Gradient Decent Algorithm

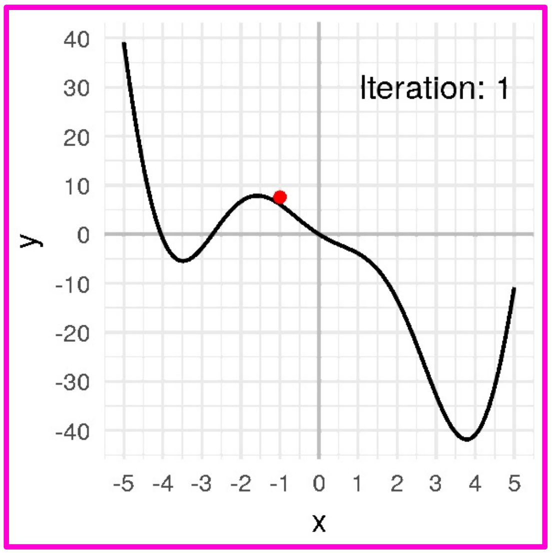
- We use Gradient Decent to approximate the minimizer of a function. minimizer is the point where the function takes its minimum value.

- GD is the core algorithm that is behind ALL of the deep learning models, including Chat GPT.

- We can summarize GD in one sentence
"To find the minimizer of a function, move in the opposite direction of the derivative of your function"

$$x_{n+1} = x_n - r\,\nabla f(x_n) \quad , \; n = 0, 1, 2, \dots N$$

$f(x) \rightarrow$ target, loss, cost, objective function
$\nabla f(x) \rightarrow$ gradient of $f(x)$
$x_0 \longrightarrow$ initial guess
$N \longrightarrow$ number of epochs
$r \longrightarrow$ learning rate



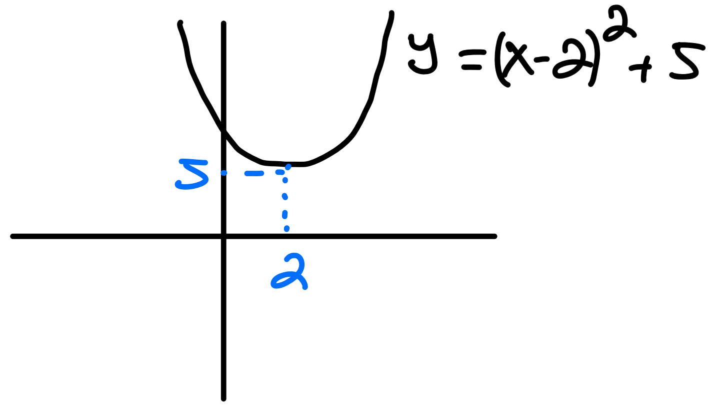
* If f is a function of one variable, i.e $f(x)$;
$\nabla f(x) = f'(x)$

* If we have $f(x,y) \Rightarrow \nabla f(x,y) = [\partial f/\partial x, \partial f/\partial y]$
is a vector.

