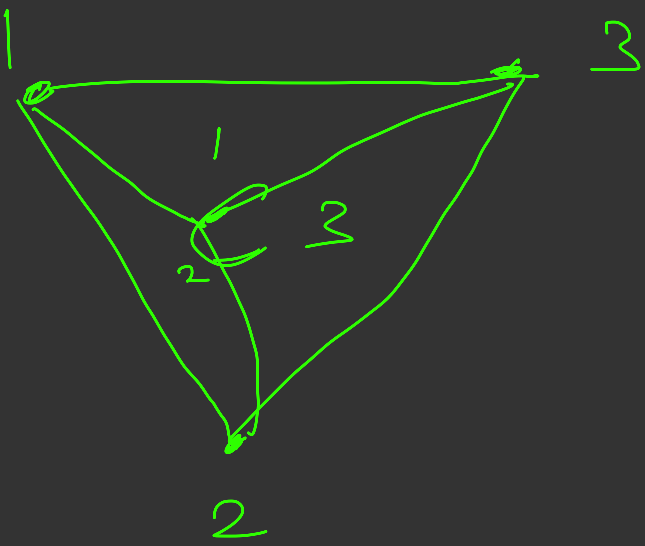
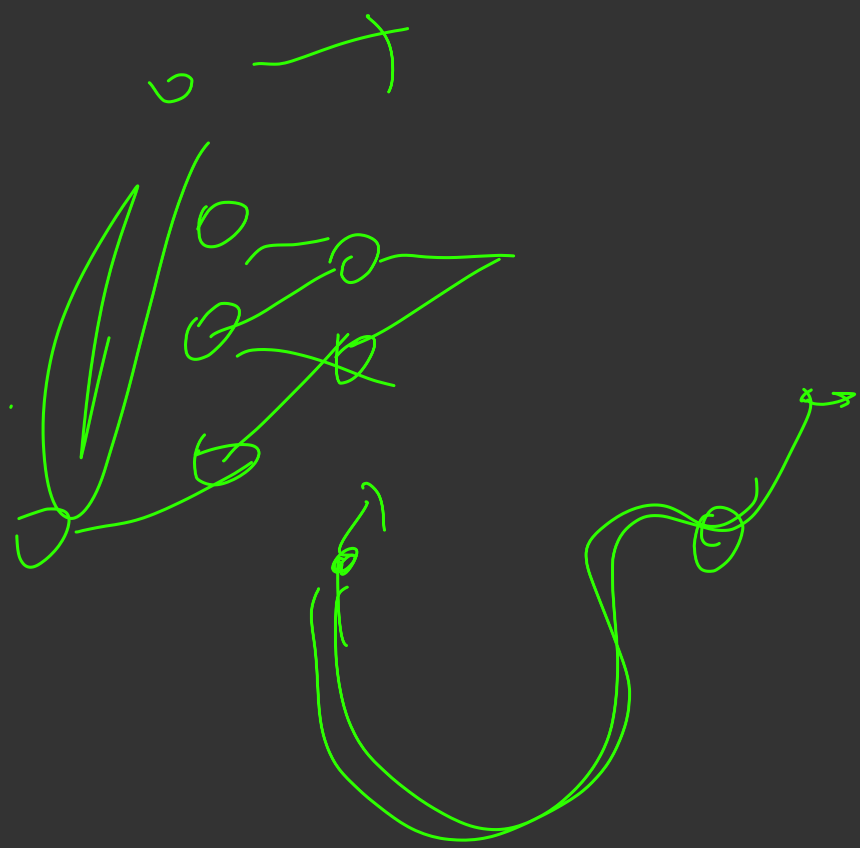
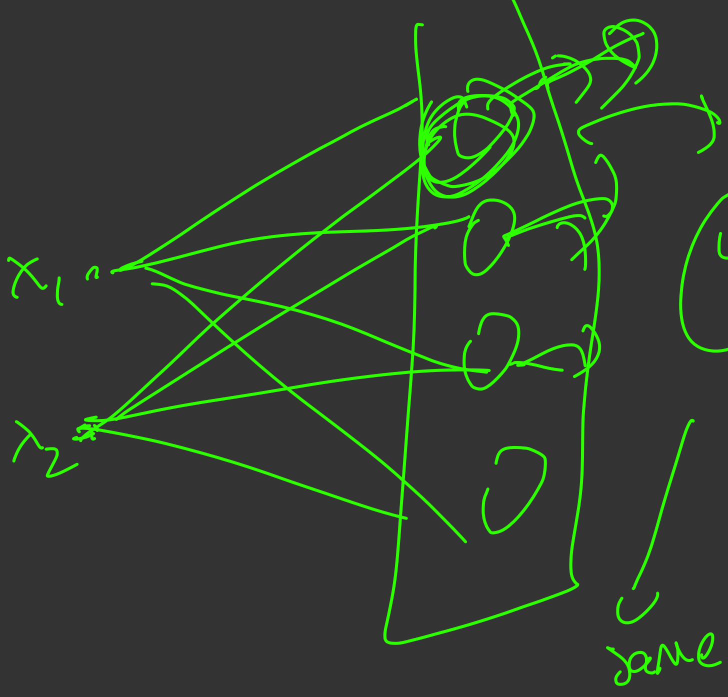


$${}^{n}C_{i}\,(1-t)^{n-i}\,t^{i}$$



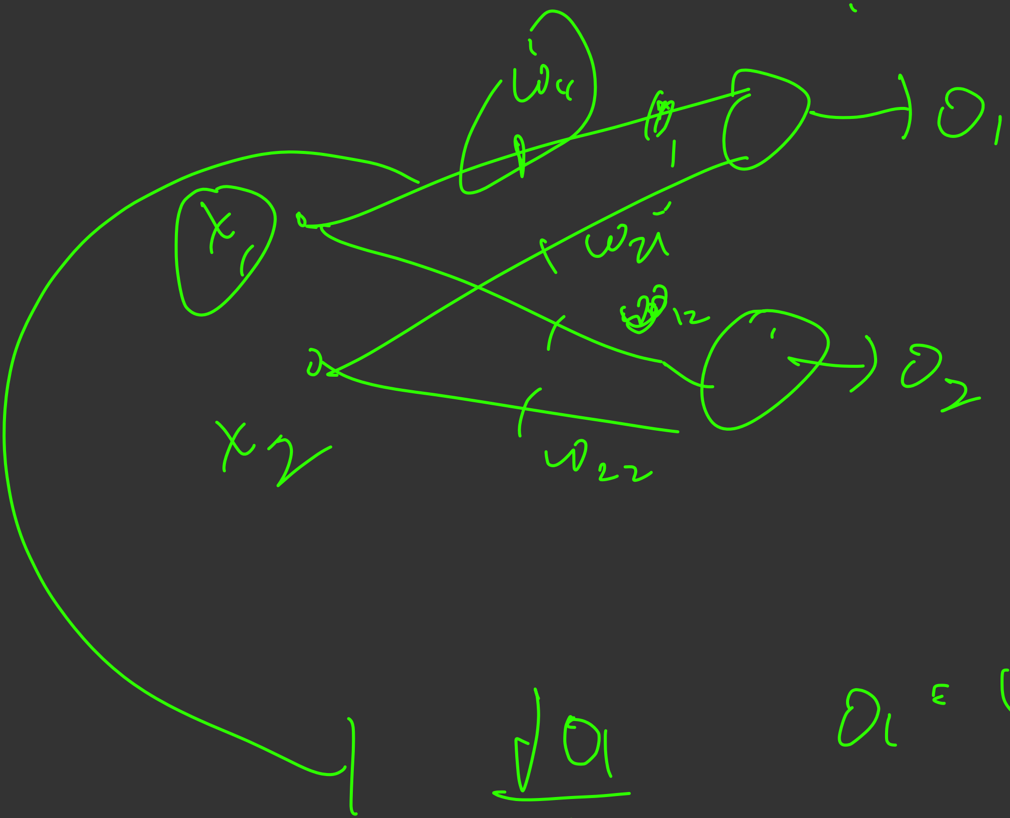


$w_n = w + \frac{dL}{dw}$

same

same

$\frac{dL}{dw}$



$O_1 = w_{11} x_1 + b$

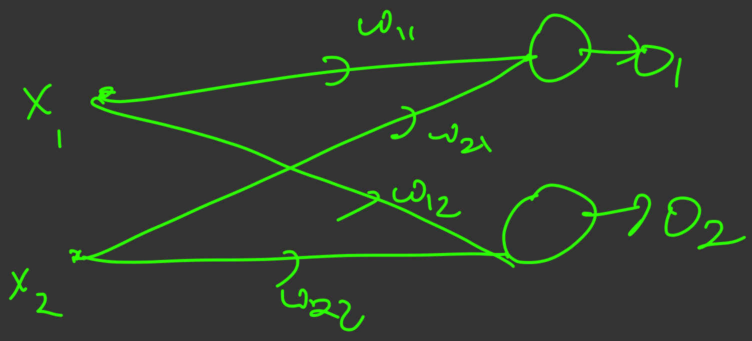
$\frac{dO_1}{dw_1}$

$w_{11} = w_{11} - \eta x_1$

$w_{21} = w_{21} - \eta x_2$

$w_{12} = w_{12} - \alpha x_1$

$w_{22} = w_{22} - \eta x_2$

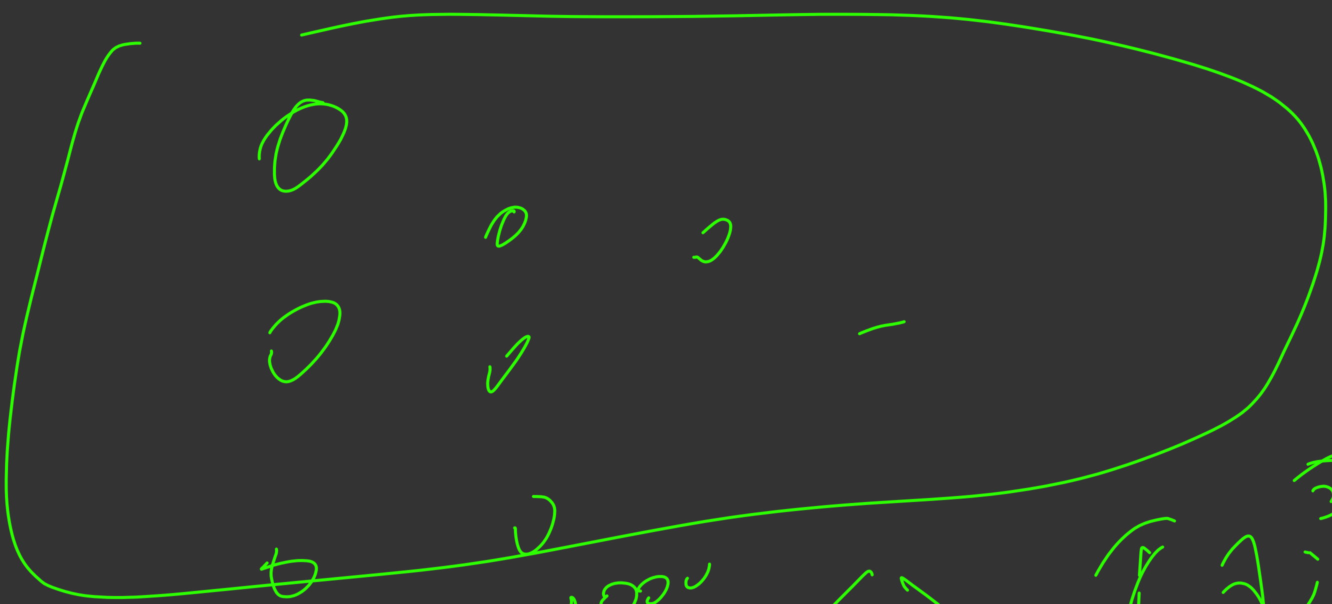


$w_{11} = w_{12} = w_{21} = w_{22}$

$$w_{11} = w_{11} - \underbrace{\frac{dO_1}{dw_{11}}}_{x_1} \cdot n \cdot k$$

$$w_{11} = w_{11} - n \cdot k \cdot x_1 \qquad w_{12} = w_{12} - n \cdot k \cdot x_1$$

$$w_{21} = w_{21} - n \cdot k \cdot x_2 \qquad w_{22} = w_{22} - n \cdot k \cdot x_2$$



