# Pre-class Assignment - 24

Reading :-

① Drawbacks of gradient descent :

* local minima : Gradient descent can potentially get stuck in local minima, where the function value is lower than its immediate neighbors but not the global minimum. this happens when the step size is too small or the function has a complex landscape with multiple minima.
* Tuning the learing rate : Choosing the optimal step size (learning rate) is crucial. A large learing rate can cause the algorithm to oscillate around the minimum or even diverge, while a small learning rate can lead to slow convergence.
* Sensitive to initialization : the final solution depends on the initial starting point. If the intial point is far from the minimum, it may take many iterations to reach it, or it might even get stuck in a local minimum.

② Gradient descent relation to quadratic optimization:

Gradient descent generalizes quadratic optimization methods. Quadratic optimization deals with quadratic polynomial objective functions, efficiently finding a unique minimum through closed-form solutions. In contrast, gradient descent extends to a wider-function range, accomodating non-quadratic ones lacking closed-form solutions. However, this broader applicability makes the minimization process iterative, demanding increased computational effort compared to the efficient closed-form solutions in quadratic optimization.

③ Factors to consider when choosing the step size:

* Function landscape: The step size should be small enough to avoid overshooting the minimum, especially in regions with high curvation.
* Convergence rate: A larger step size can lead to faster convergence, but it also increases the risk of divergence, so it needs to be balanced with stability.

* **Noise in the data:** If the data is noisy, a smaller step size might be needed to avoid overfitting the noise.

## Exercises:

① given $f(x) = x^3 - 3x^2 - x + 3$ for all $x \geq 0$

i) to find absolute minimum

$$f'(x) = 3x^2 - 6x - 1$$

Critical pts $\Rightarrow f'(x) = 0$

$$3x^2 - 6x - 1 = 0$$

$$x_1 = -0.15 \quad \& \quad x_2 = 2.15$$

$$f(x_1) = (-0.15)^3 - 3(-0.15)^2 - (-0.15) + 3$$
$$= 3.07$$

$$f(x_2) = (2.15)^3 - 3(2.15)^2 - (2.15) + 3$$
$$= -3.07$$

Hence, the absolute minimum is $-3.07$ at $x = 2.15$.

ii) Gradient descent with $\gamma = 1/50$ (3 rounds)

update rule for gradient descent

$$x_{new} = x_{old} - y \cdot f'(x_{old})$$

Round 1: $x_0 = 0$

$$x_1 = x_0 - \frac{1}{50} \cdot f'(x_0) = 0 - \frac{1}{50}(-1) = \frac{1}{50}$$

Round 2: $x_1 = 1/50$

$$x_2 = x_1 - \frac{1}{50} f'(x_1)$$

$$x_2 = \frac{1}{50} - \frac{1}{50}\left(3\left(\frac{1}{50}\right)^2 - 6\left(\frac{1}{50}\right) - 1\right)$$

$$x_2 = 0.0423$$

Round 3: $x_2 = 0.0423$

$$x_3 = x_2 - \frac{1}{50} f'(x_2)$$

$$x_3 = (0.0423) - \frac{1}{50}\left(3(0.0423)^2 - 6(0.0423) - 1\right)$$

$$x_3 = 0.0672$$

iii) given $\gamma = 5$

Round 1: $x_0 = 0$

$$x_1 = x_0 - 5 f'(x_0)$$

$$= 0 - 5\left(3(0) - 6(0) - 1\right)$$

$$x_1 = 5$$

<u>Round 2:</u> $x_1 = 5$

$$x_2 = x_1 - 5f'(x_1)$$

$$= 5 - 5\left(3(5)^2 - 6(5) - 1\right)$$

$$x_2 = -215$$

<u>Round 3:</u> $x_2 = -215$

$$x_3 = x_2 - 5f'(x_2)$$

$$= -215 - 5\left(3(-215)^2 - 6(215) - 1\right)$$

$$x_3 = -700035$$

The issue here is that the large learning rate causes the approximation to diverge and get worse rather than converging to the minimum. Smaller learning rates are generally better to carefully approximate the minimum.

② given, $f(x) = x^4 - x^3 - x^2 + 1$ & $\gamma = 1/100$

i) gradient descent: (start at $x = -1$)