Ex: let's start with a toy problem

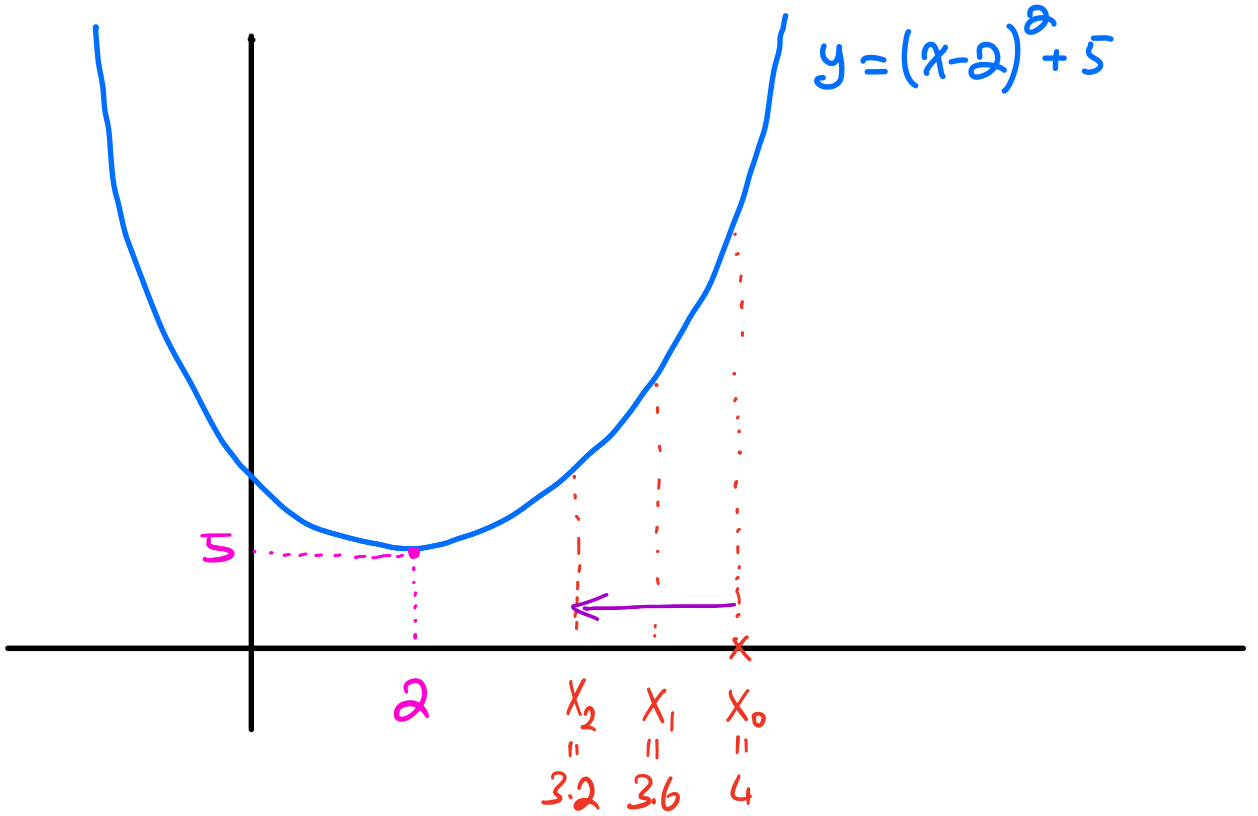
$$y = f(x) = (x-2)^2 + 5$$

We know that $\bar{x} = 2$ is the minimizer. let's approximate that.



