Machine Learning

Supervised learning:

Train a model with correct answers to produce correct answers.

Regression

Continues valued o/p

classification

Discrete valued o/p (0/1)

0, 1, 2, 3 . .

unsupervised learning:-

Given the data, the model will find a structure (cluster)

(eg) Google news.

Linear Regression:

let

$m \rightarrow$ # of training exs

x's $\rightarrow$ input variable/features

ys $\rightarrow$ o/p variable/ target feature

$(x, y) \rightarrow$ single training eg.

$(x^{(i)}, y^{(i)}) \rightarrow$ i^{th} traing eg

Training set

↓

Learning algorith

↓

Size of the house (x) ⟶ h (hypothesis) ⟶ Estimated price (y)

h → maps from x's to y's.

$$h_\theta(x) = \theta_0 + \theta_1 x$$

↳ $h(x)$. θ_i's → parameter.

<u>Cost function</u>:-

$$\underset{\theta_0,\theta_1}{\text{minimise}} \left[\frac{1}{2m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]^2\right]$$

$$J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]^2$$

$$\min_{\theta_0,\theta_1} \underbrace{J(\theta_0,\theta_1)}_{\text{cost fn (or) squared error fn.}}$$

Cost function:

Intution-I

let $\theta_0 = 0$

$$h_\theta(x) = \theta_1 x$$

$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right]^2$$

$$\min_{\theta_1} J(\theta_1).$$

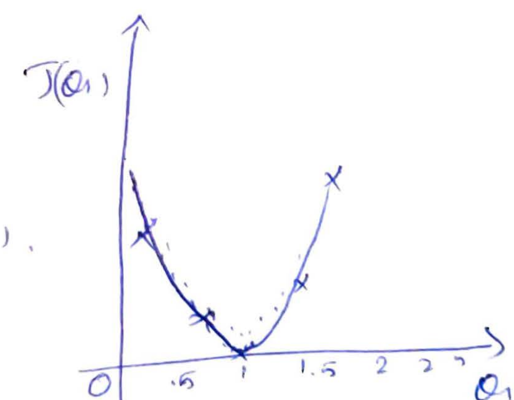
$h_\theta(x) \rightarrow$ for fixed value of θ_1, this is fn. of x

$J(\theta_1) \rightarrow$ fn. of parameter θ_1

change θ_1 in $J(\theta_1)$ and find $h_{\theta_1}(x)$.

aim is to find a line that is close to real line.

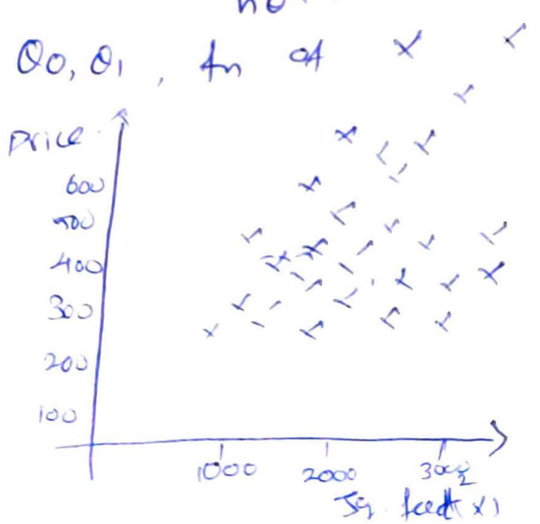
Points in $h_\theta(x^{(i)})$ is close to $y^{(i)}$.



Intution-II

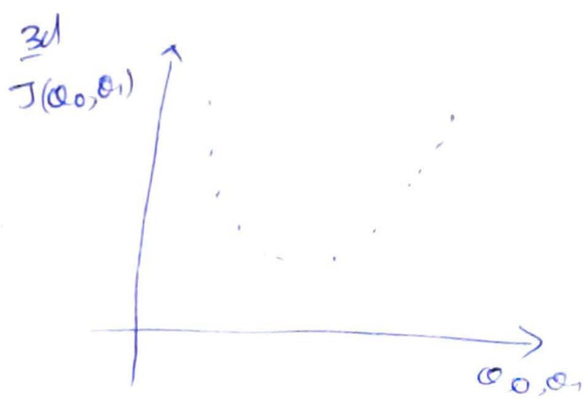
$h_\theta(x)$

fixed θ_0, θ_1, fn of



cost fn $J(\theta_0, \theta_1)$

(fn. of θ_0, θ_1)



Gradient descent :-

An: $J(\theta_0,\theta_1)$: aim is to $\min\limits_{\theta_0,\theta_1} J(\theta_0,\theta_1)$

→ start with some (θ_0,θ_1)

→ keep changing θ_0,θ_1 to reduce $J(\theta_0,\theta_1)$ untill we end up at a minimum.

Gradient descent algorithm :

repeat untill convergence {

$$\theta_j := \theta_j - \underbrace{\alpha}_{\text{learning rate}} \underbrace{\boxed{\frac{\partial}{\partial \theta_j} J(\theta_0,\theta_1)}}_{\text{derivative term}} \quad \text{for } j=0 \ \& \ j=1$$

}

Simultaneous update :-

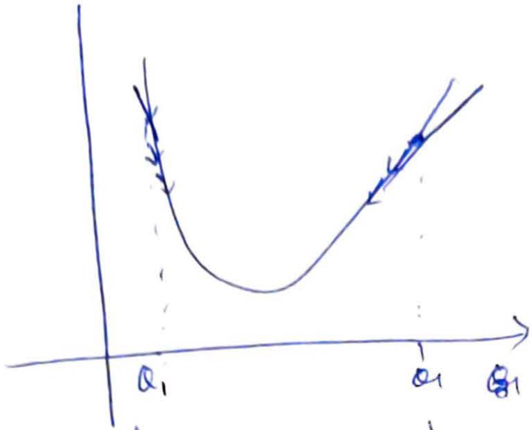
$$temp0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0,\theta_1)$$

$$temp1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0,\theta_1)$$

$$\theta_0 := temp0$$

$$\theta_1 := temp1$$

$J(\theta_1) \quad (\theta_1 \in \mathbb{R})$



$$\theta_1 := \theta_1 - \alpha \boxed{\frac{d}{d\theta_1}} J(\theta_1)$$

derivative

| derivative is slope |

If θ_1 is on this side.

If θ_1 on this side

→ If α → too small, GD can be slow.

→ α → too large, GD can overshoot → may fail to converge (or) even diverge.

→ GD can converge to a local minimum, even with learning rate 'α' is fixed.

no need to dec. 'α'

$$\frac{\partial}{\partial\theta_j} J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_j} \cdot \frac{1}{2m} \sum_{i=1}^{m} \Big[\underbrace{(h_\theta x^{(i)})}_{\theta_0+\theta_1 x^{(i)}} - y^{(i)}\Big]^2$$

$$\Big\{ \quad \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta x^{(i)} - y^{(i)})$$

$$\theta_1 := \theta_1 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \, x^{(i)} \quad \Big\}$$

cost fn of linear fn is always bowl shaped.
There is no local min. only Global min.