0.01 0.01 0.001

$w = w - \frac{dL}{dw} \cdot n$

0.001

$\frac{dL}{dw}$

small big

$w = 0.01,\ 0.02$

0.05

$\frac{dL}{dw}$ 0.0 -1

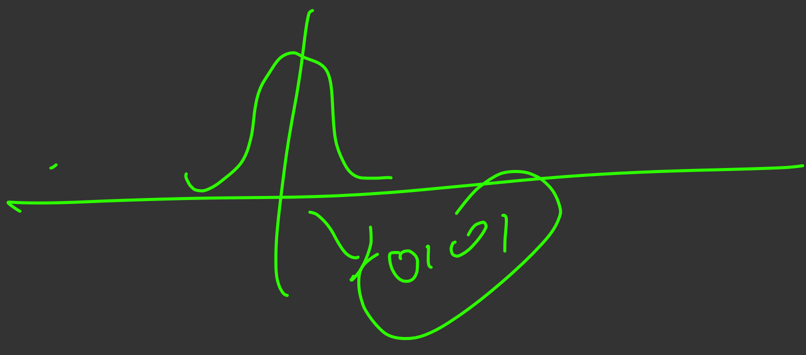
great weights

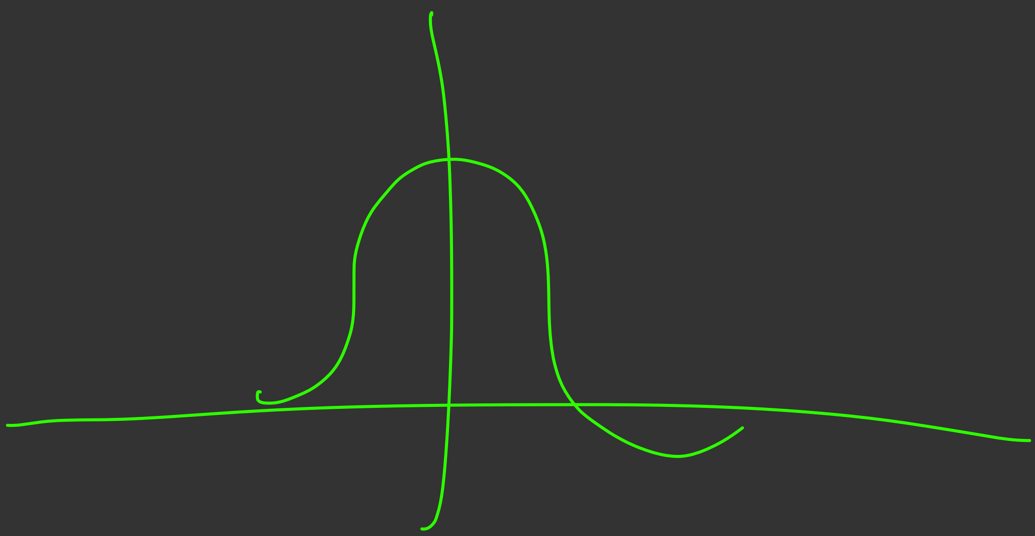
$w + \theta$

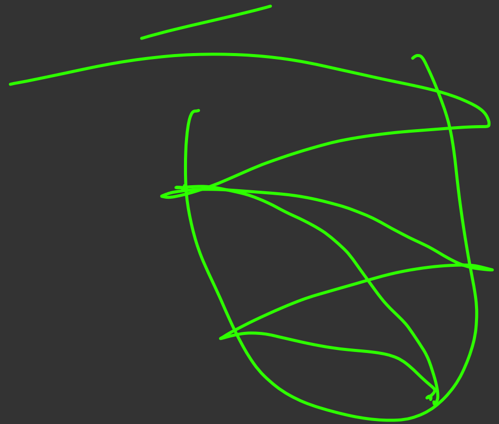
θ

$(-1, 1)$

$(-1, 1)$ 1 -1 $(-1, 1)$







$-0.$ $0.$

-0.1 0.1

-2 2

0.01 0.01

~~0.05~~

0.1 $0-0$

Imp *

For weights initialisation

if constant → all neurons converge as 1

if small random → VANISHING GRADIENT PROBLEM

if Big random → EXPLODING GRADIENT PROBLEM

Range (-0.5, 0.5) → small

Big

---

1) Xavier

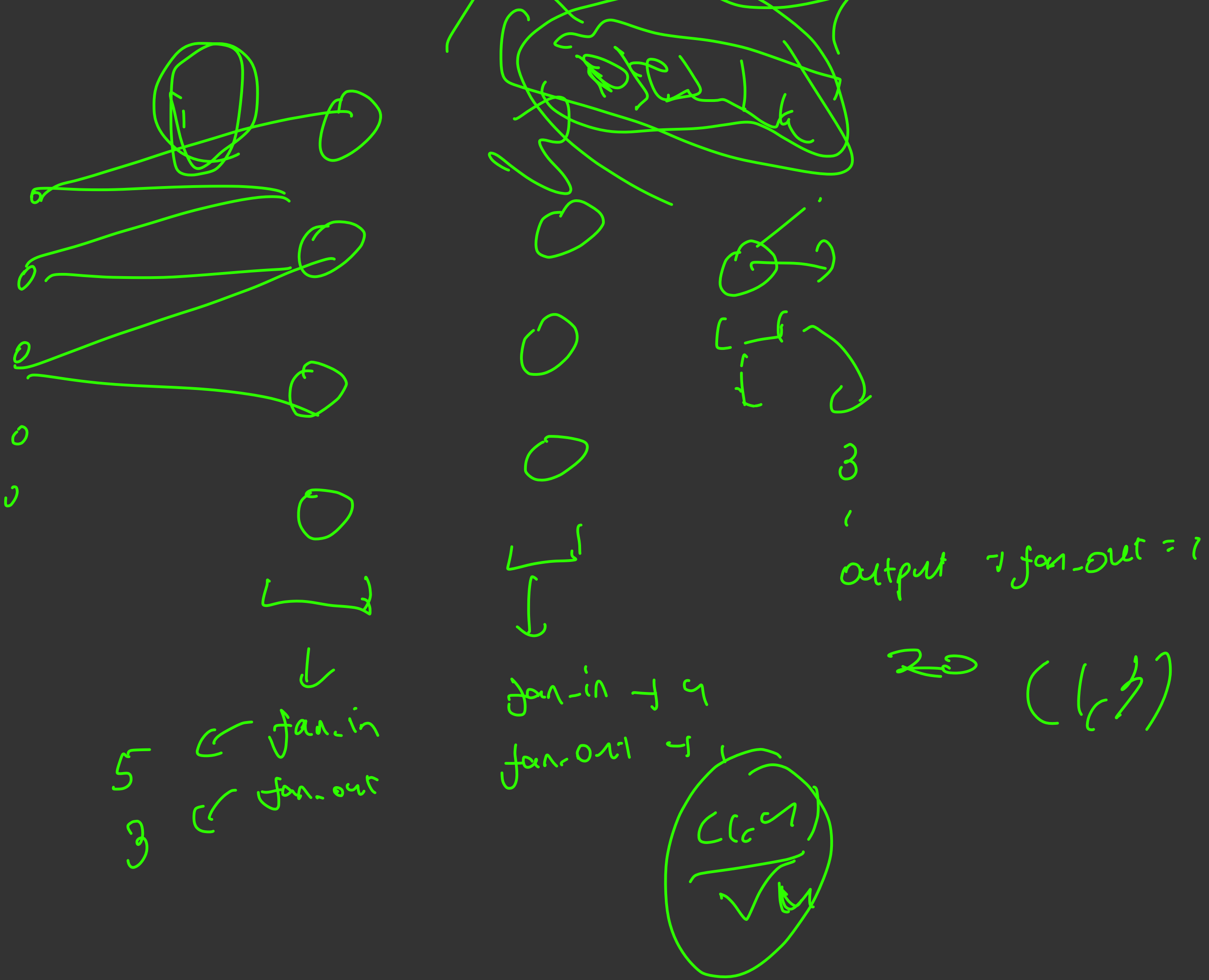
2) He

Calibrating the variances with 1/sqrt(fan_in)

$W = np.random.randn(fan\_in, fan\_out)/np.sqrt(fan\_in)$

Reasonable initialization.
(Mathematical derivation assumes linear activations)

$w = \left(np.random(fan\_in, fan\_out)\right) x$

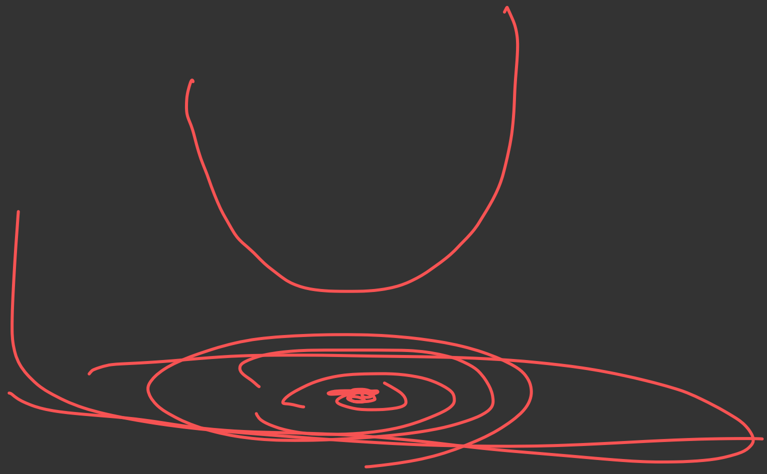


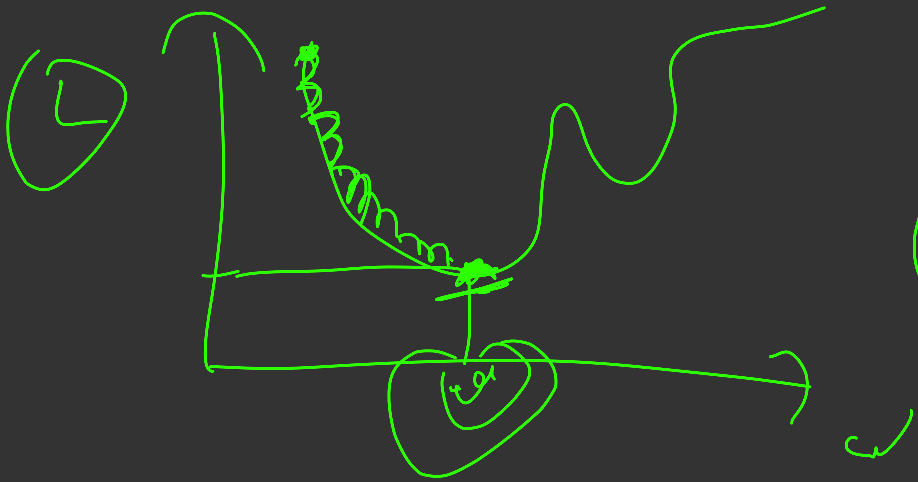
Xavier :

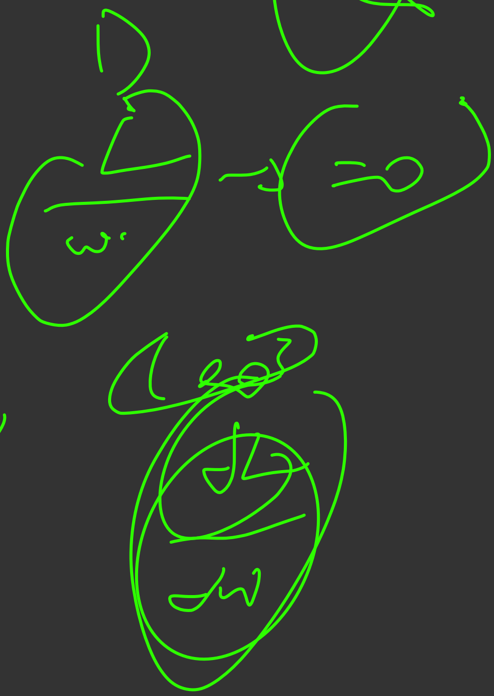
$$W = \frac{np.random(fan\_in, fan\_out)}{\sqrt{fan\_in}}$$

$$\text{Xavier Improved} = \text{Xavier} \cdot \sqrt{2}$$

$$\text{He} \rightarrow \text{Rand}() \cdot \sqrt{\frac{2}{fan\_in + fan\_out}}$$





