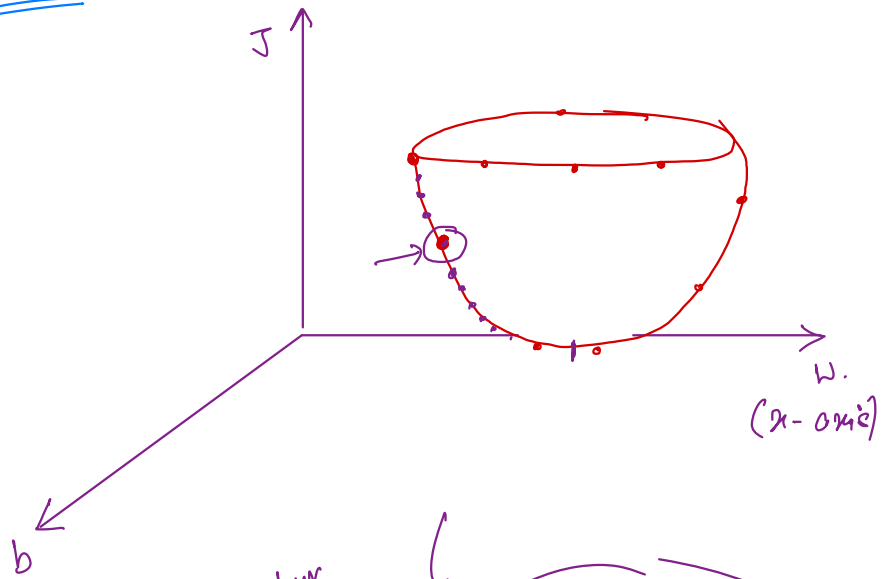
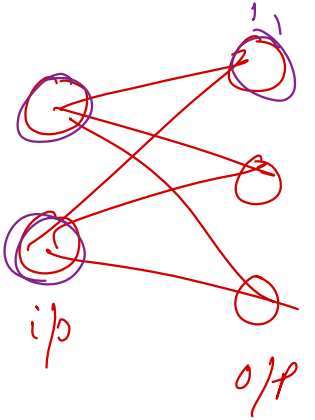
Optimizers

