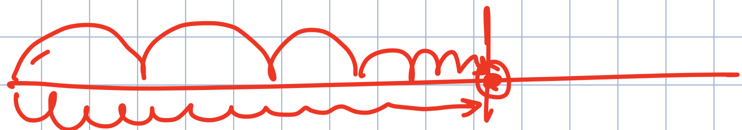
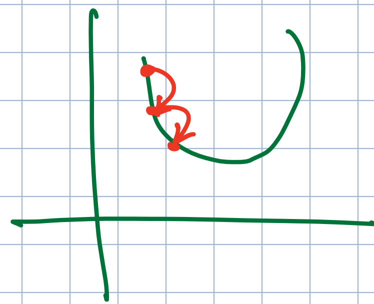
$\eta \rightarrow$ learning rate adaptive





A travel agency wants an automated system to predict travel costs. The agency has the following data available with it.

(x) Table II (y) → Regression

	S. No.	Distance (in Km)	Travelling Cost (in Rupees)
x^1	1	1	2.75
x^2	2	2	3.5
x^3	3	3	4.25
x^4	4	4	5
x^5	5	5	5.75

Page **1** of **3**

Formulate the above problem as a linear model $h(x) = w_0 + w_1 x$ to predict the travelling cost for a given distance. The parameter w_0 is 2 (optimal). Apply gradient descent algorithm to find optimal parameter w_1. The learning rate for the first epoch is 0.073, and for the second epoch and later, the learning rate is 0.091. Let the initial value of w_1 is 0.5.

$$\theta_1 = \theta_1 - \eta \frac{\partial J(\theta)}{\partial \theta_1} \rightarrow \theta_1 = \theta_1 - \frac{\eta}{m} \cdot 2 \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right) x^{(i)}$$

initialize → 0.5

do

$J(\theta)$

update θ → $\eta = 0.073$

$\eta = 0.091$

while (converge)

1st epoch

$x^{(i)}$	$h_\theta(x^{(i)})$ $\hat{y}^{(i)}$	$y^{(i)}$	$\hat{y}^{(i)} - y^{(i)}$	$(\hat{y}^{(i)} - y^{(i)}) x^{(i)}$
1	$2 + 0.5 * 1 = 2.5$	2.75	-0.25	-0.25
2	$2 + 0.5 * 2 = 3$	3.5	-0.5	-1

3	$2+0.5*3 = 3.5$	4.25	-0.75	-2.25
4	$2+0.5*4 = 4$	5	-1	-4
5	$2+0.5*5 = 4.5$	5.75	-1.25	-6.25

$$\Sigma = -13.75$$

$$\Sigma \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

$$h_\theta(x^{(i)}) = \theta_1 x^{(i)} + \theta_0$$
$$= 0.5\, x^{(i)} + 2$$

$$\theta_1 = \theta_1 - \frac{\eta * 2}{m} \Sigma \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

$$\theta_1 = 0.5 - \frac{0.073 * 2}{5} (-13.75)$$

$$\theta_1 = 0.9$$

2nd epoch

repeat same process with $\theta_1 = 0.9$

Linear Regression with Multiple features

Eg: House Price Prediction

Features (x) | (Prediction) (y)

	#Area (x'_1)	#floors (x'_2)	#Bedrooms (x'_3)	#Age (x'_4)	Price
$x^{(1)}$	200	2	3	10	2.5
$x^{(2)}$	100	3	2	20	2
⋮	⋮				

$$X = \begin{bmatrix} \text{———} & x^1 & \text{———} \\ \text{———} & x^2 & \text{———} \\ & \vdots & \\ \text{———} & x^m & \text{———} \end{bmatrix}$$

m examples

$$X = \begin{bmatrix} x^1_1 & x^1_2 & x^1_3 & \cdots & x^1_n \\ x^2_1 & x^2_2 & x^2_3 & & x^2_n \\ \vdots & & & & \\ x^m_1 & x^m_2 & x^m_3 & \cdots & x^m_n \end{bmatrix}$$

$m \times n$ → examples, features

x^i_j = ith example jth feature

Hypothesis

$$\hat{y} = h_\theta(x) = \theta_0 + \underbrace{\theta_1 x}_{\text{single feature}}$$

(LWS) x
0
0
0
0

x_1	x_2	x_3 ...	x_n

$$\hat{y} = h_\theta(x) = \underbrace{\theta_0}_{\text{bias}} + \underbrace{\theta_1 x_1 + \theta_2 x_2 + \cdots \theta_n x_n}_{\text{weight assign feature}}$$

$$h_\theta(x) = \theta_0 + \sum_{i=1}^{n} \theta_i x_i$$

$$h_\theta(x) = \theta_0 x_0 + \sum_{i=1}^{n} \theta_i x_i \qquad x_0 = 1$$

$$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i$$

n+1 features

x_0	x_1	x_2	. . .	x_n
1	–	–	– –	–

one example →

$$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i = \theta_0 x_0 + \theta_1 x_1 + \dots \theta_n x_n$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \qquad x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\theta^T = [\theta_0 \ \theta_1 \ \theta_2 \ \dots \ \theta_n] \qquad x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\theta^T x = \theta_0 x_0 + \theta_1 x_1 + \dots \theta_n x_n$$

$$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x$$

- Random
- Good your θ is ? → error
- Update your θ → G.D

Error / Loss Fxⁿ: $J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$

Update θ:

$$\theta = \theta - \eta \frac{\partial J(\theta)}{\partial \theta}$$

Gradient

$\nabla_\theta J(\theta)$

$\theta = [\theta_0 \ \theta_1 \ \theta_2 \ \ldots\ldots \ \theta_n]$

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

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

$$= \frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial \theta_j}\left(\theta_0 x_0 + \theta_1 x_1 + \ldots \theta_j x_j + \ldots \theta_n x_n - y^{(i)}\right)^2$$