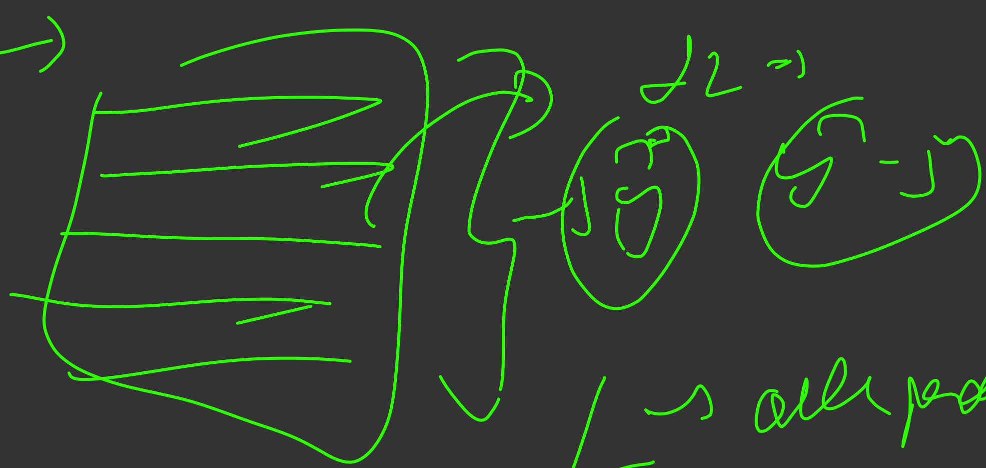


epoch

s

$$w = w - \eta \frac{dL}{dw}$$

Batch →

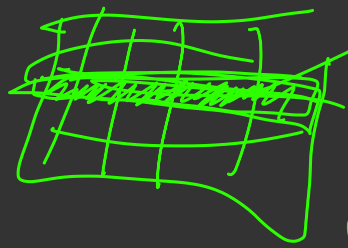


↳ all parallel

$\frac{dL}{dw}$

$\sum (y - \hat{y})^2$

Batch, Mini, SGD



10,000

1000

$\hat{y} + 1000$

for

for epoch

1000

100

$w_n = w_o - \eta \frac{dL}{dw}$

B

optional

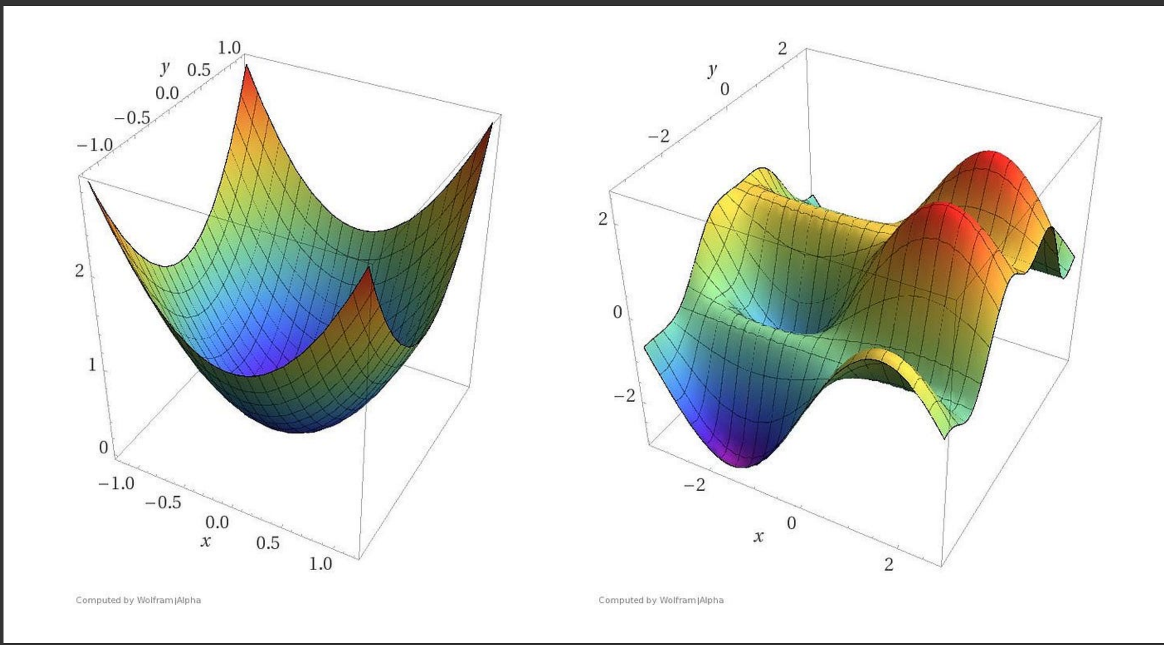
compute

MB

SGD

opt

for



$w - 10$

dw

$w = \frac{dL}{dw} = 0$

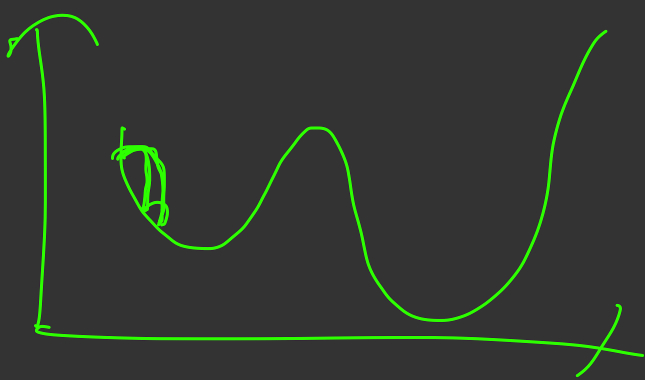
$b = 0$

$w = w - \frac{dL}{dw}$

$b = b - \frac{dL}{db}$

1) local minima

2) saddle point



Optimizers

1) SGD + momentum
2) Adagrad
3) RMSprop
4) Adam

SGD + MOMENTUM

Normal GD : $w_t = w_t - \eta_i \frac{dL}{dw_t}$

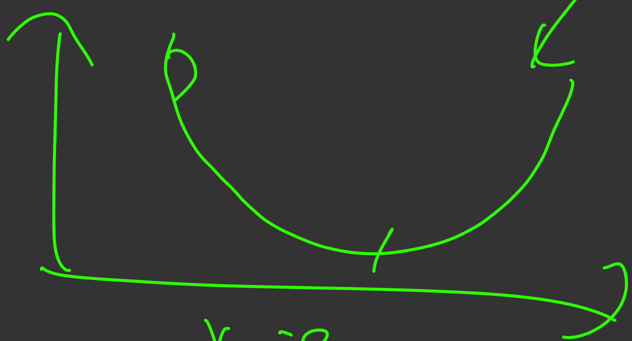
SGD + Momentum :

$$\boxed{\begin{aligned} w_{t+1} &= w_t - \alpha v_{t+1} \\ v_{t+1} &= \rho v_t + \frac{dL}{dw_t} \end{aligned}}$$

bias correction term

SGD+Momentum

$$v_{t+1} = \rho v_t + \nabla f(x_t)$$
$$x_{t+1} = x_t - \alpha v_{t+1}$$



$v_0 = 0$

$w_t = w_t - \frac{dL}{dw} - \alpha \cdot \rho \cdot v_{t-1}$

$w_{t+1} = w_{t+1} - \frac{dL}{dw} - \alpha \cdot \rho \cdot v_t$

$$w_{t+1} = w_t - \alpha V_t$$

$$V_t = \beta V_{t-1} + \frac{dL}{dw}$$

$V_0 = 0$ $\quad V_t = \frac{dL}{dw}$

$$w_{t+1} = w_t - \alpha \frac{dL}{dw}$$

1 epoch

$$V_t = \frac{dL}{dw}$$

$$w_{t+1} = w_t - \alpha \cdot \frac{dL}{dw}$$

2 epoch

$$w_{t+2} = w_{t+1} - \alpha \cdot V_{t+1}$$

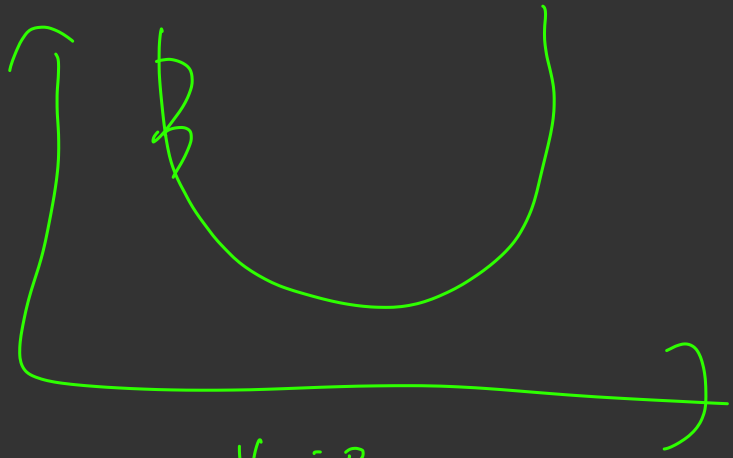
$$V_{t+1} = \beta V_t + \frac{dL}{dw_{t+1}}$$

$$V_{t+1} = \beta \cdot \underset{\downarrow\ -ve}{\frac{dL}{dw_t}} + \underset{\downarrow}{\frac{dL}{dw_{t+1}}}$$

2) ADAgrad

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{V_t + \epsilon}} \cdot \frac{dL}{dw}$$

ϵ = 1e-7 $\quad 10^{-7}$



$v_0 = 0$

$$v_t = v_{t-1} + \left(\frac{dL}{dw}\right)^2$$

$$v_{t+1} = v_t + \left(\frac{dL}{dw_{t+1}}\right)^2$$

$$w = w - \alpha \frac{dL}{dw}$$

$$b = b - \alpha \frac{dL}{db}$$

Step I : $\frac{dL}{dw}$

Step II = $v_t =$

$v \quad b \downarrow \quad \downarrow$

$$\frac{\alpha}{\sqrt{v_t + \epsilon}}$$

$$v_t = \left(\frac{dL}{dw}\right)^2 \qquad \frac{\alpha}{\sqrt{v_t+\epsilon}} \quad (0.01)$$

$$w = w - \alpha \frac{dL}{dw} \qquad 0.1 \qquad v_t + (0.1)^2$$

$$b = b - \alpha \frac{dL}{db} \rightarrow 1 \qquad \cancel{\sqrt{v_b+\epsilon}} \quad v_b + (1)^2$$

1

$\frac{dL}{dw}$ = compute_gradient (w)

grad_squared += $\left(\frac{dL}{dw}\right)^2$

$$\alpha_{new} = \frac{\alpha}{\sqrt{v_t + \epsilon}}$$

$$= \frac{\alpha}{\sqrt{v_t} + \epsilon}$$

optim

3) RMSprop

$w_t - v_t$

$$\frac{1}{\sqrt{v_t}}$$

$$v_t \mathrel{+}= \left(\frac{dL}{dw}\right)^2$$

0.00001

$$v_t = \beta\, v_{t-1} + (1-\beta)\left(\frac{dL}{dw}\right)^2$$

.05

$\beta = 0.95$

0.95 + (0.05)

0.05 + 0.01

RMSprop + momentum ⇒ Adam

v_t

$$-\frac{\alpha}{\sqrt{s}} \cdot v_t$$

$$w_{t+1} = w_t - \frac{\alpha \cdot v_{t_1}}{\sqrt{v_{t_2} + \epsilon}}$$

$$v_{t_1} = \beta_1 \cdot v_{t_1 - 1} + (1-\beta_1) \cdot \frac{\partial L}{\partial w}$$

$$v_{t_2} = \beta_2 v_{t_2 - 1} + (1-\beta_2)\left(\frac{\partial L}{\partial w}\right)^2$$

$v_M \rightarrow$

$v_R \rightarrow$

$v_{M_0} = 0$

$v_{R_0} = 0$

$$v_{M_t} = \beta_1 v_{M_{t-1}} + (1-\beta_1) \frac{\partial L}{\partial w}$$

$$v_{R_t} = \beta_2 v_{R_{t-1}} + (1-\beta_2)\left(\frac{\partial L}{\partial w}\right)^2$$

$$\text{Adam} = \quad w_{t+1} = w_t - \frac{\alpha \cdot v_{M_t}}{\sqrt{v_{R_t}} + \epsilon}$$

$$V_{mB} = \frac{X_{mt}}{1-\beta_1^{epoch}}$$

$t =$ epochs

$$V_{RB} = \frac{V_{Rt}}{1-\beta_2^{epoch}}$$

$$\text{Adam} \quad w_{t+1} = w_t - \frac{\alpha \cdot V_{mB}}{\sqrt{V_{RB}} + \epsilon}$$

$x = [1, 1, 1, 1]$

$w_{if} = [1, 0, 0, 0]$?

$w = [0.25, 0.25, 0.25, 0.25]$?