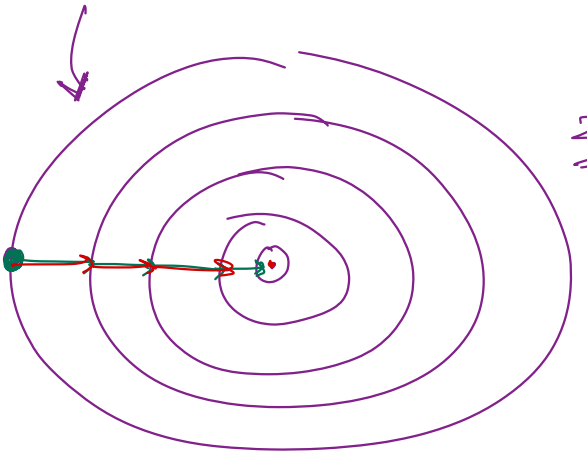


$J(w,b)$

3D



In a contour



2D

Batch gradient descent

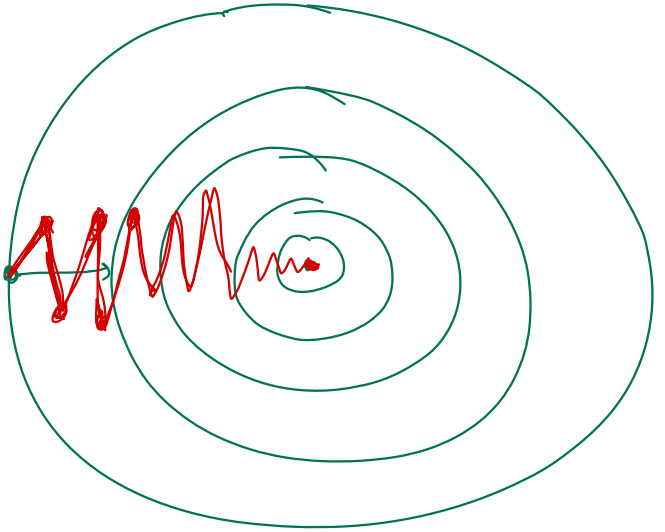
faster

Calculation

5 steps → global minima

⇒ 5 × 300 ⇒ 1500

Mini batch GD

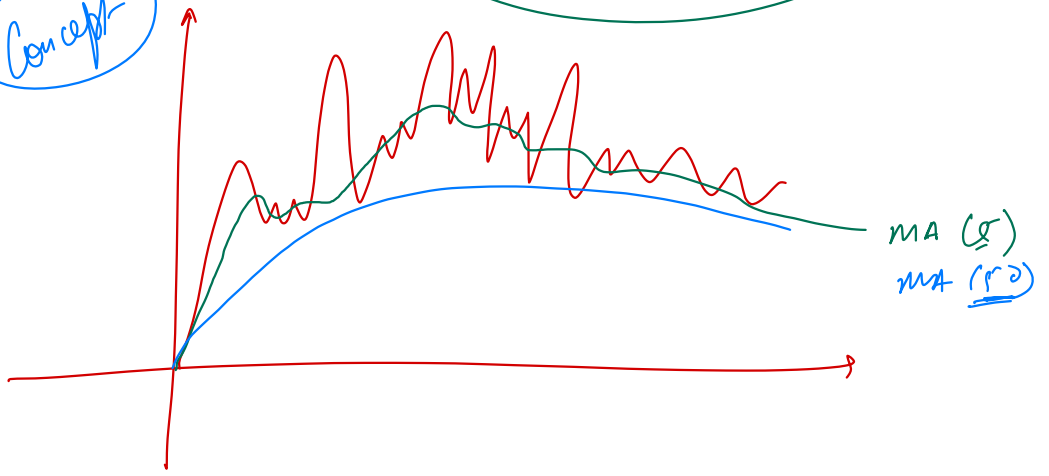


100 data points

10 steps

$\Rightarrow$ 10,00 $\Rightarrow$

Concept



Exponential weighted average

$t(1) \rightarrow 40$

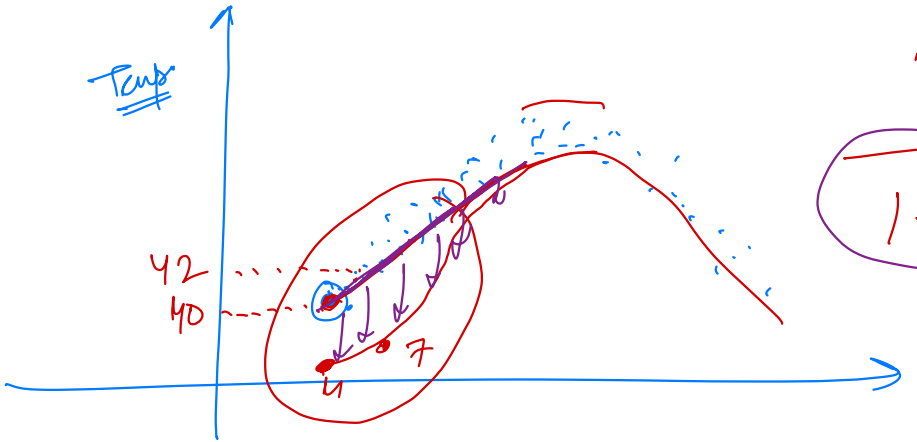
$t(2) \rightarrow 42$

$t(3) \rightarrow 50$

.

.

$t(360) \rightarrow 45$



$$\frac{V_t}{1-\beta^t} = \frac{4}{1-(0.9)^1}$$

$$\Rightarrow \frac{4}{1-0.9}$$

$$\Rightarrow \frac{4}{0.1}$$

$$\Rightarrow 40$$

gd with momentum

$V_0 = 0$

$V_1 = (0.9)\,V_0 + (0.1)\,t_1$

$\Rightarrow (0.9) \times 0 + (0.1)\,40 \Rightarrow 4$

$V_2 = 0.9\,(40) + (0.1)\,42 \Rightarrow 7.4$

$= 36 + 4.2 \Rightarrow 40.2$

$$\frac{7}{1-\beta^t}$$

$$V(t) = \beta V_{t-1} + (1-\beta)\, t$$

Bias corrected

Calculation of EMA.

