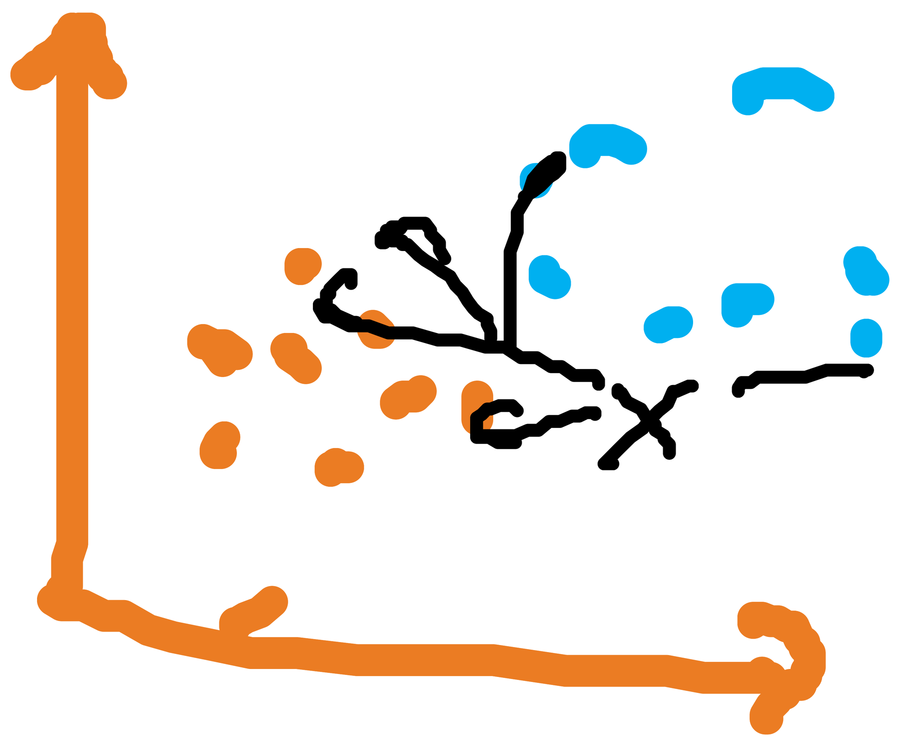
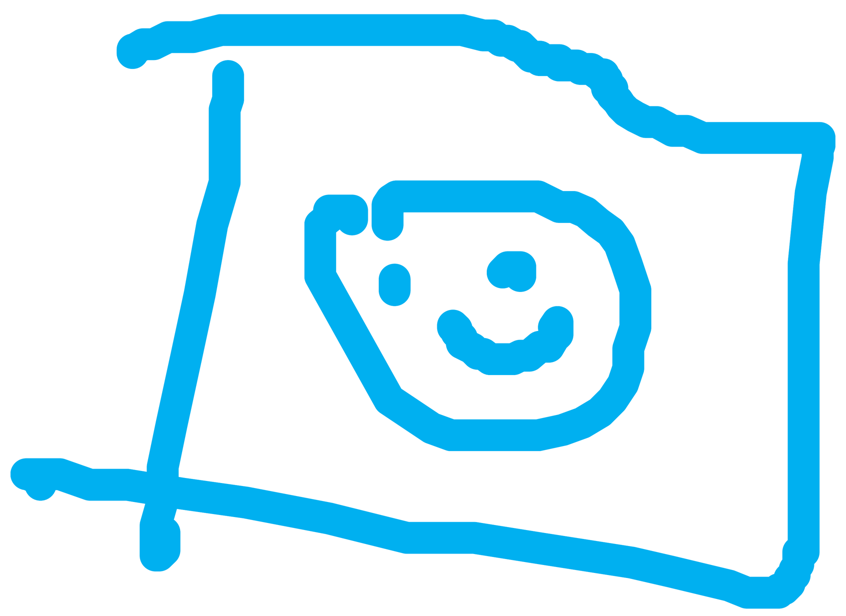
Apple = [x, y]