$w^T x + b$

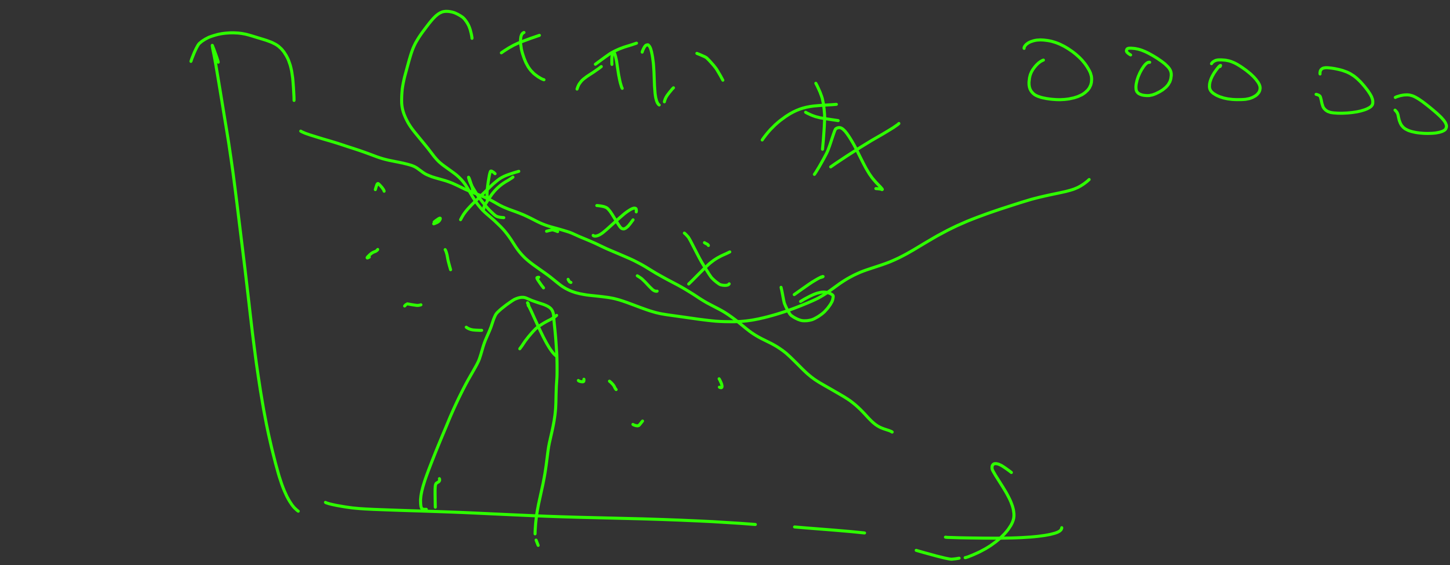
$[0 \; 0 \; 0] + b$

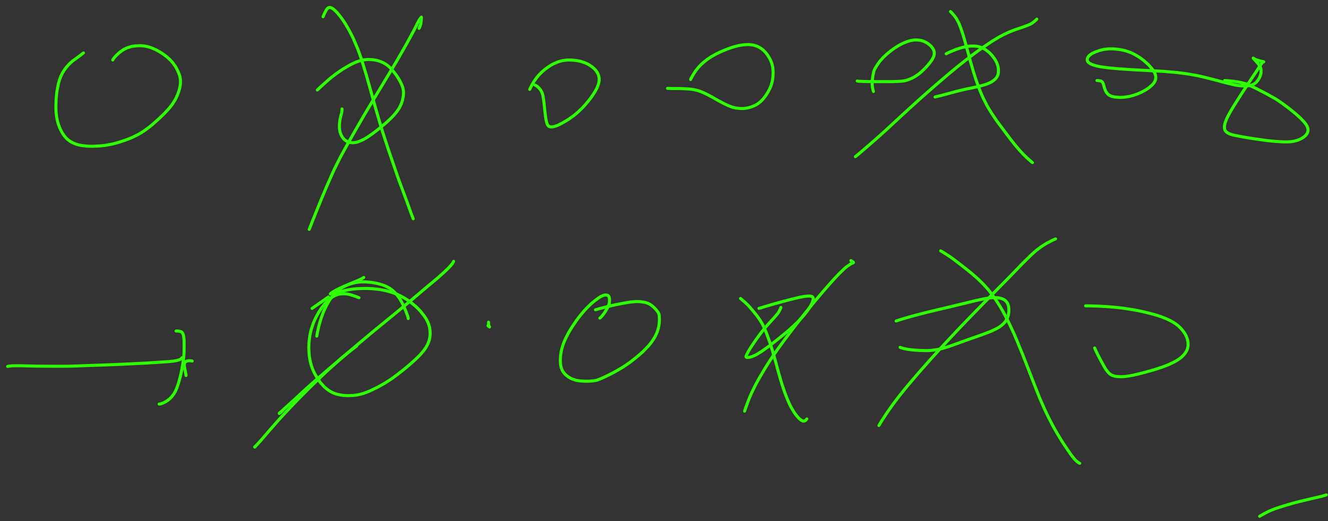
$[0.25 \; 0.5 \; 0.5 \; 0.2 \; \ldots] + b$

