Om Shete
Group 7
2103163

Experiment No: 2

Aim: Implement Gradient descent algorithm to minimise given objective function.
For e.g., $f(x_1, x_2) = x_1^3 + 6x_2^2$.

Theory:

- A gradient is nothing but derivative that defines the effects on outputs of the function with little bit of variations in Input.
- Gradient descent stands as a cornerstone orchestrating the intricate dance of model optimization.
- As its core, it is numerical optimization algorithm that aims to find the optimal parameters - weights and biases of a neural network by minimizing a defined cost function.
- It works by iteratively adjusting the weights or parameters of the model in the direction of negative gradient of the cost function until the minimum of the cost function is reached.
- The cost function evaluates between the difference between the actual and predicted outputs.
- Gradient descent is a fundamental optimization technique in ML used to minimise the cost or loss function during model

training.

i) It iteratively adjust model parameter by moving in the direction of the steepest decrease in the cost function.

ii) The algorithm calculates gradients, representing the partial derivative of the cost function concerning each parameter.

The key steps are:

i) Initialize the parameters a and b with some initial values.

ii) Compute the gradient of the function w.r.t each parameter.

iii) Update the paramters

$$a = a - \alpha \cdot \frac{\partial f}{\partial a}$$

$$b = b - \alpha \cdot \frac{\partial f}{\partial b}$$

where,

α = learning rate.

iv) Repeat ~~step~~ (ii) and (iii) until convergence.

- For e.g, $f(x_1, x_2) = -x_1^3 + 6x_2^2$

$$\frac{\partial f}{\partial x_1} = \frac{\partial(-x_1^3 + 6x_2^2)}{\partial x_1}$$

$$= -3x_1^2$$

$$\frac{\partial f}{\partial x_2} = \frac{\partial(-x_1^3 + 6x_2^2)}{\partial x_2}$$

$$= 12x_2$$

- Update the parameters:

$$x_1 = x_1 - \alpha(-3x_1^2)$$

$$x_2 = x_2 - \alpha(12x_2)$$

$$\therefore x_1 = x_1 + 3\alpha x_1^2$$

$$x_2 = x_2 - 12\alpha x_2$$

27/9/24