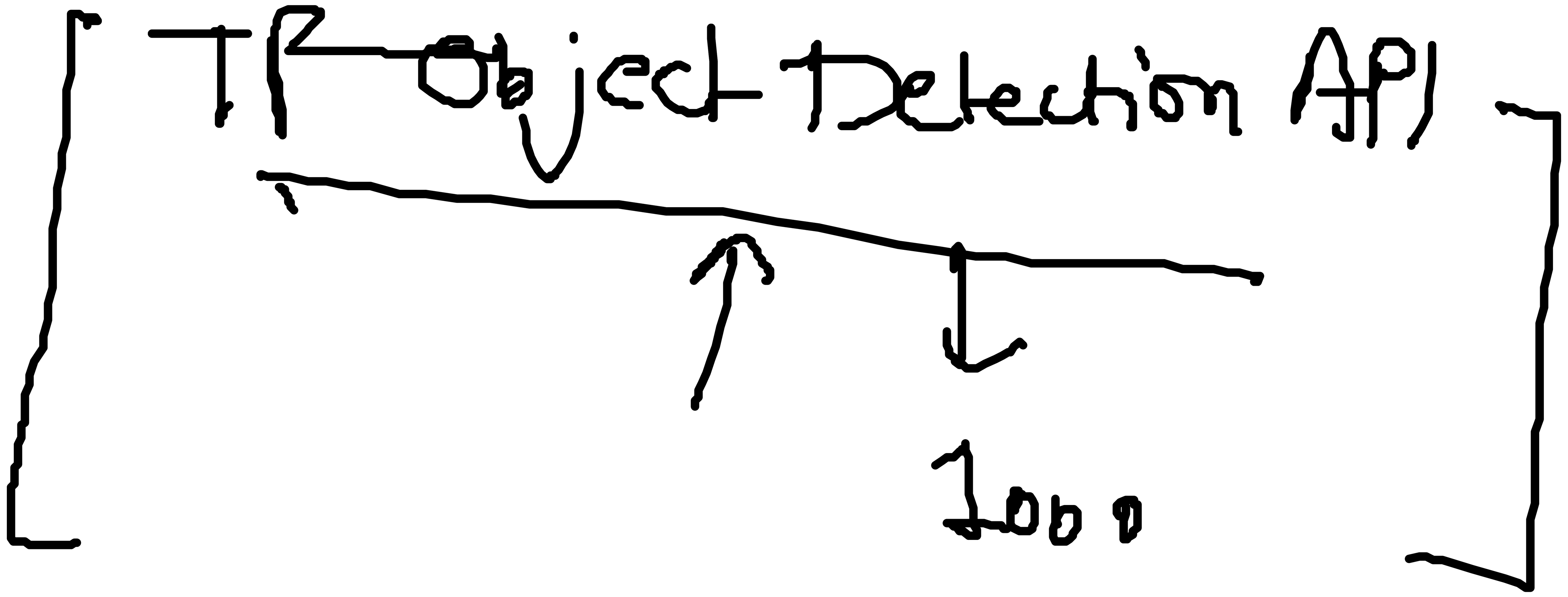
2dim





40x40

1600 dim vector



Introduction to ML

X: A	B	C	y: Marks
10	2	3	17
5	6	9	26
3	8	10	14
⋮	⋮	⋮	⋮
3	8	10	13

1. Supervised Learning
   80%.
