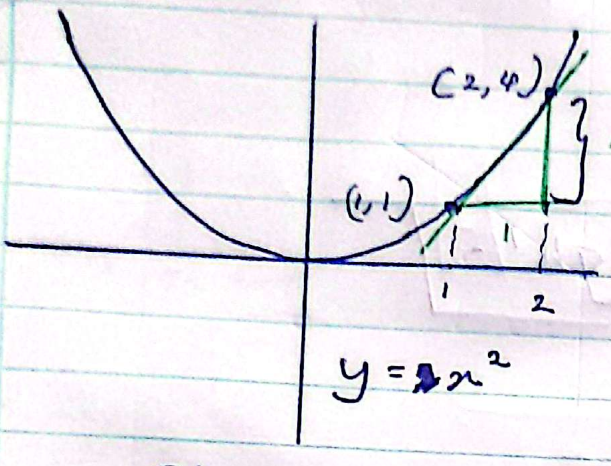
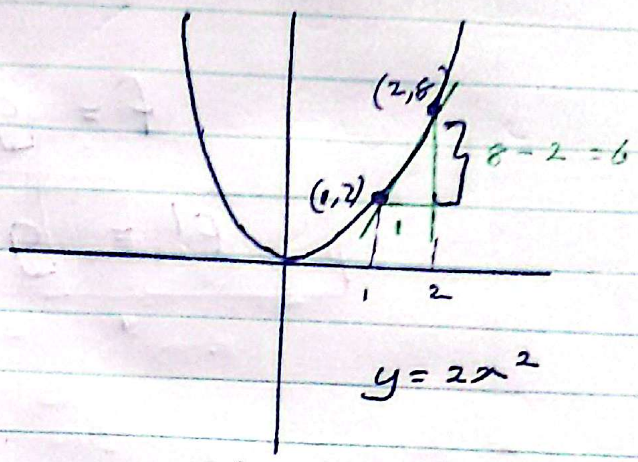
14] Multiplication by scalars



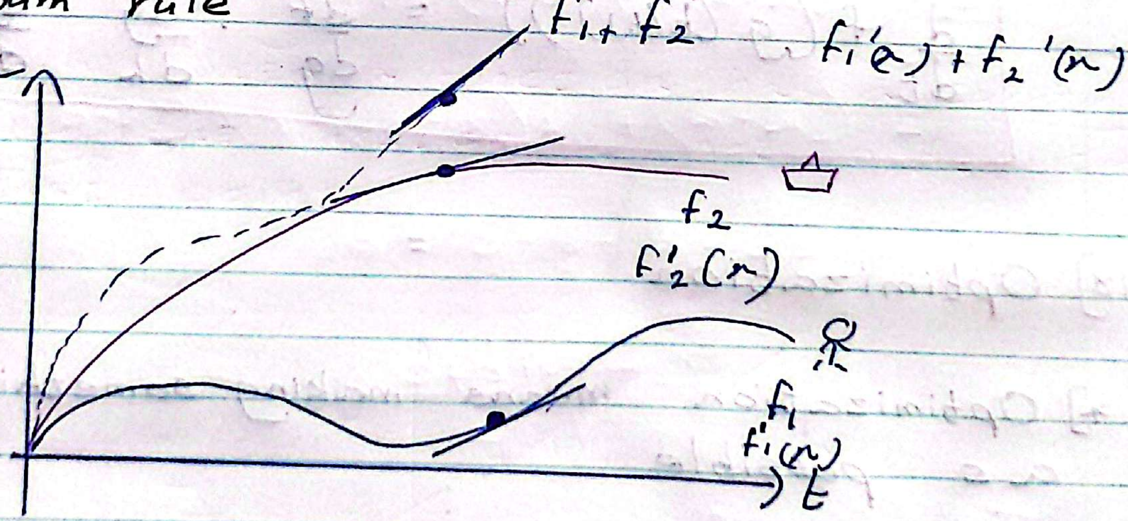


Slope = $\frac{3}{1} = 3$

Slope = $\frac{6}{1} = 6$

$$\boxed{f' = cg'}$$

15] The sum rule



$f(x) = g(x) + h(x)$

$$\boxed{f'(x) = g'(x) + h'(x)}$$

16] Product rule

$$f = gh$$

$$\boxed{f' = g'h + gh'}$$

17] Chain rule

$$\boxed{\frac{d}{dt} g(h(t)) = \frac{dg}{dh} \cdot \frac{dh}{dt}}$$

$$\boxed{\frac{d}{dt} f(g(h(t))) = \frac{df}{dg} \cdot \frac{dg}{dh} \cdot \frac{dh}{dt}}$$

18] Optimization

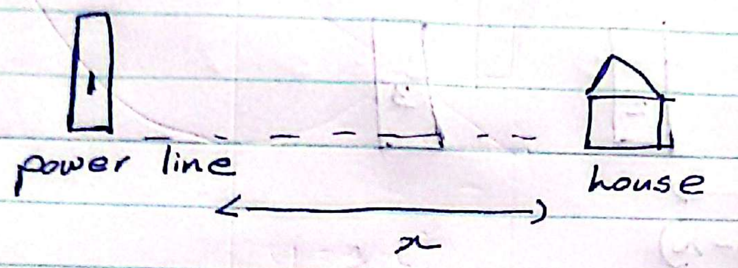
*] Optimization means making something as good as possible

