Practical - 04.

* **Aim**: Implement gradient descent algorithm to find the local minima of a function.
For example - find the local minima of the function $y = (x+3)^2$ starting from the point $x = 2$.

- **Theory** -

- The objective of the experiment is to implement the gradient descent algorithm in order to locate the local minimum of a given function. Specifically, we consider the simple quadratic function

$$y = (x+3)^2.$$

and start algorithm from the initial point $x = 2$.

- By repeatedly moving in the opposite direction of the gradient, we will demonstrate how the algorithm converges to the minimum point of the function.

1] **Gradient Descent Concept**:

- Gradient descent is an iterative optimization technique used to minimize functions, especially in ML and numerical analysis.

- It relies on the fact that the gradient (or derivative in 1D) points in the direction of steepest increase of the function.
- To find a minimum, we take steps opposite to the gradient, gradually approaching the point where the gradient is zero the minimum.

2) Algorithm steps -

- Initialize - Choose a starting point x_0.
- Compute gradient - Find $\frac{dy}{dx}$ at the current x.
- Update rule -

$$x_{new} = x_{old} - \alpha \frac{dy}{dx}.$$

where alpha is the learning rate, a small positive constant that controls the step size.

- Repeat - Continue until the change in x or the gradient becomes very small, indicating convergence.

3. Given function and gradient:

- Function : $y = (x+3)^2$
  Gradient - derivative.

$$\frac{dy}{dx} = 2(x+3)$$

- The true minimum occurs where $\frac{dy}{dx} = 0$: $x = -3$.

- Starting from $x = 2$, gradient descent should move x gradually towards -3.

4. Example update calculations -

Using a learning rate $\alpha = 0.1$:

- Iteration formula:

$$x_{new} = x_{old} - 0.1 \cdot 2(x_{old} + 3).$$

- Iteration 1: $x = 2$, gradient $= 2(2+3) = 10$, new $x = 2 - 0.1 * 10 = 1.0$.

- Iteration 2: $x = 1.0$, gradient $= 2(1+3) = 8$, new $x = 1 - 0.8 = 0.2$

- Iteration 3: $x = 0.2$, gradient $= 2(0.2+3) = 6.4$
  new $x = 0.2 - 0.64 = -0.44$
  and so on converging to $x = -3$.

5. Importance in machine learning -

- Gradient descent underpins training algorithms for models like linear regression, logistic regression and neural networks.

- It works for high dimensional functions as well, using vector gradients.

- Proper selection of learning rate and stopping criteria is critical for convergence.

* CONCLUSION -

This experiment shows that the gradient descent algorithm can be successfully applied to minimize a simple differentiable function. Starting from $x=2$, by following the update rule and using an appropriate learning rate, the value of x moves closer and closer to the actual minimum at $x=-3$. This illustrates how gradient descend works. it iteratively refines the solution by moving opposite to the slope until it reaches the point of zero gradient. The same principle scales to complex, high dimensional loss functions used in modern machine learning, making gradient descent one of most fundamental optimization tool.