2. Unsupervised Learing

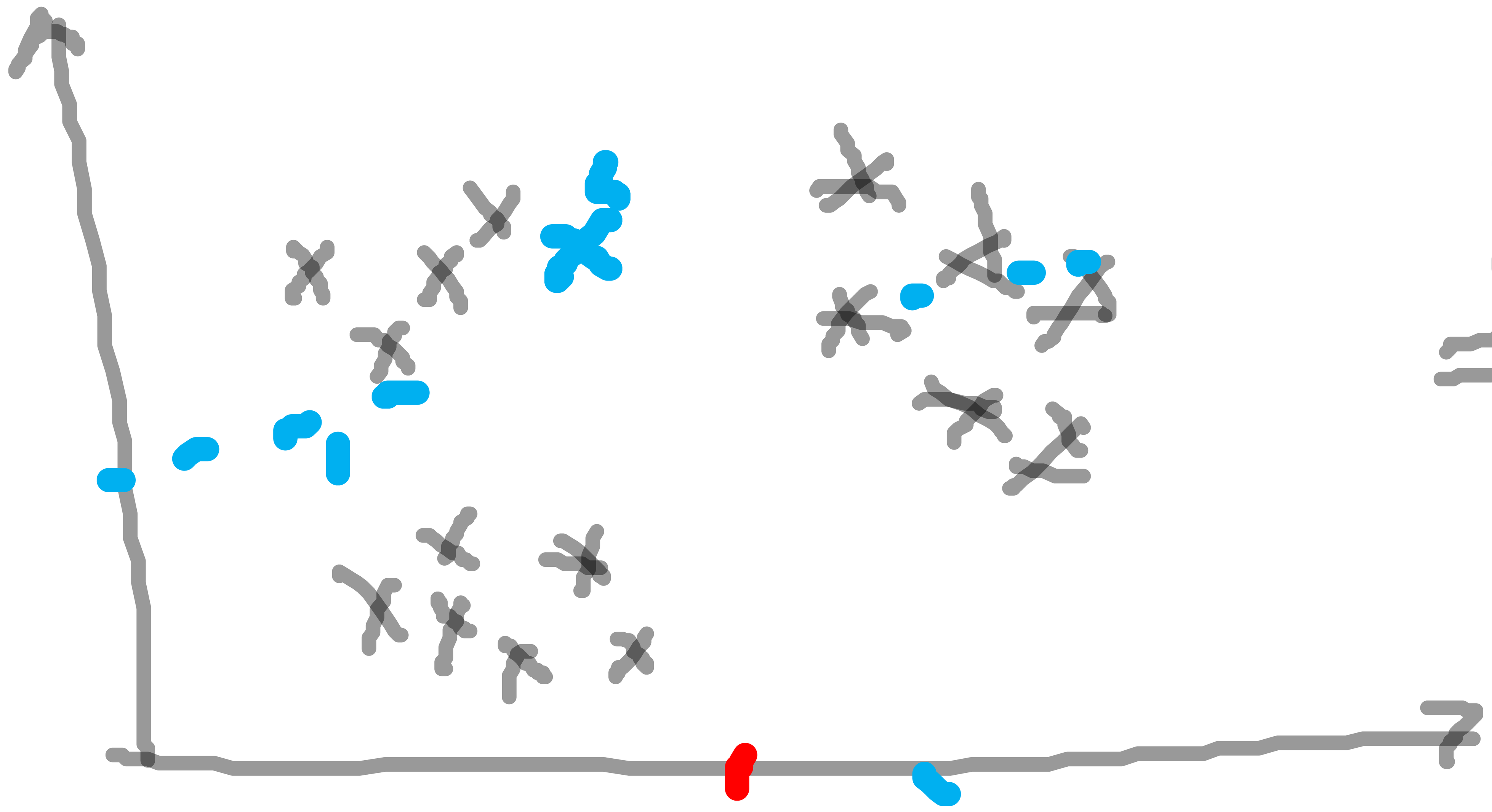
$$\underbrace{A * B - C}_{\text{Approx}}$$

$$X \longrightarrow Y$$

$$y' = f(x)$$

$\boxed{X}$

x1, y1, c1
x2, y2, c2
x3, y3, ...