$$f'(x) = 4x^3 - 3x^2 - 2x$$

<u>Round 1:</u> $x_0 = -1$

$$x_1 = x_0 - \gamma f'(x_0)$$

$$x_1 = -1 - \frac{1}{100}\left(4(-1)^3 - 3(-1)^2 - 2(-1)\right)$$

$x_1 = -0.95$

Round 2: $x_2 = x_1 - \frac{1}{100} f'(x_1)$

$$x_2 = -0.95 - \frac{1}{100}\left(4(-0.95)^3 - 3(-0.95)^2 - 2(-0.95)\right)$$

$x_2 = -0.90$

Round 3: $x_3 = -0.86$

Round 4: $x_4 = -0.82$

Round 5: $x_5 = -0.79$

Round 6: $x_6 = -0.76$

Round 7: $x_7 = -0.74$

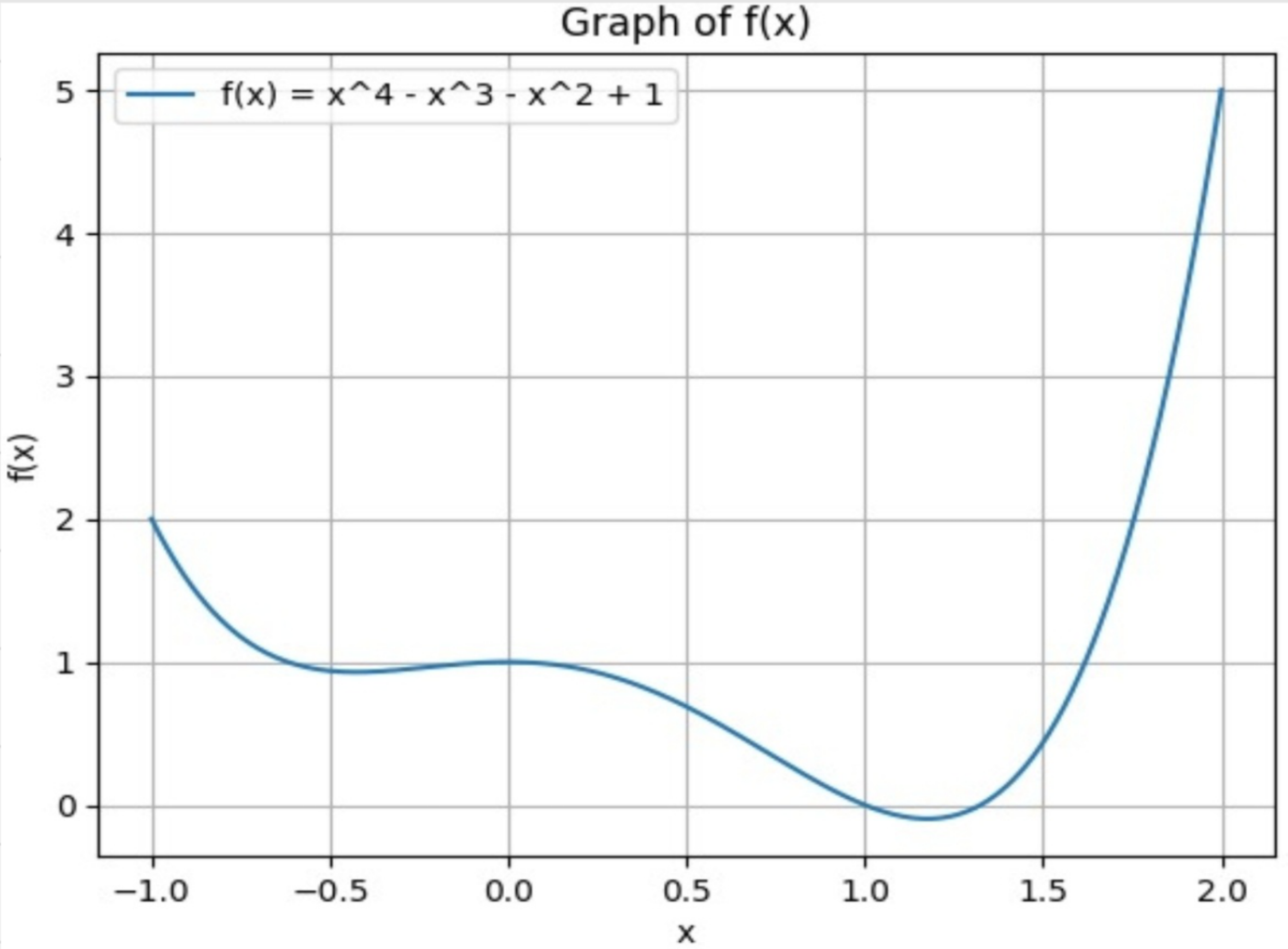
Round 8: $x_8 = -0.72$

Round 9: $x_9 = -0.70$

Round 10: $x_{10} = -0.68$

ii) Based on the computation, we seem to be getting closer and closer to $x \approx -\infty$

iii)




The graph $f(x)$ on the interval $[-1, 2]$ shows that the function is increasing within this interval. This indicates that starting at $x = -1$, gradient descent will continue moving to the left, approaching negative infinity & never reaching the actual minimum located at $x = 1$.

## What went wrong?

We started our gradient descent at $x=-1$, which is far from the actual minimum located at $x=1$. Gradient descent, while effective in finding local minima, can only converge to minima with their vicinity. Starting far outside the relevant region led us to a wrong conclusion, approaching $-\infty$ instead of the true minimum.