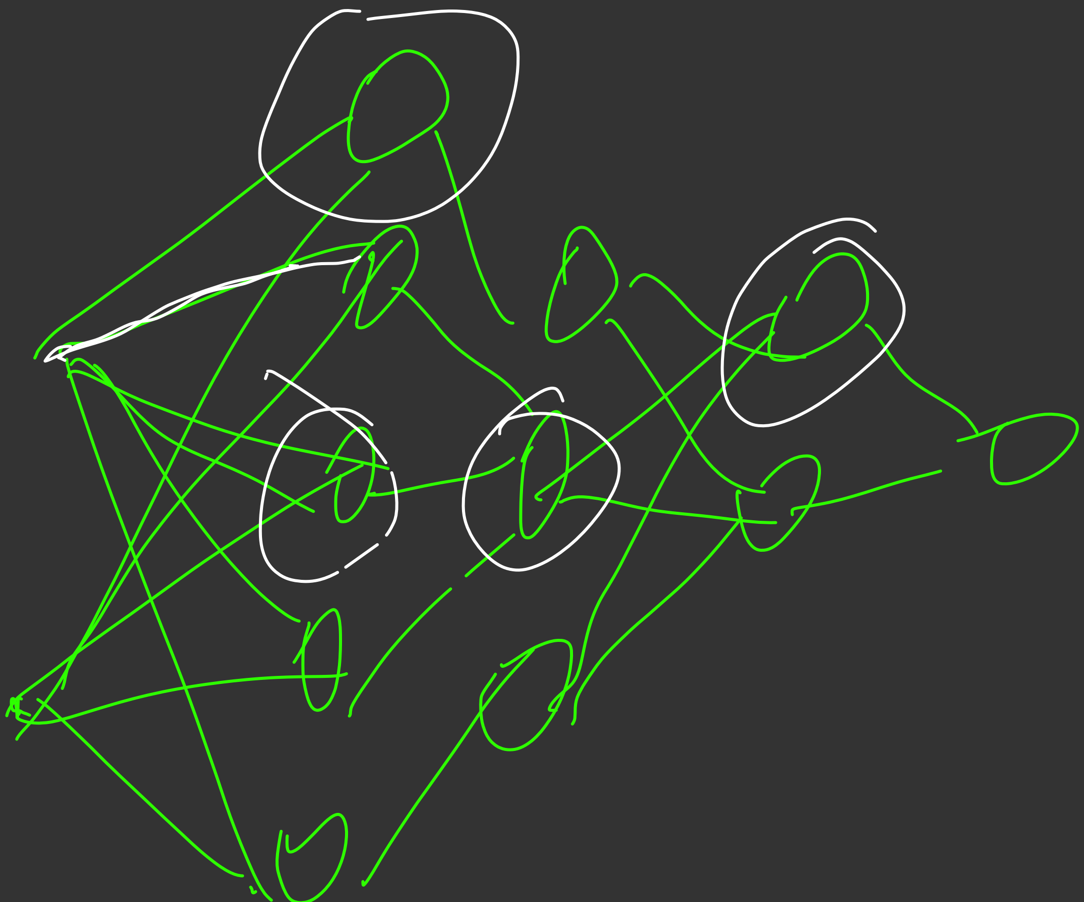
$1 - 1.5$

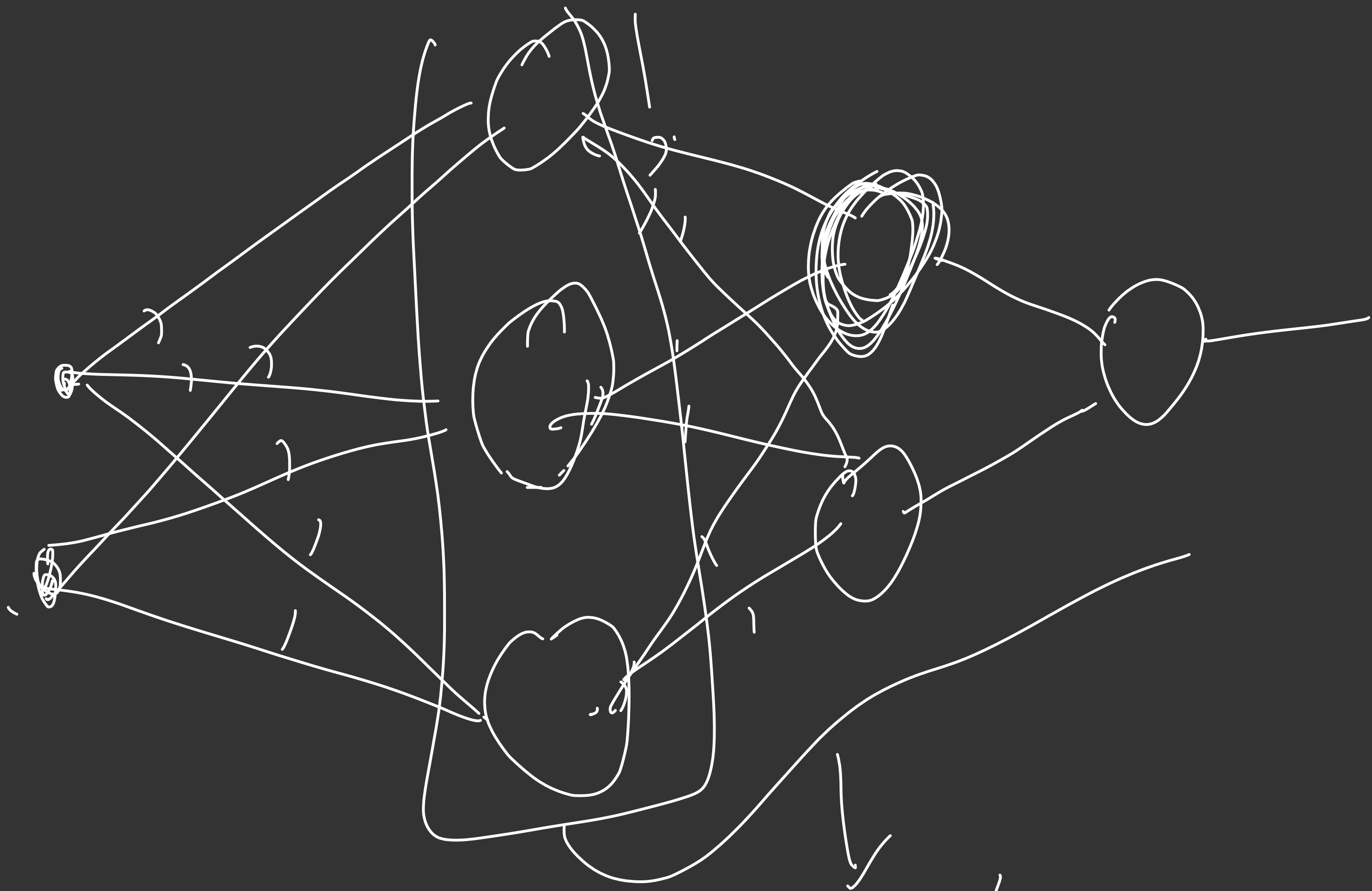
$+b$

1) Dropout
2) Dropconnect
3) Data Augmentation
4) Batch Normalisation





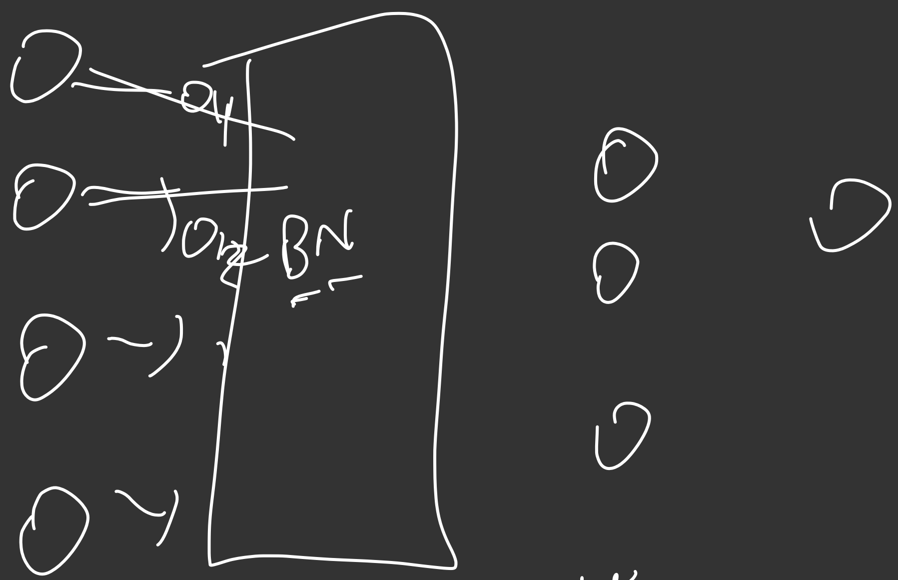




ReLU

$x \to x$

$(wx+b) \to (0,1)$



$O_1 \quad O_2 \quad O_3 \quad O_4$

mean

variance

max = mean + std

min = mea

$$O_1 = \text{for } i: \quad \frac{O_i - \text{mean}}{\sqrt{\text{var}}}$$

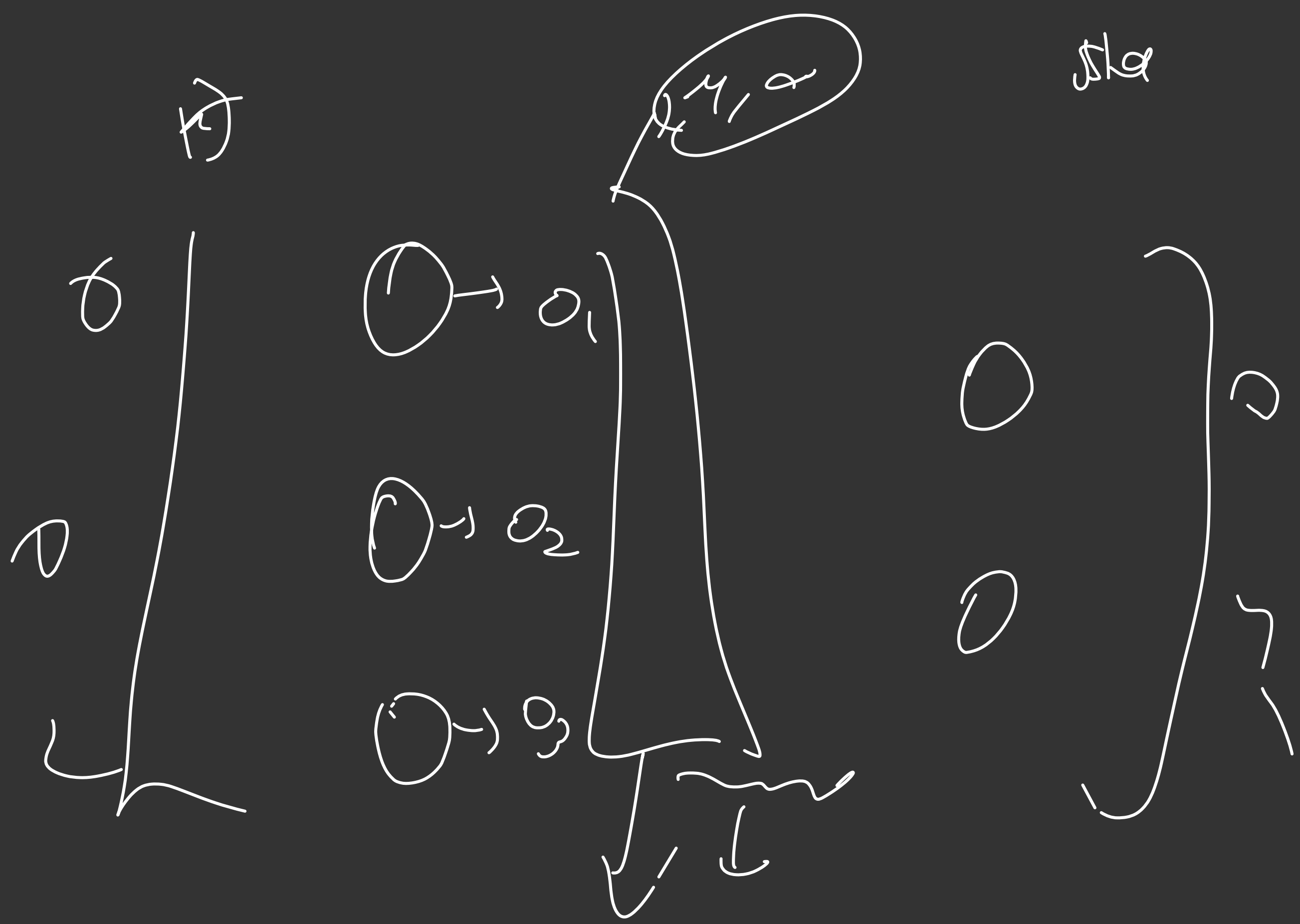
$$O_i^* = \frac{O_i - \text{mean}}{\text{std}} \rightarrow \boxed{-1, 1}$$

(O_i)

$\boxed{\gamma O_i + \beta}$

mean

std



Step I) find mean & std of

O_1, O_2 & O_3

Step II) $O_i' = \dfrac{O_i - \text{mean}}{\text{std}}$

Step III) $O_i'' = \gamma O_i' + \beta$

$\gamma := 1$

$\beta := 0$

step

$\frac{X - \text{mean}}{\text{std}}$ t X. 0

mean $\to 0$

$[-1, 1]$

std $\nearrow$

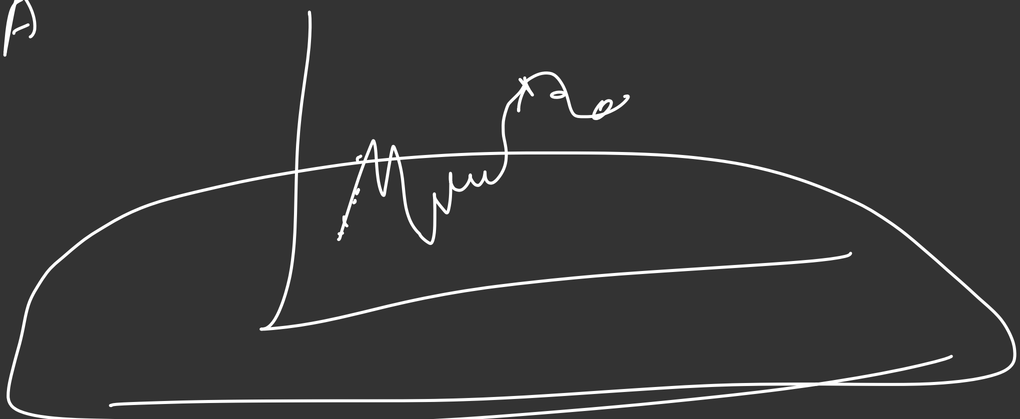
---

epoch $\to$

$\{\mu_j, \mu_{i+1}, \mu_i, \}$

$\{\sigma_1, \sigma, ---\}$

EMWA

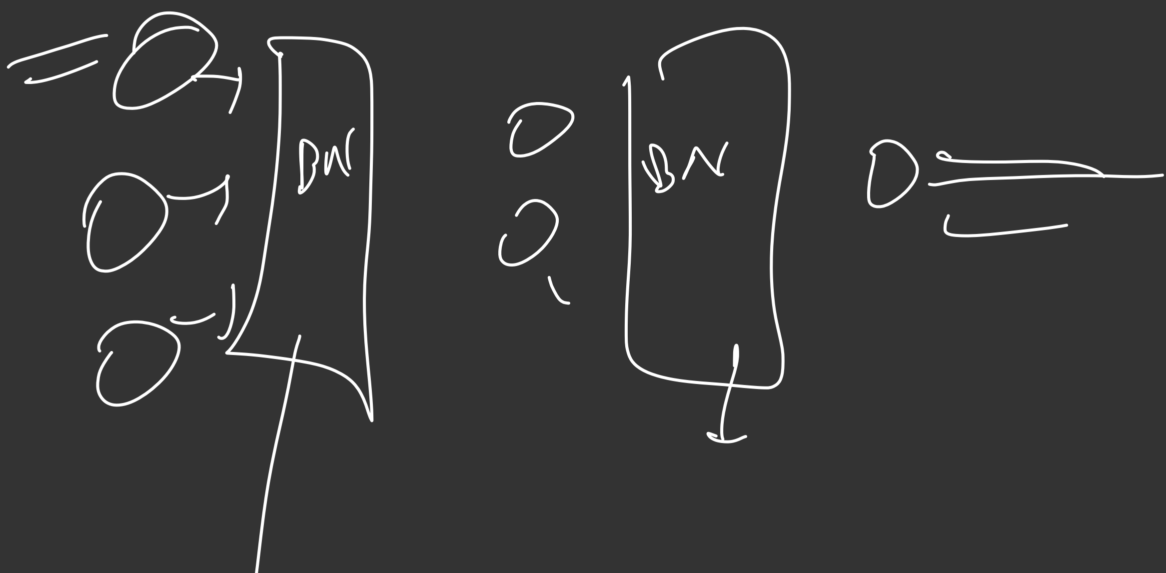


$\beta \quad \mu \to 0$

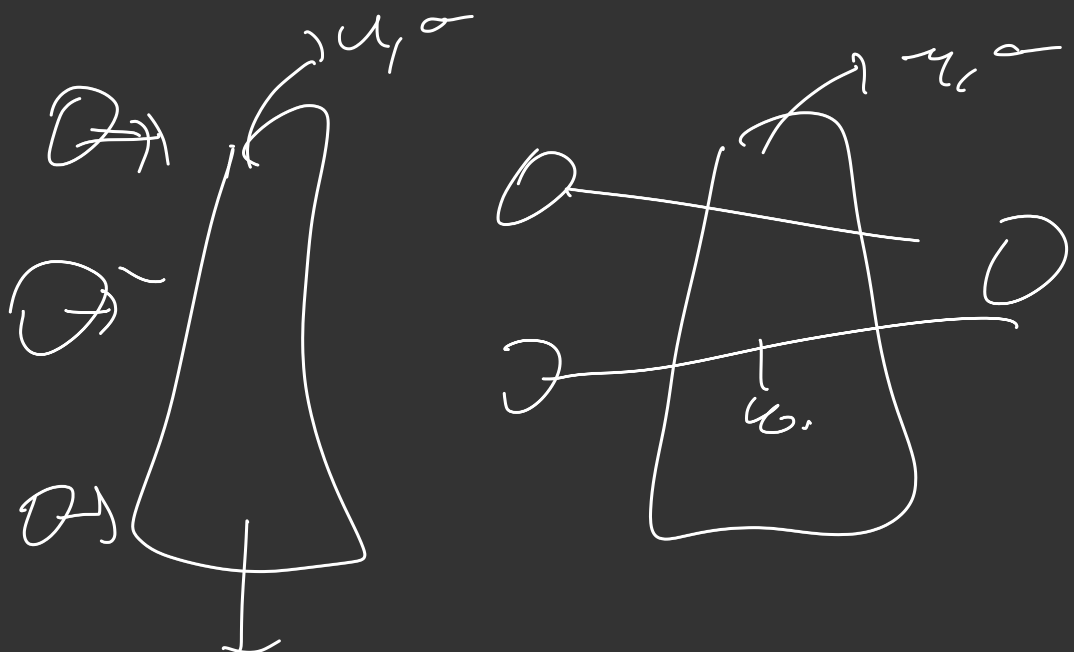
$\beta = \sigma \to 1$

$$\mu_{emma} = \beta \mu_{i-1} + (1-\beta)\mu_i$$

$$\left.\begin{matrix} O_i \to \\ O \to \\ O \to \end{matrix}\right\} \to \frac{O_i - \mu_{emma}}{\sigma_{emma}}$$