*] In order to find the best model in ML we calculate an error function. When we minimize this error function we have the best model

*] This means follow the slope (derivative) of a function downhill until you reach the lowest point (slope = 0)

18] Optimization of squared loss

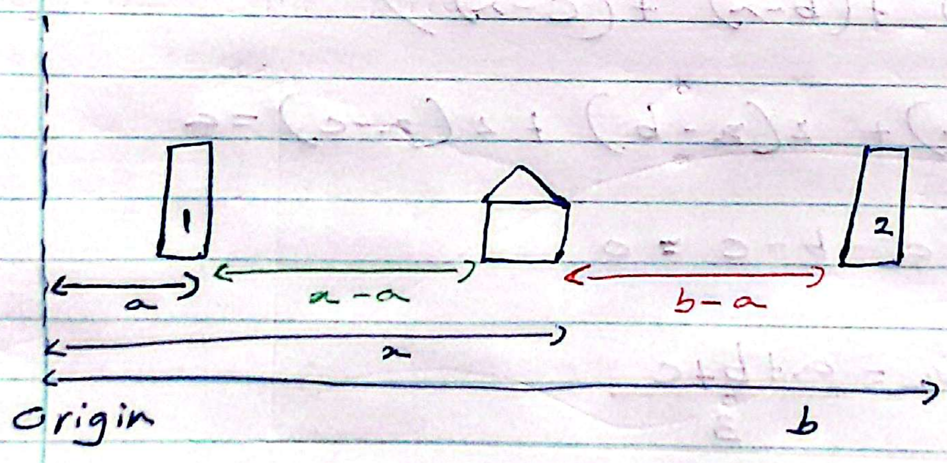
1] The one powerline problem



*] Cost of connecting power line 1 is x^2

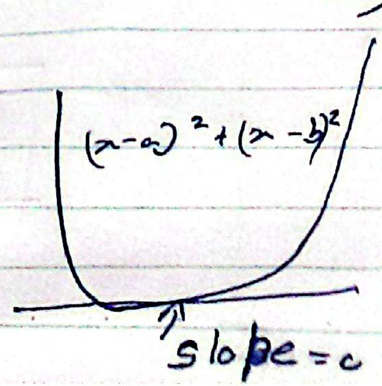
*] In order to minimize the cost x should be 0 (nearby powerline 1).

2] The two powerline problem



$$\text{Cost} = (x-a)^2 + (b-a)^2$$ (Goal: minimize the cost)

$$\frac{d}{dx}\left((x-a)^2 + (x-b)^2\right) = 0$$

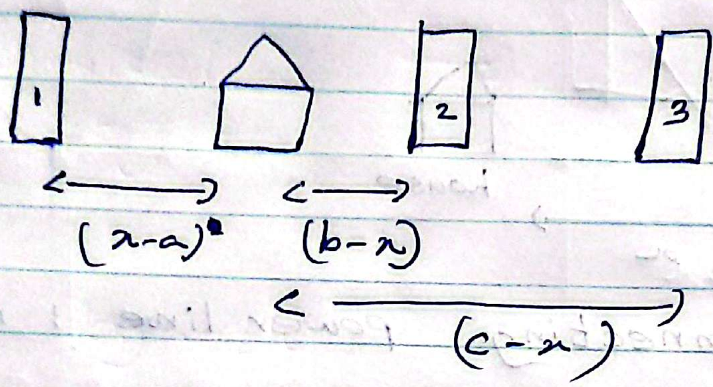


$$2(x-a) + 2(x-b) =$$

$$2x = a+b$$

$$x = \frac{a+b}{2}$$ (write in the middle)

3] The three powerline problem



Cost of $= (x-a)^2+(b-x)^2+(c-x)^2$

(Goal : minimize the cost)

$$\frac{d}{dx}\left((x-a)^2+(b-x)^2+(c-x)^2\right)$$

$$2(x-a)+2(x-b)+2(x-c)=0$$

$$3x-a-b-c=0$$

$$x=\frac{a+b+c}{3}$$

The square loss

Minimize $(x-a_1)^2+(x-a_2)^2+\dots(x-a_n)^2$

Solution $= \frac{a_1+a_2+\dots a_n}{n}$

20] Optimization of log loss

- In order to win the coin game, when the coin is tossed 10 times → 7 need to be head and 3 need to be tails.

Goal is to find which bias coin will help to maximize our wining chances.

probability of head = $p \rightarrow p^7$ (7 times)
probability of tail = $(1-p) \rightarrow (1-p)^3$ (3 times

$$g(p) = p^7 (1-p)^3$$

In order to function to be monotonic (always goes one direction) we take log

$$\log(g(p)) = \log(p^7 (1-p)^3)$$

$$= \log p^7 + \log(1-p^3)$$

$$= \boxed{7\log(p) + 3\log(1-p)} = G(p)$$

$-G(p)$ = log loss ↓

In order to find the maximum chances:

$$\frac{d}{dp}(\cancel{\ } 7\log(p) + 3\log(1-p)) = 0$$

$$\frac{7(1-p) - 3p}{p(1-p)} = 0$$

$$7(1-p) - 3p = 0$$

$$p = 0.7$$

H H H H H
H H H T T

H T

P $(1-P)$

Data

Model

Minimized logloss → $P = 0.7$

Why logarithm

1) Derivative of products is hard, derivative if sums is easy

2) Product of lots of tiny things is tiny.