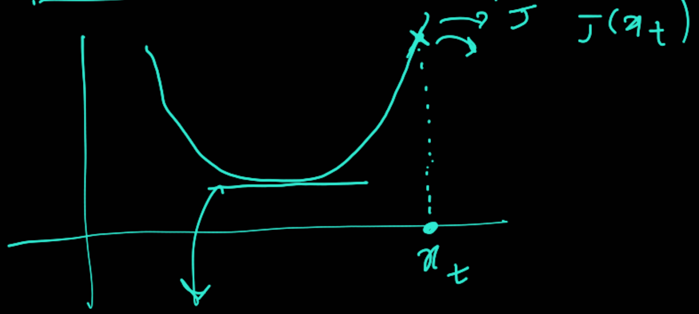
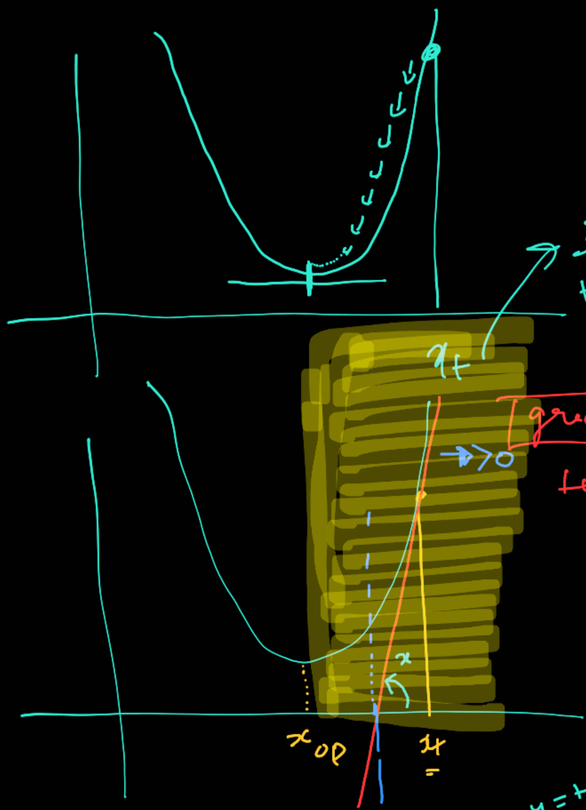
Gradient Descent



My goal is to reach optima.



I want something +ve to be deducted from x_t in this entire region.

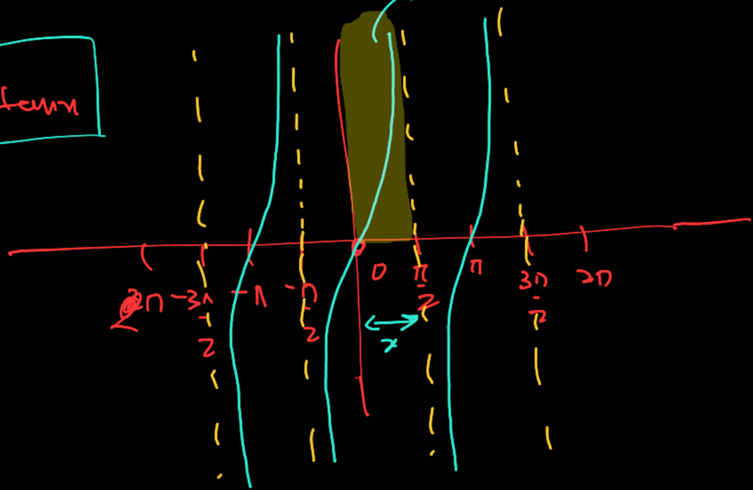
[gradient:] means tangent at that point.

+ve

$0 < x < \frac{\pi}{2}$

$y = \tan x > 0$

$y = \tan x$



$x_{op} < x_t$

$$x_t - (\boxed{\ }) = x_0$$

+ve or −ve

+ve

I need to substract some qty from x_t so that I reach optima.

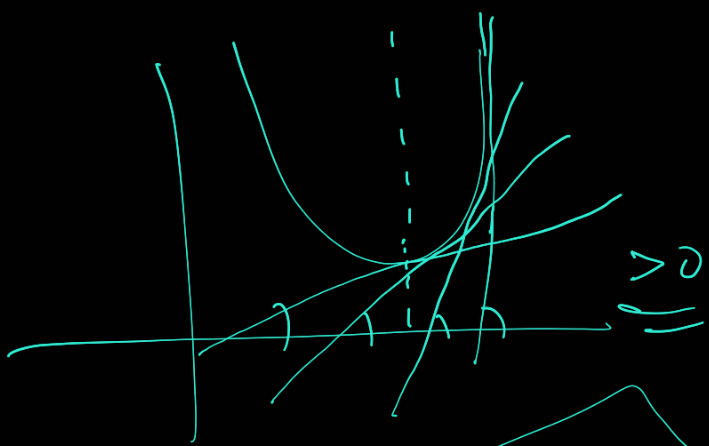
The catch is this, that, I want to reach optima in a series of steps. (iteratively)

x_0 x_t

x_{t+1}
x_{t+2}
x_{t+3}
x_{t+4}
x_{t+5}
x_{t+6}

$$x_t - (+ve) = x_{t+1}$$

x_0 x_{t+1} x_t



> 0

$x_{t+1} = x_t - (+ve)$ step size "learning rate"

$$x_{t+1} = x_t - \eta \cdot \nabla f(x_t)$$

(<1)