$f'(x) = 2(x-2)$, let's pick $r = 0.1$, $x_0 = 4$

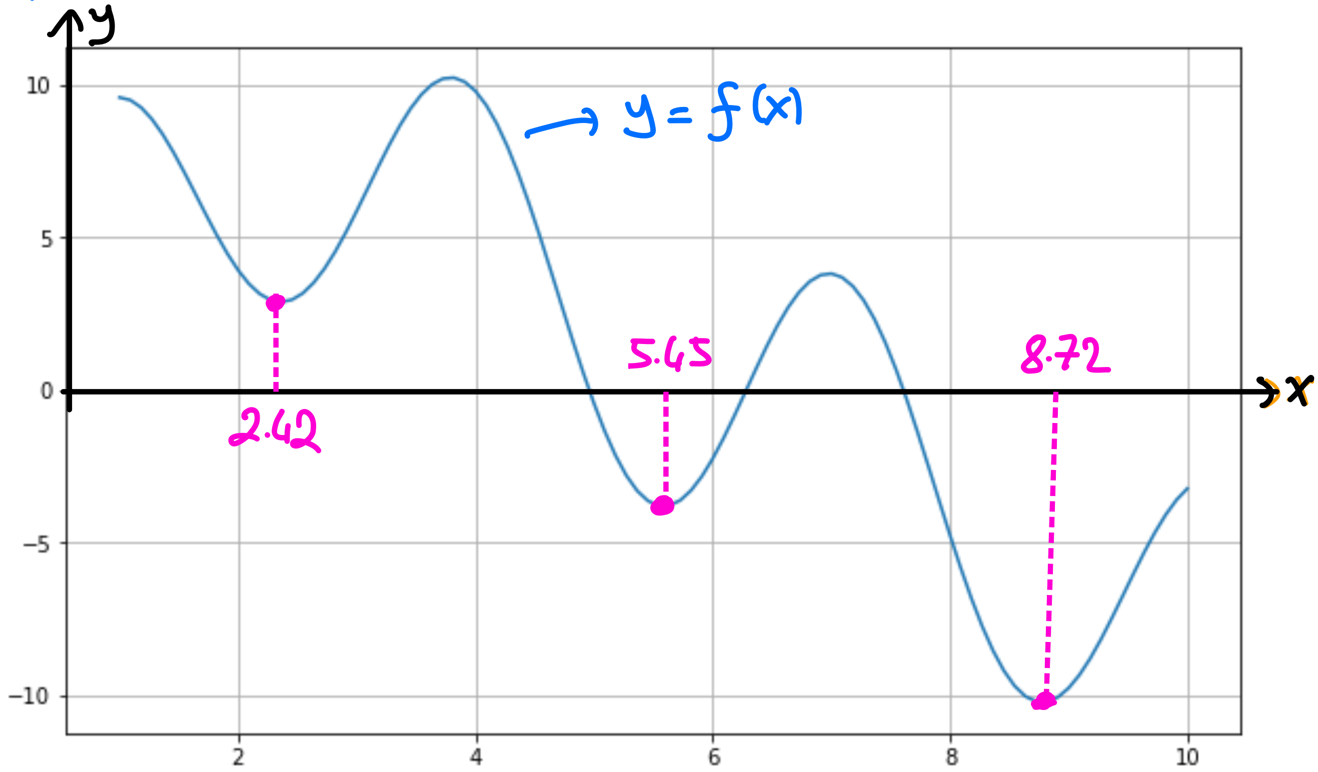
$n=0 \Rightarrow x_1 = x_0 - r f'(x_0) = 4 - 0.1 \cdot 2(0.1-2) = 3.6$
$n=1 \Rightarrow x_2 = x_1 - r f'(x_1) = 3.6 - 0.1 \cdot 2 \cdot (3.6-2) = 3.2$
$n=2 \Rightarrow x_3 = x_2 - r f'(x_2) = 3.02$
$\vdots$
$n=19 \Rightarrow x_{19} = 2.02$



Note: 1) In a real application, we cannot visualize $f(x)$. moreover we do not know if we have a global minimum. Thus we don't know if GD works but it works.

2) x_0, r, N are all important, we will see play with them in our Phyton code.

Ex: $f(x) = 2\sin x + 5\sin 2x + 7\sin(x/2)$



Jump into Pyhton code to see if you can converge to 8.72

Ex: Find the minimizer of

$$f(x,y) = (x-2)^2 + (y-3)^2 + 10$$

$$\vec{\nabla} f = [\partial f/\partial x, \partial f/\partial y] = \begin{bmatrix} 2(x-2) \\ 2(x-3) \end{bmatrix}$$