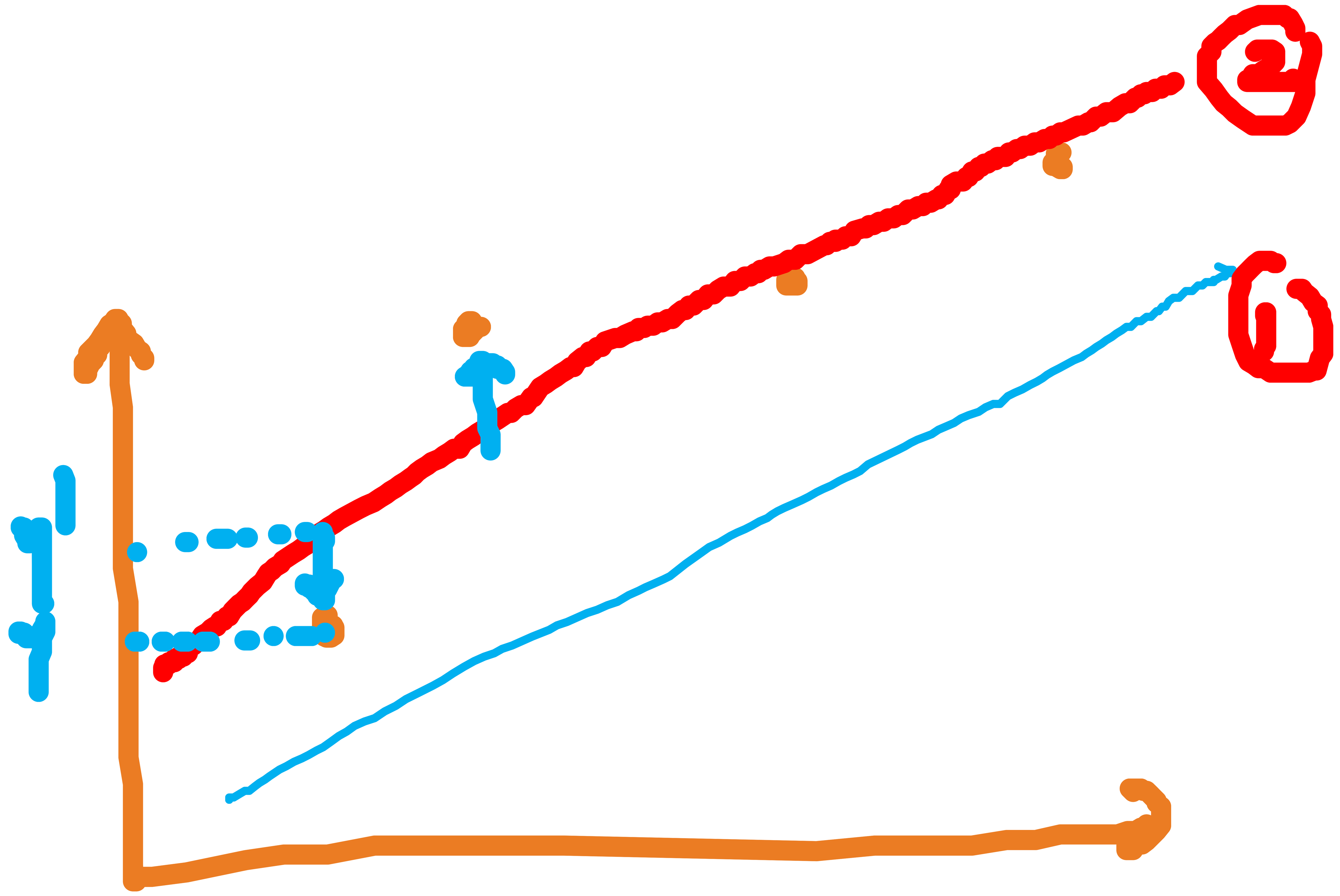


3 - ~~3~~

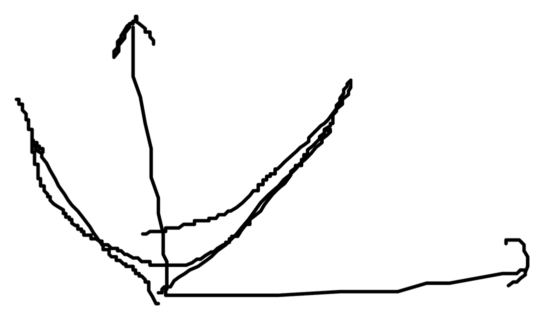
X	Y
Area	Price
10	100
20	220
40	420
50	490
⋮	
80	760