Test



epoch

μ_1, μ_2 $\quad$ μ_1, μ_2

σ_1, σ_2 $\quad$ σ_1, σ_2

test

θ_1 —

θ_2 —

θ_3 —

4

EMWA → exponentially weighted moving avg

μ_1 μ_2

σ_1 σ_2

$\mu_{emwa} = \mu_1$

$\mu_{emwa} = \beta \mu_t + (1-\beta) \frac{\partial L}{\partial w}$

6

$u_{emwA} = u_1$

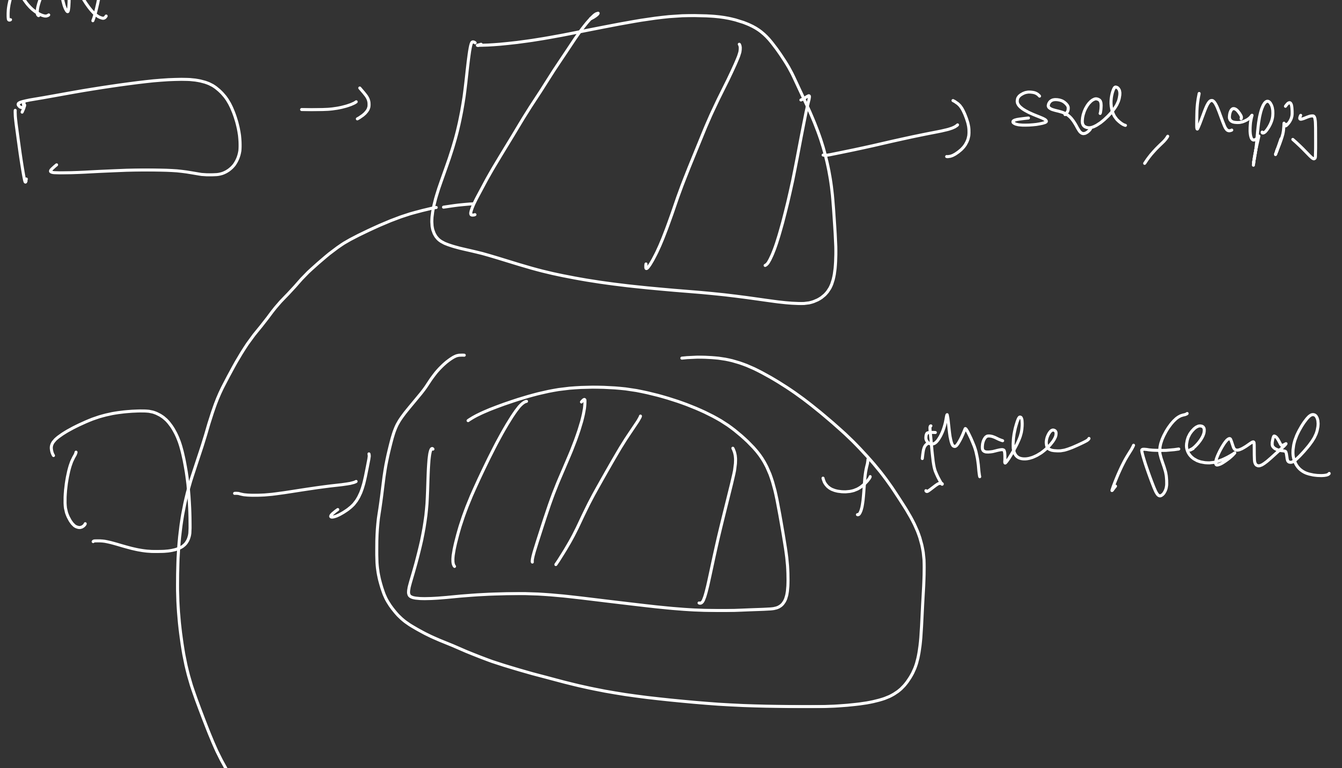
$u_{emw} = \beta u_{emwA} + (1-\beta) u_2$

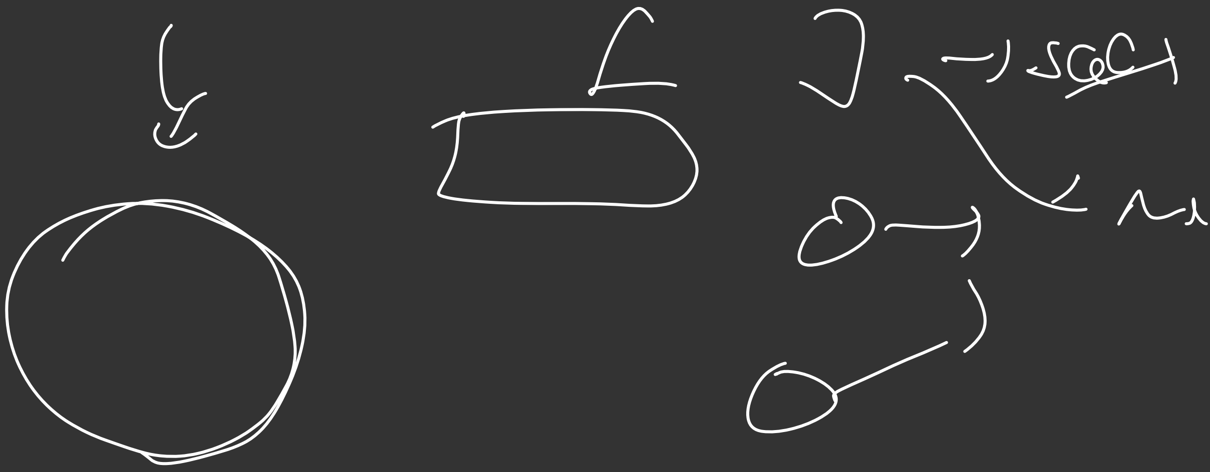
$u_{\beta mw} = \beta u_{emwA} + (1-\beta) u_3$