$$\vec{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad r = 0.1$$

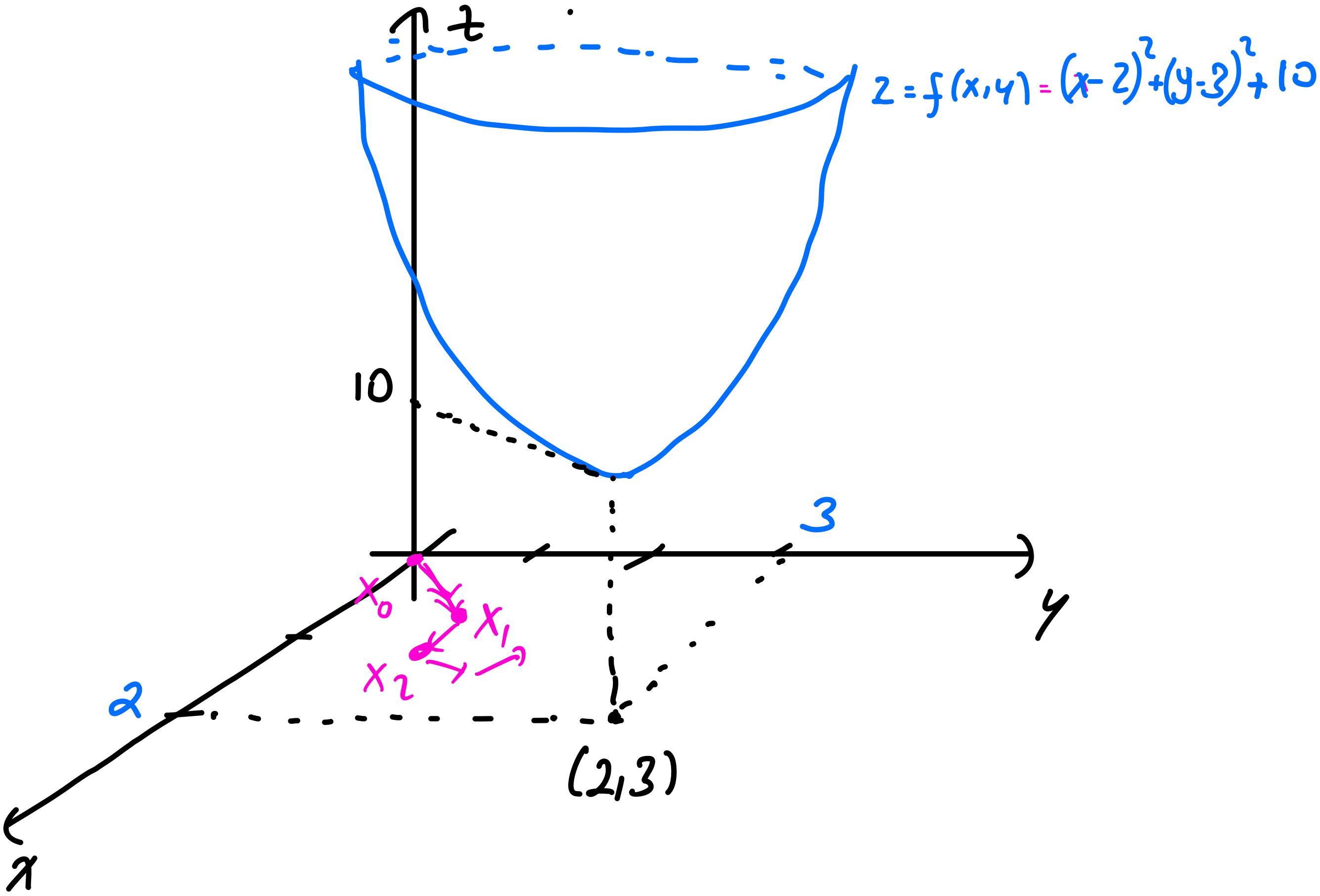
$$n=0: \quad \vec{x}_1 = \vec{x}_0 - r\nabla f(\vec{x}_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} 2(0-2) \\ 2(0-3) \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$$

$$n=1: \quad \vec{x}_2 = \vec{x}_1 - r\nabla f(\vec{x}_1) = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} - 0.1 \begin{bmatrix} 2(0.4-2) \\ 2(0.6-3) \end{bmatrix} = \begin{bmatrix} 0.72 \\ 1.08 \end{bmatrix}$$

$$\vdots$$

$$n=20: \quad \vec{x}_{20} = \begin{bmatrix} 1.9815 \\ 2.9723 \end{bmatrix} \approx \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

actual minimizer


z
$z = f(x,y) = (x-2)^2 + (y-3)^2 + 10$
10
3
y
2
x_0
x_1
x_2
(2,3)
x