$$V_{dw} = \beta V_{dw} + (1-\beta)\, dw$$

$$dw = \frac{\partial J}{\partial w}$$

$$w^{new} = w^{old} - \alpha\, dw$$

wavyness

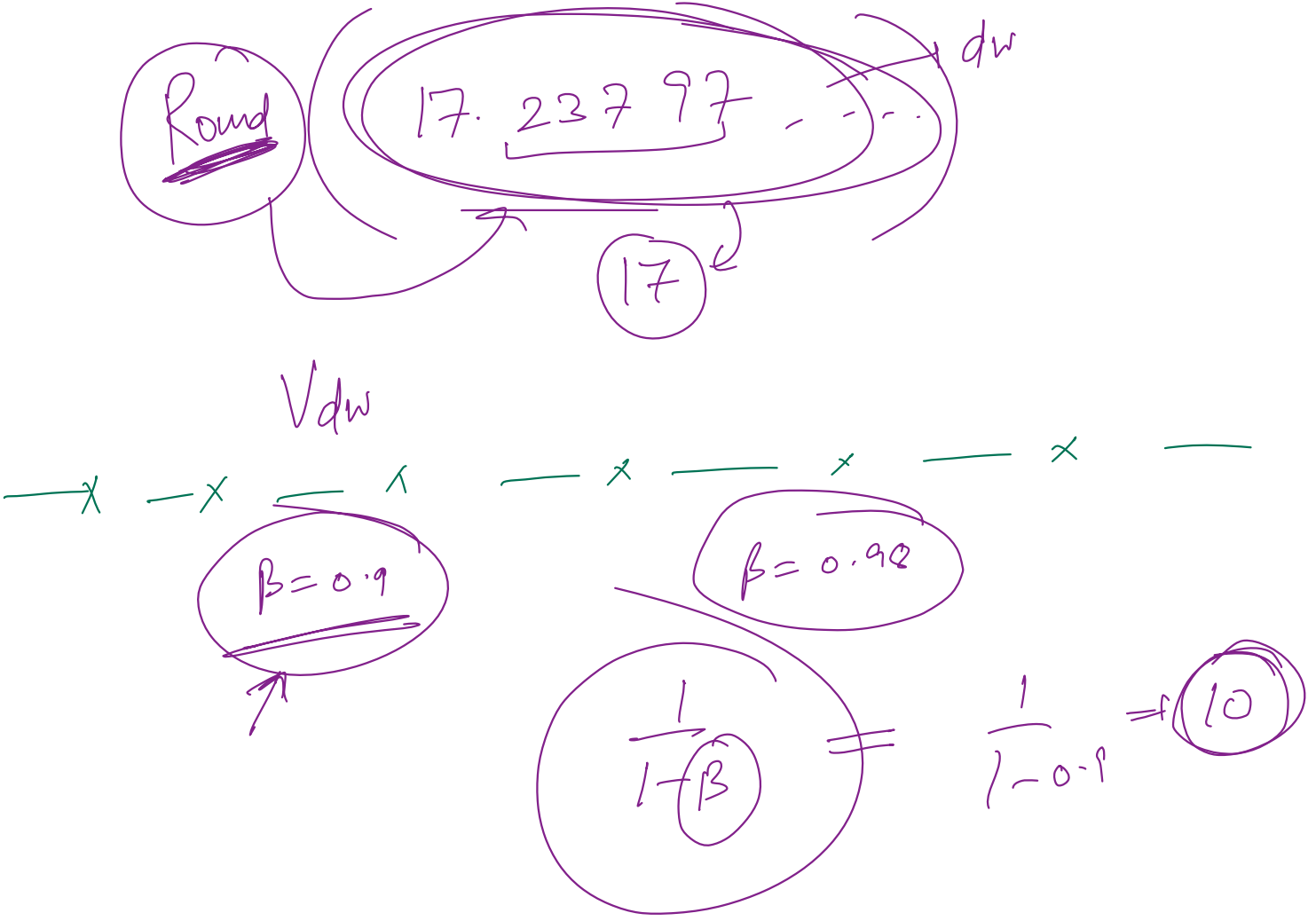
$$w^{new} = w^{old} - \alpha\, V_{dw}$$

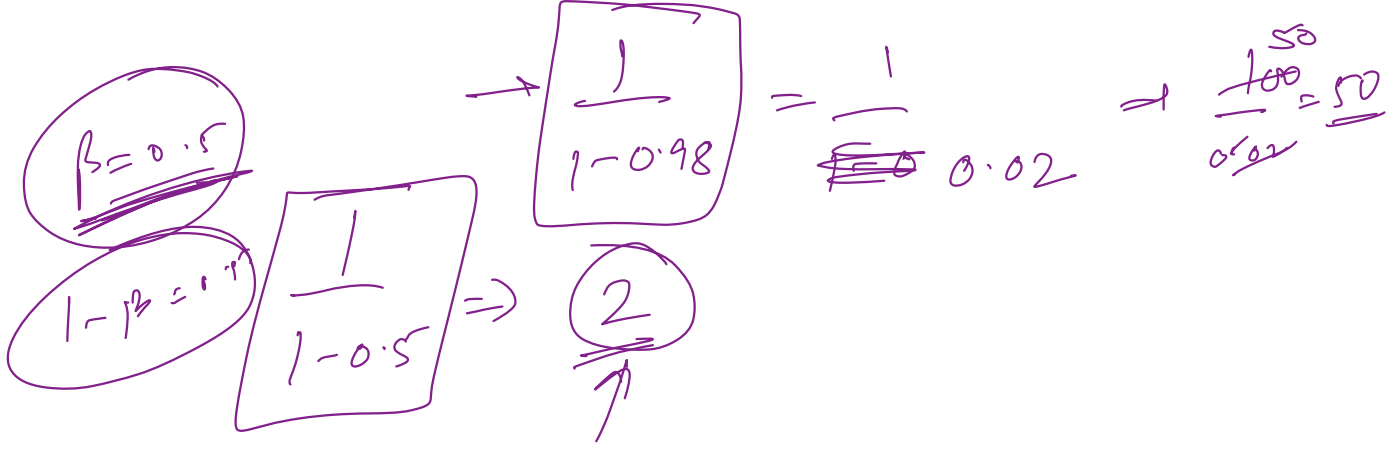
$$db = \frac{\partial J}{\partial b}$$

$$V_{db} = \beta V_{db} + (1-\beta)\, db$$

friction — momentum — acceleration

$$b^{new} = b^{old} - \alpha V_{db}$$


Round
17.23797.....
dw
17
Vdw
β=0.9
β=0.98
1/1-β
1/1-0.9
10


β = 0.5
1 − β = 0.5
1 / 1 − 0.5
⇒ 2
1 / 1 − 0.98
= 1 / 0.02
⇒ 100/2 = 50