Memo

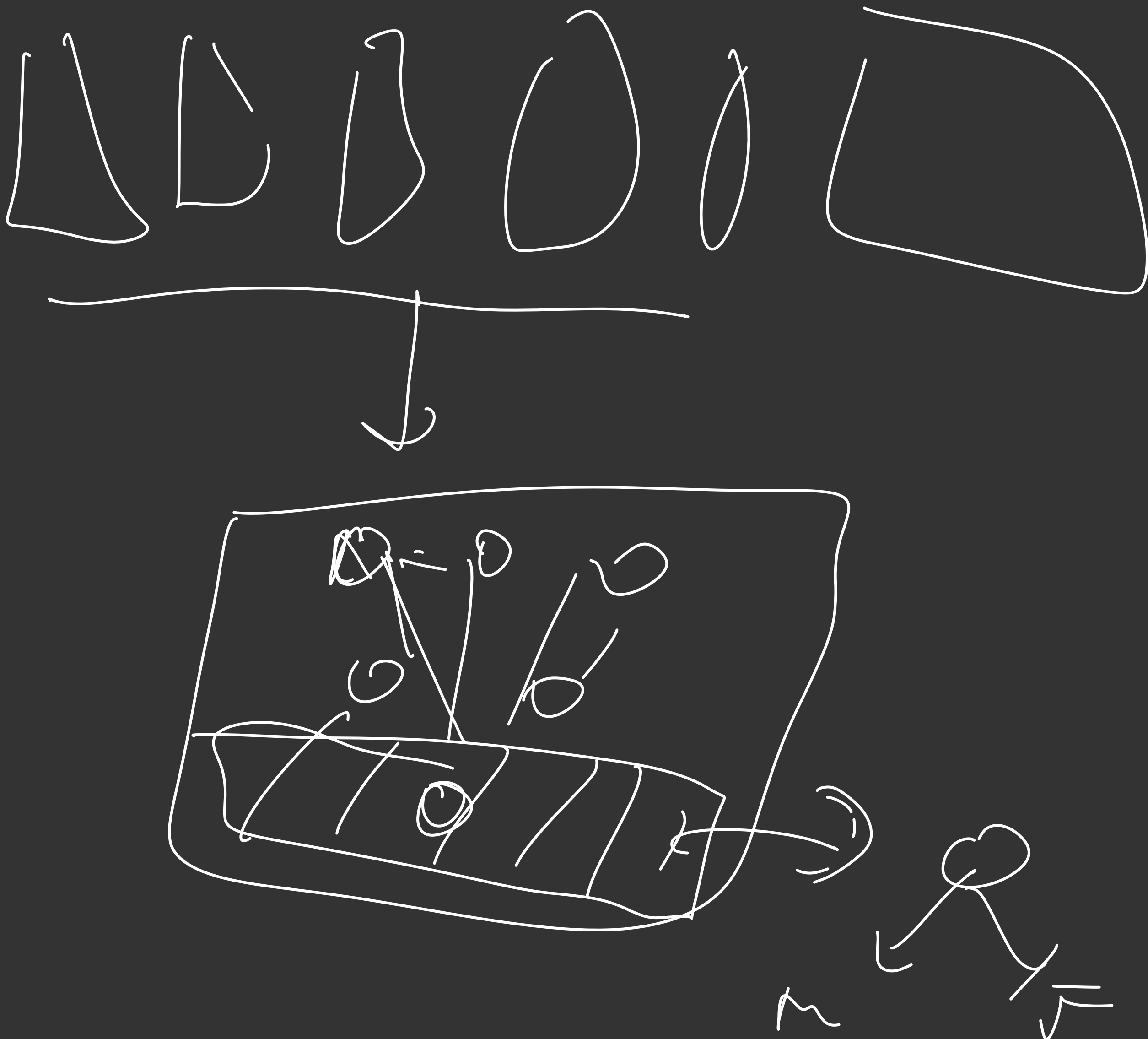
1) IA talks INTERNAL COVARIATE SHIFT

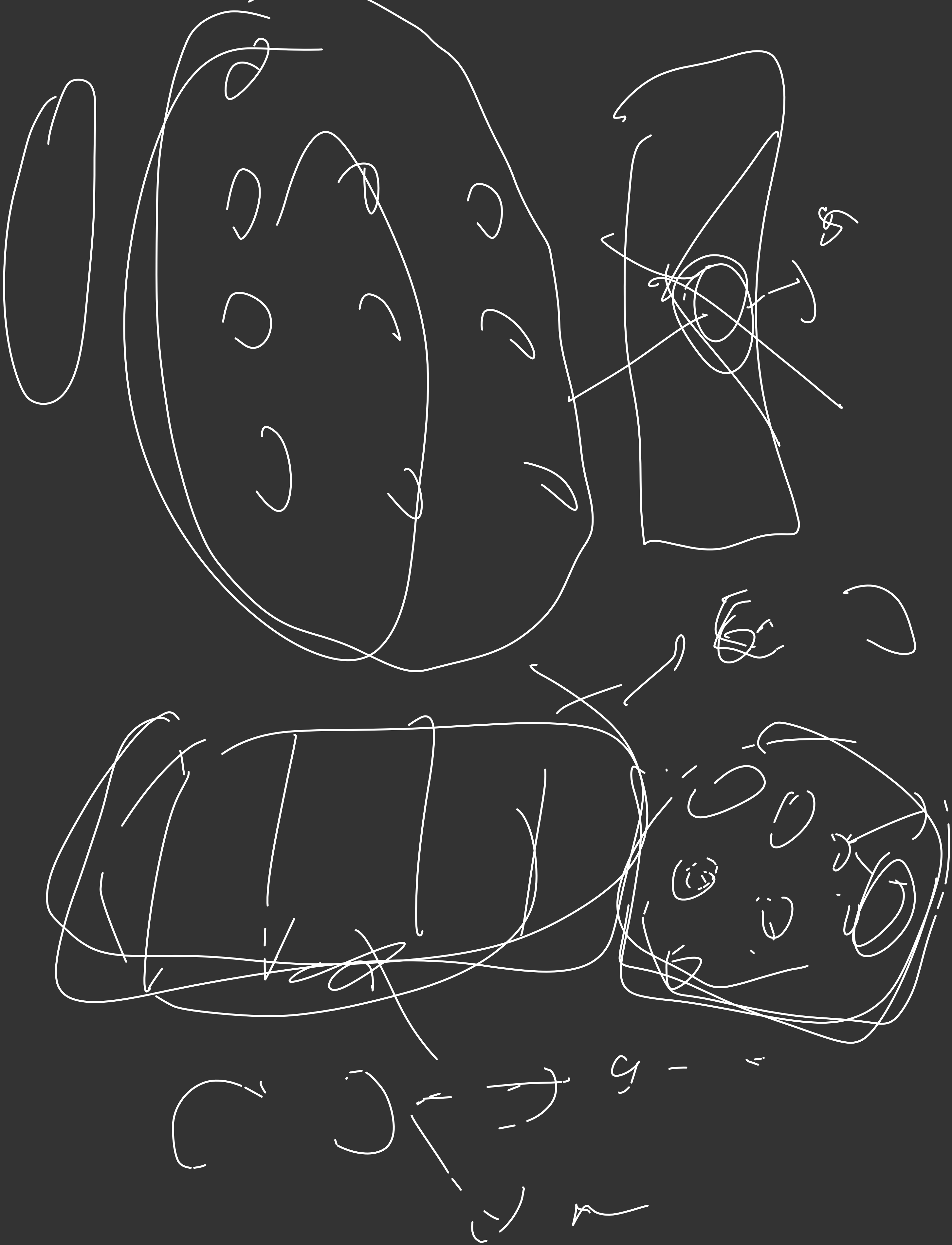
CNN

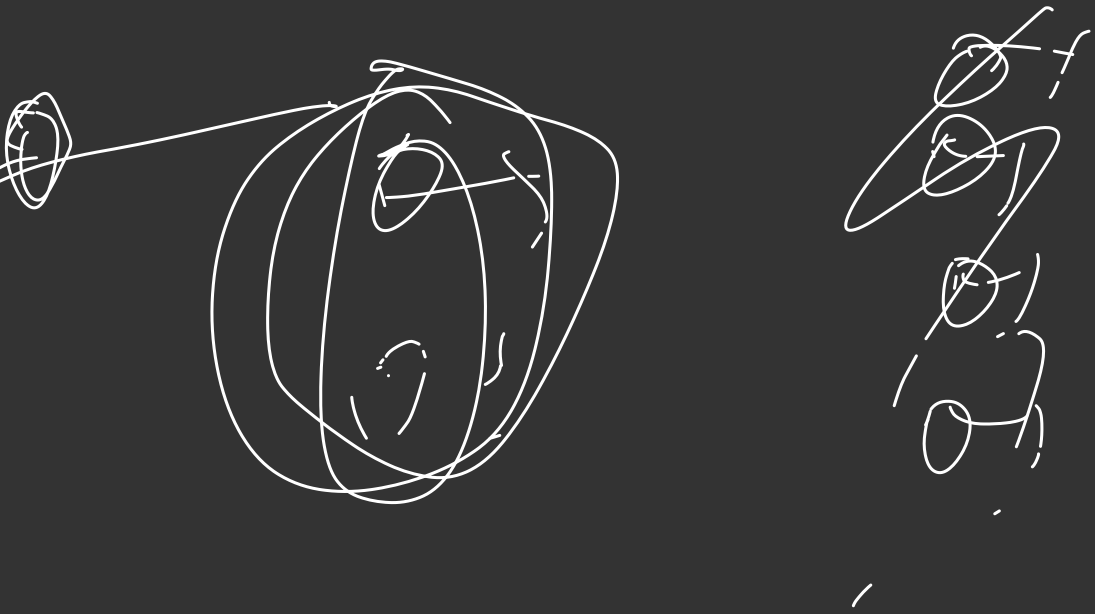


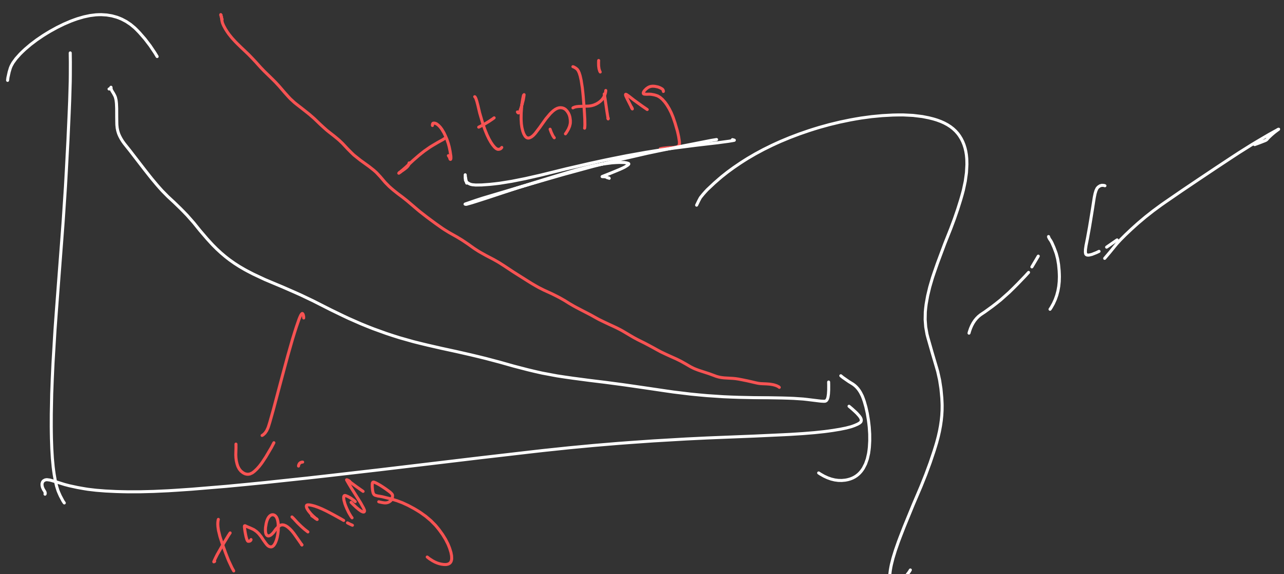


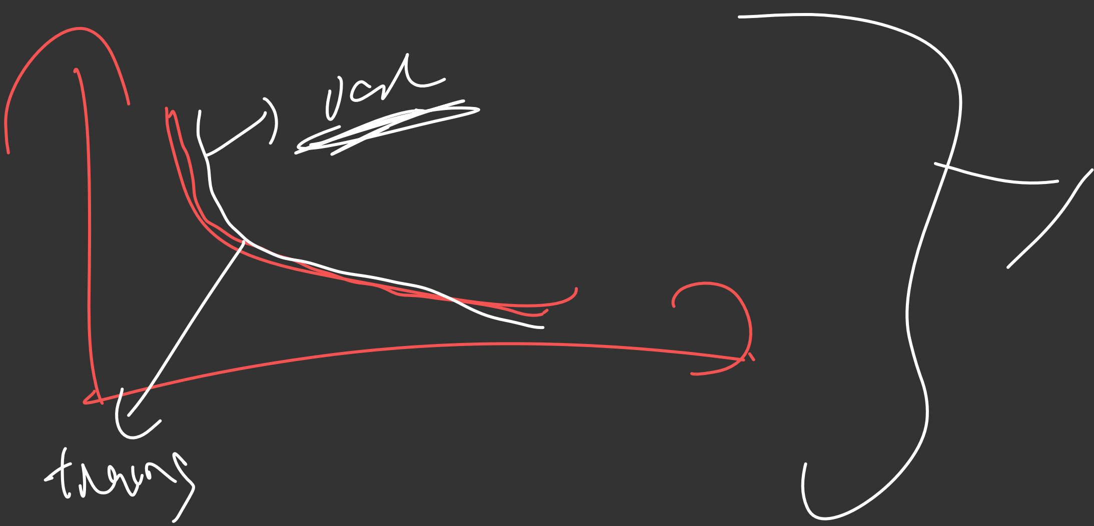

male
female

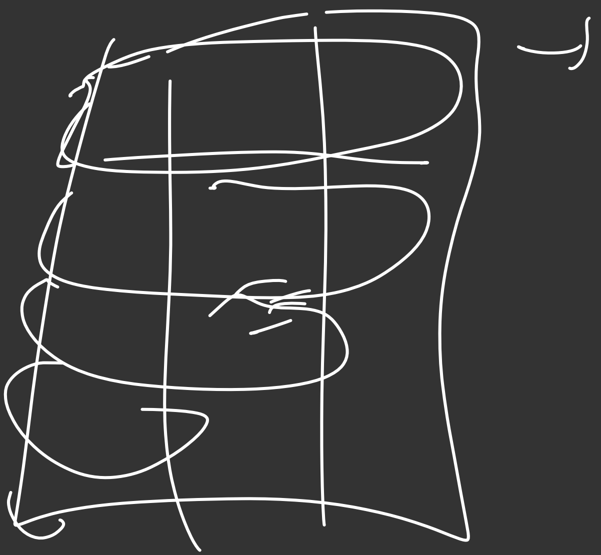

M
F






testing


val



100MB

200MB

C1

C2

{1, 3}

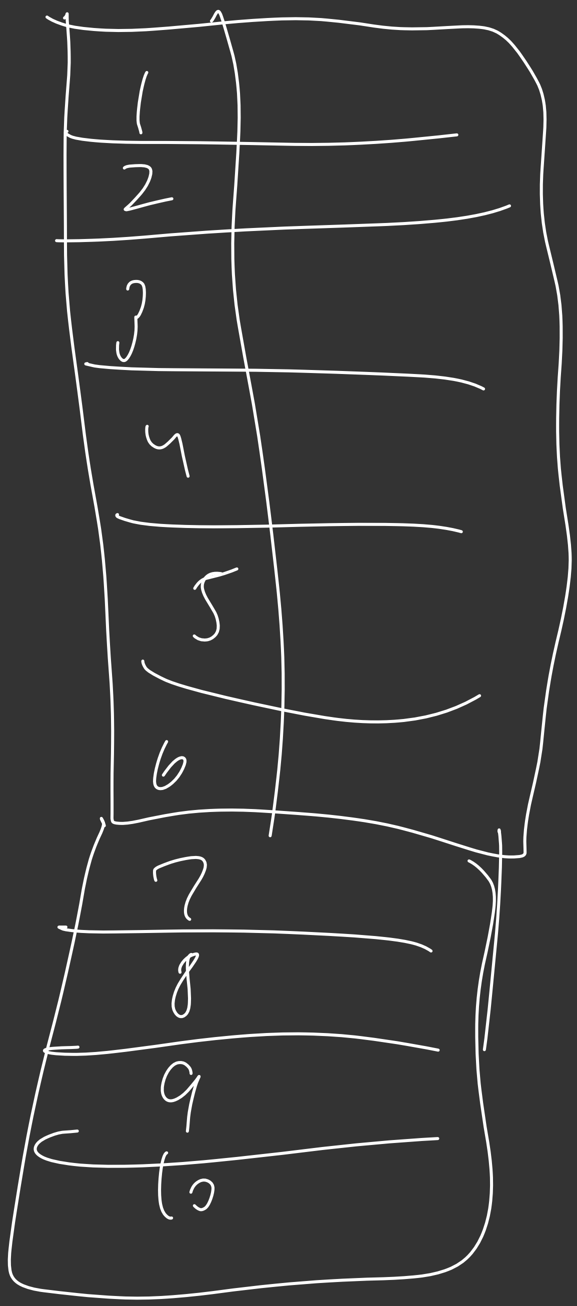
5]|0

{1, 2, 3, 4, 5}

5|0

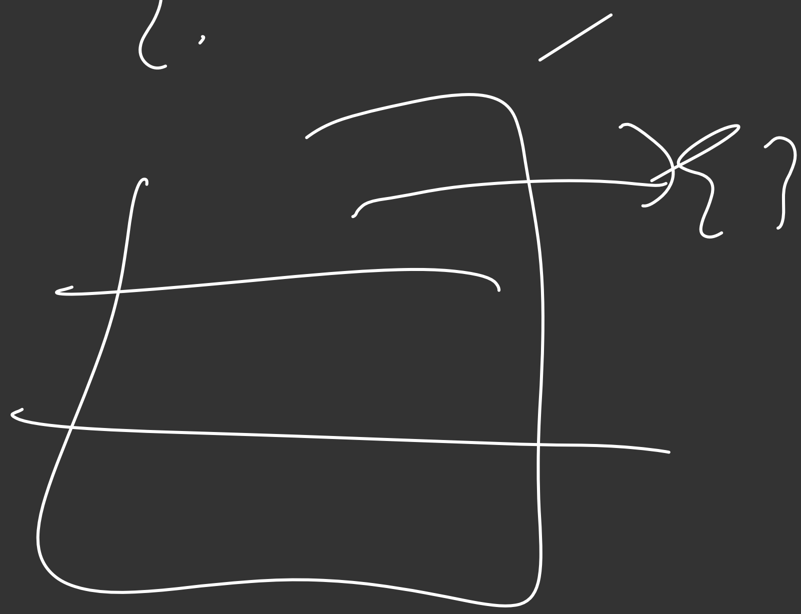
5+5 = 25]|0

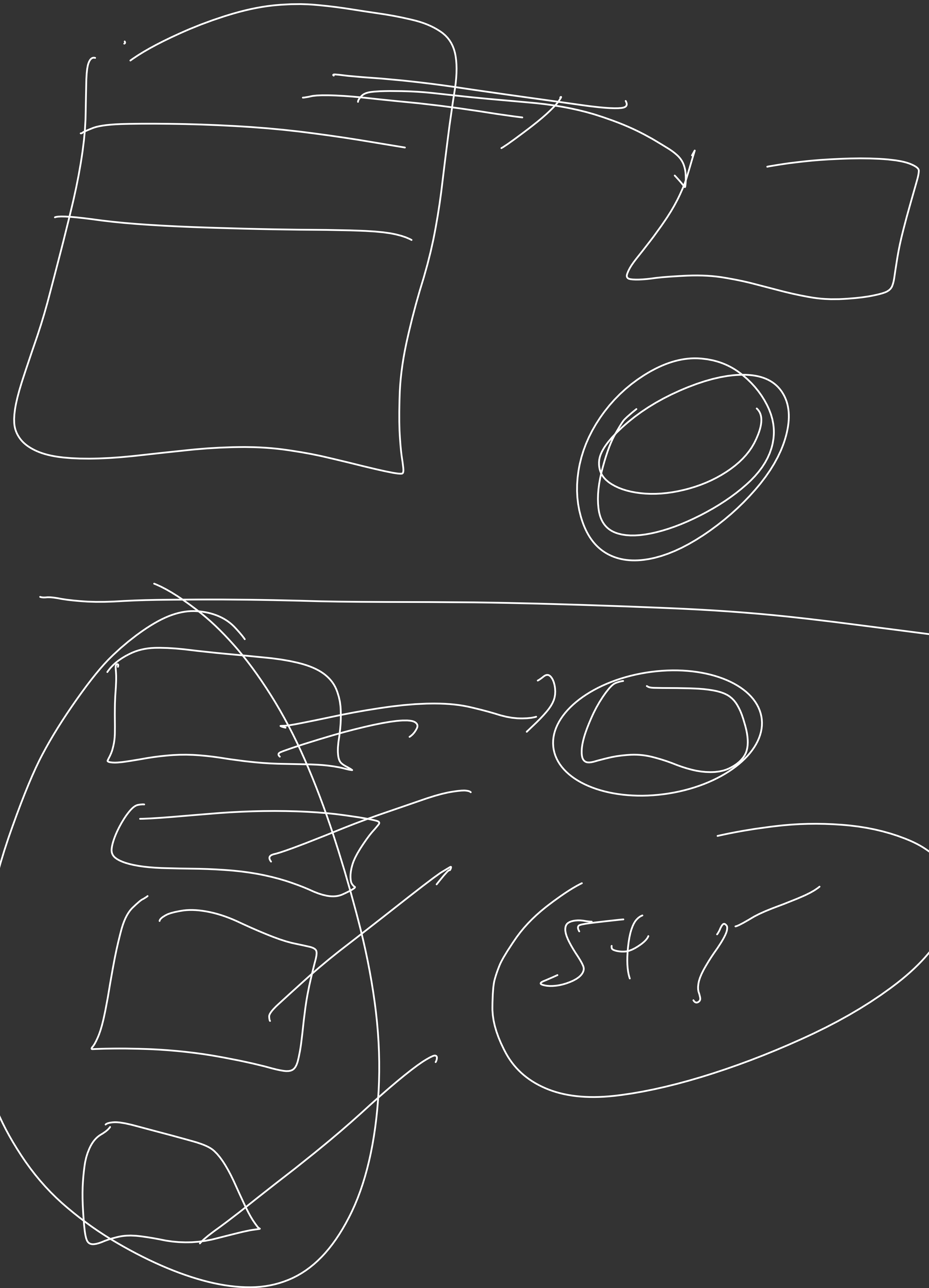
25]|0

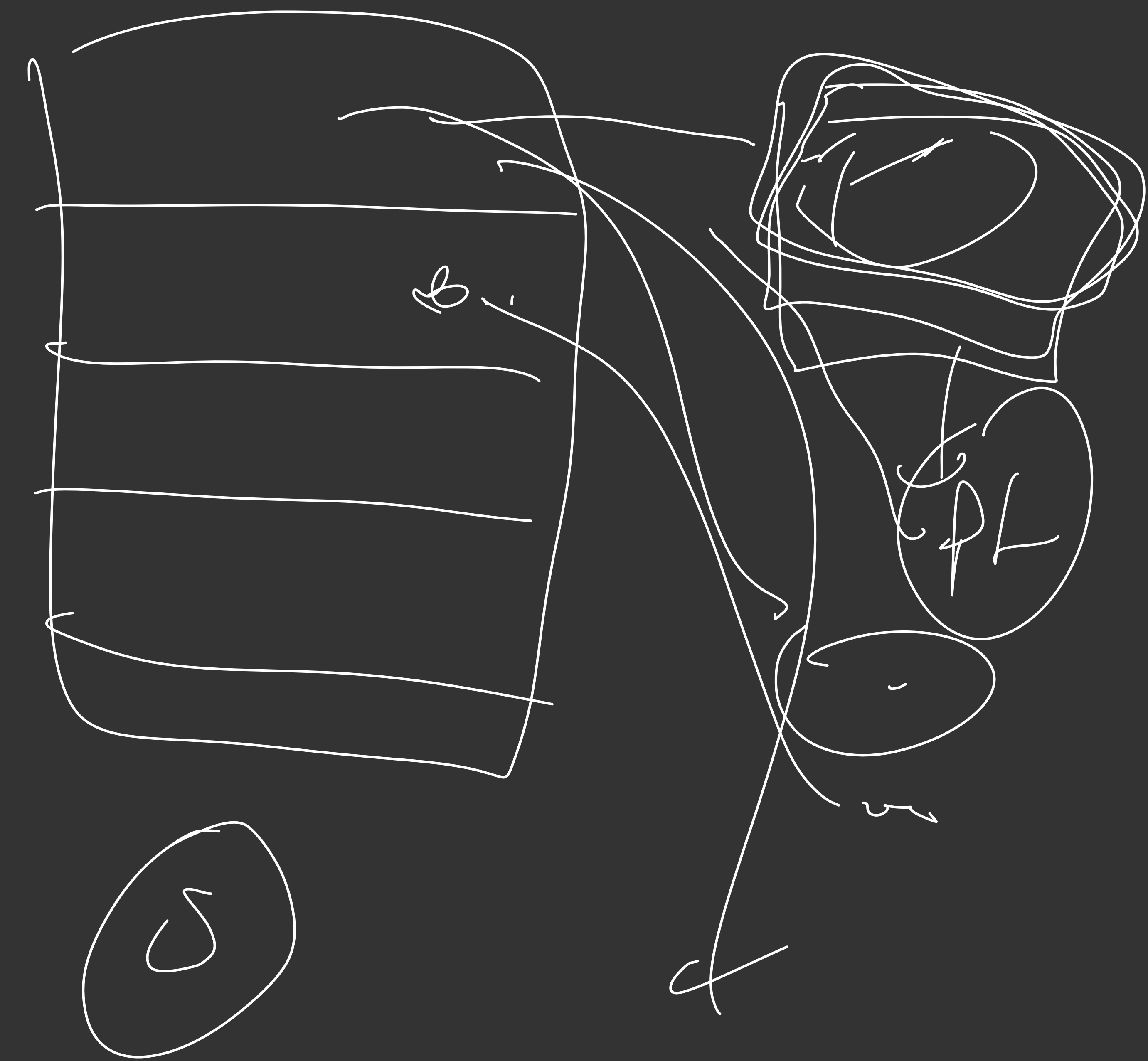


Štv

2.







$$P'_1 \cup Pr_2 \cup P$$