$\eta/2$ <0 $\eta<0$

$$x_{t+1} = x_t - \eta \cdot \nabla f(x_t)$$

$$\boxed{\theta_j^{(t+1)} \leftarrow \theta_j^{(t)} - \eta \cdot \nabla_j J(\theta)\Big|_{\theta_t}}$$

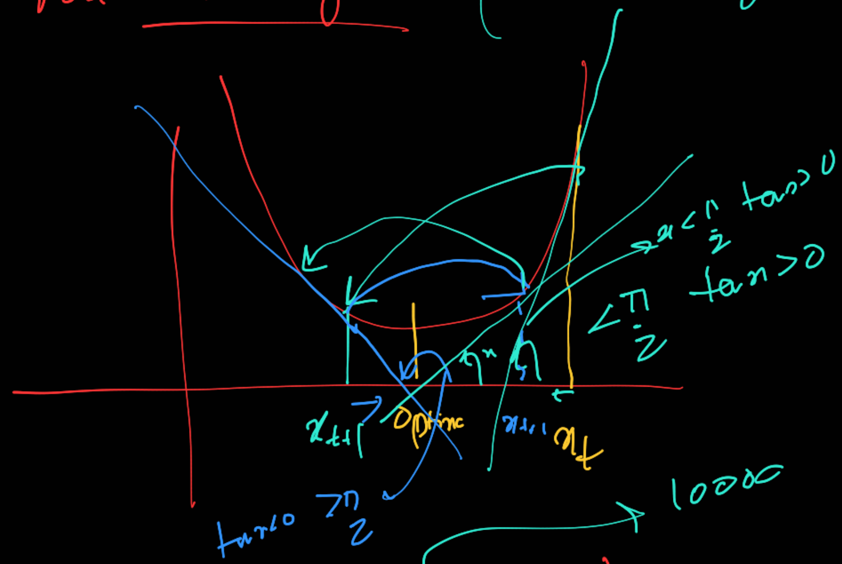
In this case

$x_0 > x_1$

$$x_{t+1} = x_t - (-ve)$$

$$\boxed{\theta_j^{(t+1)} \leftarrow \theta_j^{(t)} - \eta \cdot \nabla_j J(\theta)\Big|_{\theta^{(t)}}}$$

A note on learning rate. (Overshooting min^n)



$$x_{t+1} = x_t - \eta \cdot \nabla f|_{x_t}$$

10000

low high

$$x_{t+1} = x_t - [\quad]$$

We tend to keep our learning rate low, so that we do not "overshoot the min"

When to use NEqns & when to use GD vs Newton's Method.

(i) $$\boxed{\theta = (X^T X)^{-1} X^T y} \rightarrow \text{NE}$$

(ii) $$\theta_j^{(t+1)} \leftarrow \theta_j^{(t)} - \eta \cdot \nabla_{\theta_j} J(\theta)\Big|_{\theta^{(t)}} \quad \text{GD}$$

(iii) $$\theta_j^{(t+1)} \leftarrow \theta_j^{(t)} - (H)^{-1} \nabla J(\theta)\Big|_{\theta^{(t)}} \quad \text{NM}$$

In NE → we have inverse.

"Computationally expensive"

guarantee is also difficult.

pseudo-inv

$\boxed{NT = mn}$

$O(n^3)$

$m = 1000$

$n = 1000$

$O(N^3)$

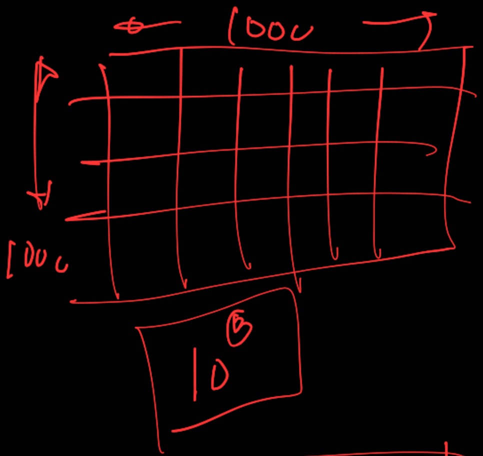
$\Rightarrow mn = 1000 \times 1000$

$= 10^6$

$O((10^6)^3)$

$\boxed{= O(10^{18})}$

entries in the matrix



GD

SGD → Batch

$\theta_j^{(t+1)} = \theta_j - \eta \nabla J(\theta)$

Δ | m → It is very fast"

But involves second order term → Hessian

"Inv"

Feynman's Learning Method

Learn in 3 stages (passes)

1. Pass 1 → Just go through the material w/o thinking about it, that what is it trying to say. Some things will go inside head. (20%)
2. Pass 2 : Now pay attention & read

2. Pass 2 : Now pay attention, ...
carefully this time. Try to fill in gaps
from prev pass. Now you will have
proper context, and made some improvements
over content from Pass 1.

3. Pass 3 : Now pay most attention,
and try to see holistic picture, solving
everything.

→ DOUBT CLEARING
→ well founded

Explain to some "novice"