$60 \rightarrow$ Price

Linear Regression



$$\text{Error} = \sum_{i=1}^{n} (y' - y)^2$$



$x1, y1$
$x2, y2$
$x3, y3$
$\vdots$
xn, yn

classification — Regression

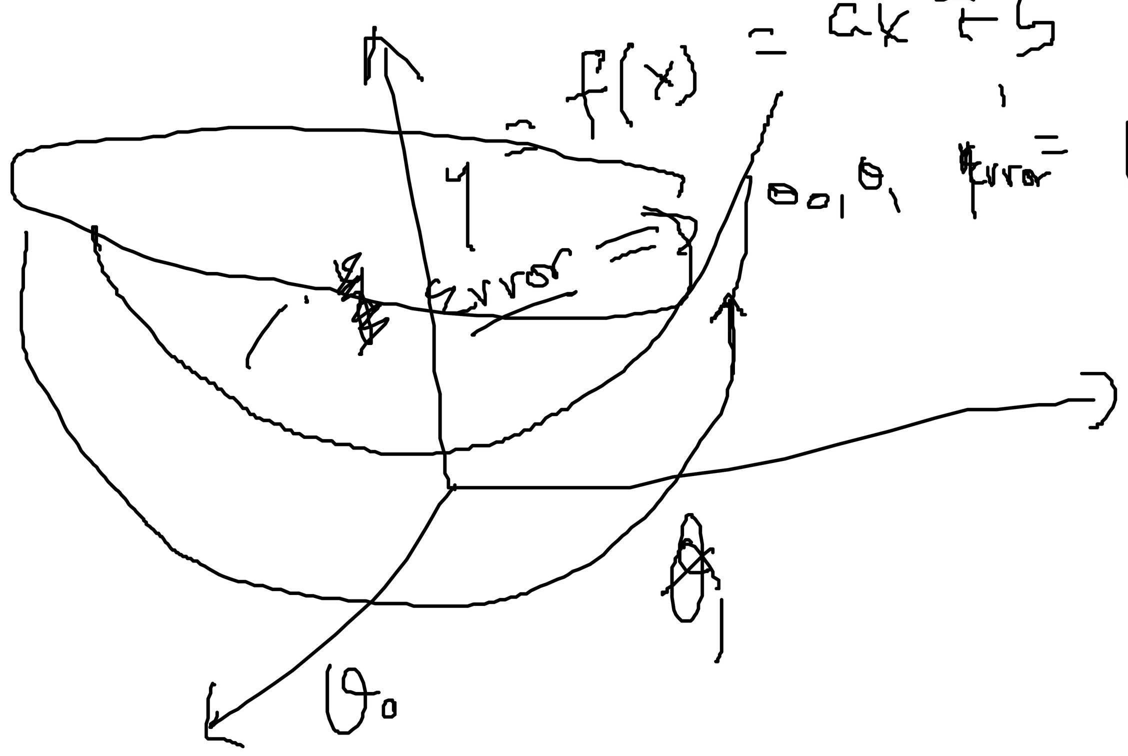
Discrete

0,
1,
2

1,2...k

10.56
10882
100.1

conti value



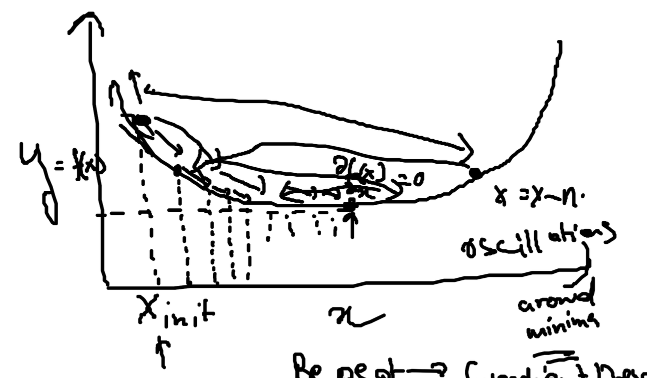
$f(x) = ax^2 + b$

θ_0, θ_1 error $= (y' - y)^2$

$y' = \theta_1 x + \theta_0$

Find θ_1, θ_0 s.t. error is minimum

Mimization wrt one variable



$y = x^2 - 10x + 25$

$= (x-5)^2$

↓

minima

Minima

$2x - 10 = 0$

$x = \frac{10}{2} = \boxed{5}$

Repeat → Gradient Decent

$X = X_{init}$

$$\left[x = x - \eta \cdot \frac{\partial f(x)}{\partial x} \right]$$

↑ update

$x = x - \Delta x$ Converged to local minima

very small = 0.001

Loss function

$$\varepsilon = \sum_{i=1}^{m} [\hat{y}_i - y_i]^2 \qquad \hat{y}_i = \theta_1 x + \theta_0$$

grad
$$\frac{\partial \varepsilon}{\partial \theta_0} = \sum_{i=1}^{m} 2[\hat{y}_i - y_i]$$

