v}_t = \gamma \mathbf{v}_{t-1} + \eta \nabla_\theta J(\theta) \tag{2}$$

$$\theta := \theta - \mathbf{v}_t \tag{3}$$

The term γ is called *momentum*, which is usually set to 0.9 and η is the learning rate.

In this assignment, both methods have been implemented to test and *Gradient Descent with momentum* has been chosen for final submission.
In a visualization of the change of $f(x)$ value with number of iterations or steps its observed that both methods seem to converge to a single value but with momentum method it converges faster. (Figure 1)

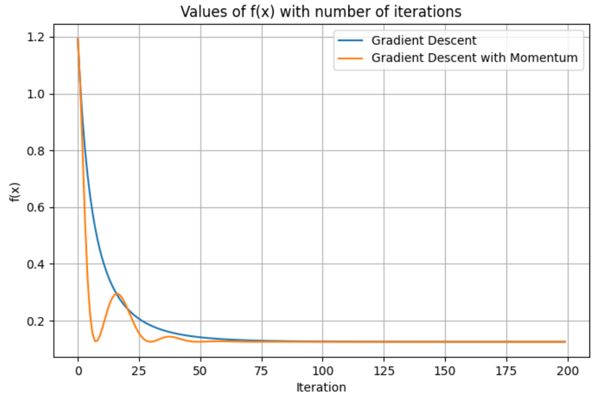

Figure 1: Comparison of convergence of $f(x)$ for gradient descent and gradient descent with momentum.

The first derivative is also checked for Gradient Descent with Momentum (Table 1). Following parameters was used

- Initial Value: $x_0 = 1$
- SR Number: `SR_No` = 24528
- Learning Rate: $\eta = 0.01$
- Momentum: $\gamma = 0.9$

Table 1: Results using gradient descent with momentum

Iterations	**x* (Obtained Minima)**	**f(x*) (Minimum value)**	**f′(x*) (First derivative)**
10	0.1101261433760414	0.1732214496495718	-0.44641438373689474
100	0.31145426719809	0.12548970282353852	-0.011593134572461805
1000	0.31610467883603666	0.1254627055118043	-5.268913394473884e-08
10000	0.31610467883603666	0.1254627055118043	-5.268913394473884e-08

It is also observed that the first derivative values at the obtained minima remain close to zero which is expected for a minimum and do not change after 1000 iterations. So, we can conclude that the function has converged to its minimum or a local minimum.

Question 2

$x^* = 0.31610467883603666$

Question 3

$f(x^*) = 0.1254627055118043$

References

Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016. URL `http://arxiv.org/abs/1609.04747`.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

Wikipedia. Gradient descent wikipedia, the free encyclopedia, 2025. URL `https://en.wikipedia.org/wiki/Gradient_descent`.