$$P_L = \mathcal{U}$$

12 rows of data

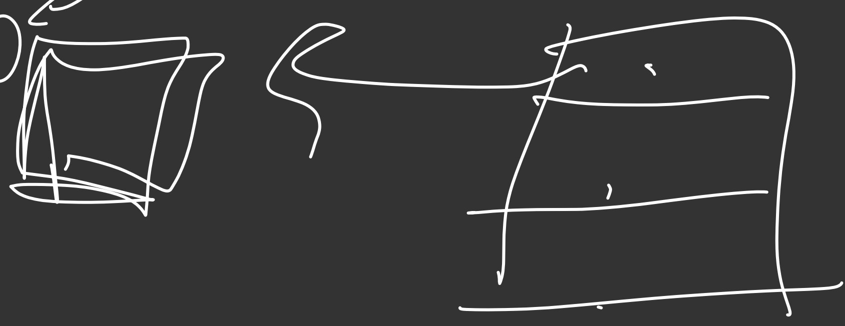
1 - of data or

12 scans

Pos

AOI

AOI



Memory → 120 GB

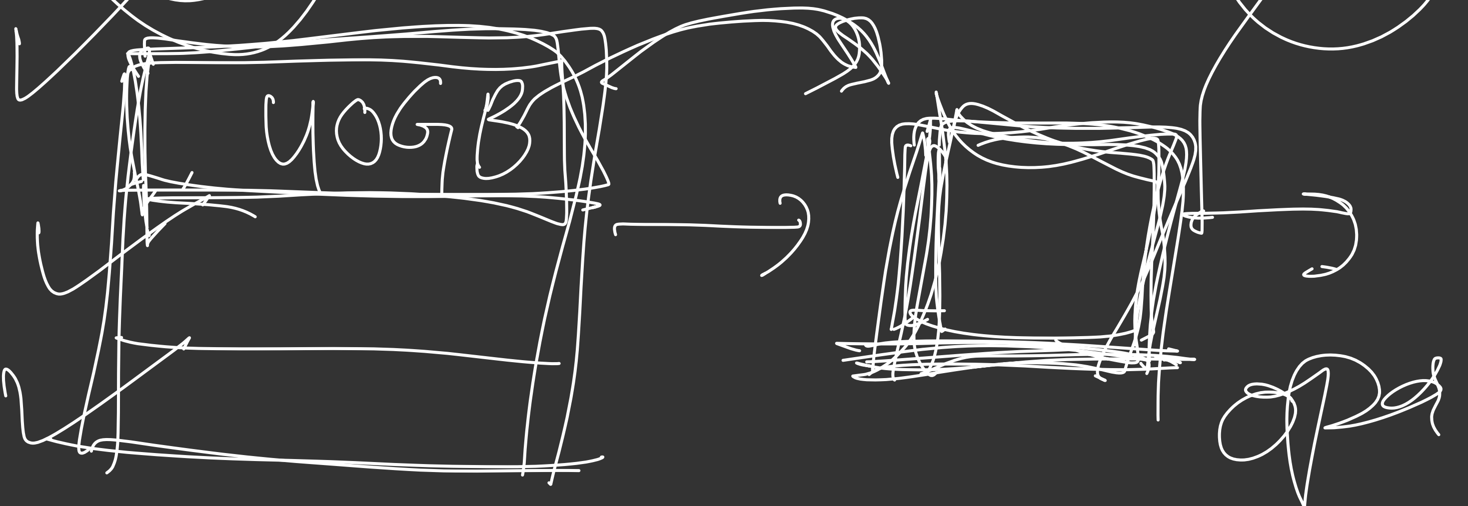
Ram → 40 GB

Apriori scans = 15

I/O → Apriori = 3 × 15

3

Partitioning = 6



5 Marks Done

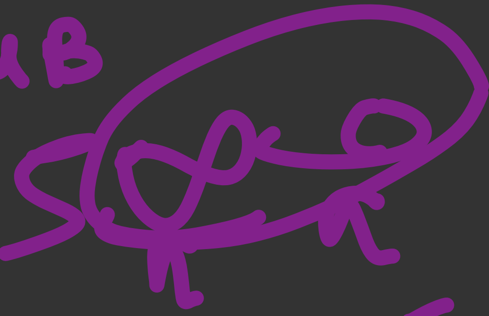
2 apm (10 secs)

1 part.

100 MB

20 MB

= 5



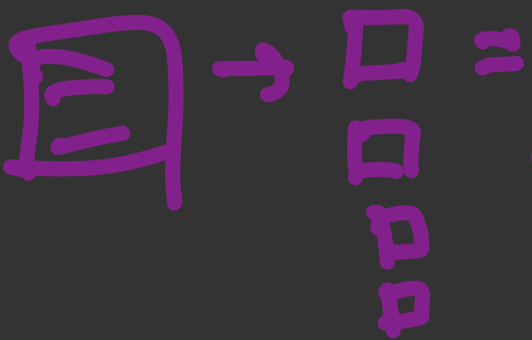
5 x 10 x 2 + 10

4

6

5

6



5x6 + 5x6

1) FP

2) Tr. Red

3) Lift Z Chi ^2

4) Data Preprocessing

5) Theory Z Lama

X - mean
/ std

Data ⊰

DA

1) FP growth

2) Transaction

3) Shaty

4) Seides