$$\frac{\partial \varepsilon}{\partial \theta_1} = \sum_{i=1}^{m} 2[\hat{y}_i - y_i] \frac{\partial \hat{y}_i}{\partial \theta_1} \quad = x$$

$$= \sum_{i=1}^{m} 2[\hat{y}_i - y_i] * x$$

$\partial y / \partial x$

$y = f(x)$

$E = f(\theta_0, \theta_1)$

$\frac{\partial \varepsilon}{\partial \theta}$

$= \frac{\partial \varepsilon}{\partial \theta_0}$

$\frac{\partial \varepsilon}{\partial \theta_1}$

Gradient descent

$\theta_0, \theta_1 = \text{random}()$

update Rule

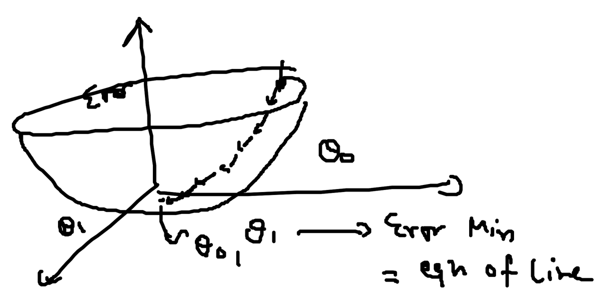
while ( ) {

$\theta_0 = \theta_0 - \eta \cdot \frac{\partial \varepsilon}{\partial \theta_0}$

$\theta_1 = \theta_1 - \eta \cdot \frac{\partial \varepsilon}{\partial \theta_1}$

}

Gradient Descent update Rule



N Multivariable Regression

Many Features

$y = f(\underline{x})$

$= \boxed{\theta_0, \theta_1, \theta_2, \theta_3}$

\# Area

\# Rooms

\# Locality

$$y = \underbrace{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3}_{\text{scalar}}$$

$x_j^{(i)}$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}_{(4,1)} \qquad x^{(i)} = \begin{bmatrix} 1 \\ x_1^{(i)} \\ x_2^{(i)} \\ x_3^{(i)} \end{bmatrix}_{(4,1)}$$

$$y = \boxed{\theta^T X} \rightarrow \text{avoid loop}$$

$1 \times 4 \quad 4 \times 1$

$$= [\theta_0 \; \theta_1 \; \theta_2 \; \theta_3] \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

np.dot(x, y)

↑ fast

- Less time

$1-2$

$$= \underset{1}{\theta_0} + \underset{2}{\theta_1 x_1} + \underset{3}{\theta_2 x_2} + \theta_3 x_3$$

$$\hat{y}^{(i)} = \sum_{j=0}^{n} \theta_j x_j^{(i)}$$

where $x_j^{(i)} = 1$ if $j == 0$

$$E = \text{Loss} = \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial E}{\partial \theta_j} = \sum_{i=1}^{m} 2 * \left(\hat{y}^{(i)} - y^{(i)}\right) x_j^{(i)}$$

$\theta^T X \rightarrow$ vector

θ = np.zeros(n)

while(____)

~~θ = [ ]~~

for j in range(n):

$$\theta_j = \theta_j - \eta \cdot \frac{\partial E}{\partial \theta_j}$$

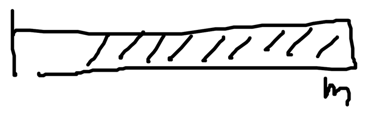
$$= \theta_j - \eta \cdot \sum_{i=1}^{m} 2 * \left(\hat{y}^{(i)} - y^{(i)}\right) x_j^{(i)}$$

Gradient Descent

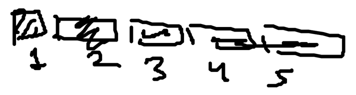
Batch

Mini

Stochastic Gradient Descent



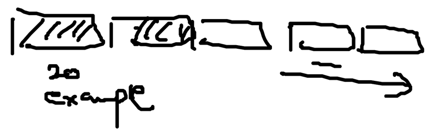
$\rightarrow$ g.D.



$\rightarrow$ Stochastic gradient Descent

grd for ith example

$\theta_j = \theta_j - \eta \cdot \text{grd}$



20 example

SGD

GD

faster convergence

$\rightarrow$ close to minima

better minima