Linear Regression

01 October 2021 11:32 PM

ML

Supervised | Unsupervised

Regression | Classification

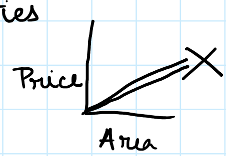
$X \rightarrow Y$ some examples known

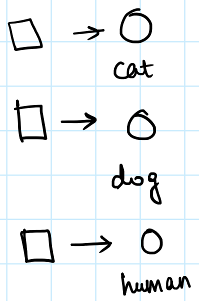
5 items $\rightarrow$ 20

8 " $\rightarrow$ 60

Price of house — Y

$\boxed{\text{Area}}$ location, amenities — X

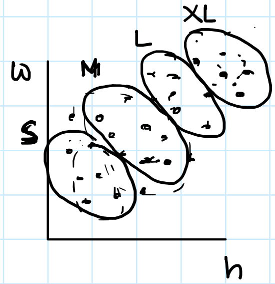




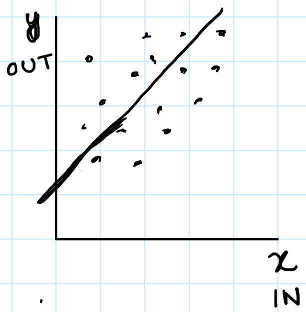
Unsupervised —

height, weight

S, M, L, XL



Linear Regression:



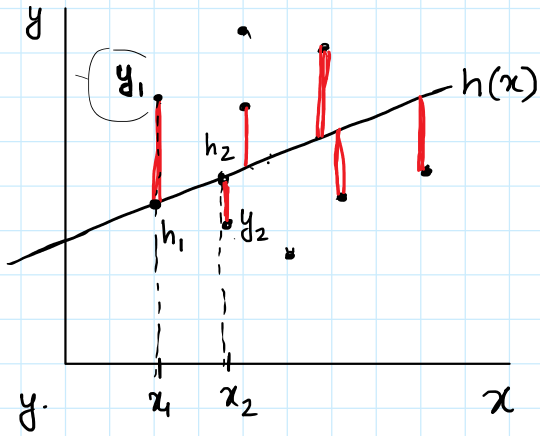
Training set (DATA)

$\downarrow$

Learning algorithm

$\downarrow$

$x \rightarrow$ Hypothesis (h) $\rightarrow$ Prediction



$h(x) = \theta_0 + \theta_1 x \rightarrow$ univariate

θ_0: y-intercept, θ_1: slope

$\theta_0, \theta_1 \rightarrow$ parameters

choose θ_0, θ_1 such that $h(x)$ is "close" to y.

distance b/w $h(x)$ and y is minimum. $\Rightarrow \frac{1}{2m}\left((h(x_1) - y_1)^2 + (h(x_2) - y_2)^2 + \dots\right)$

$x^{(i)} \rightarrow$ example

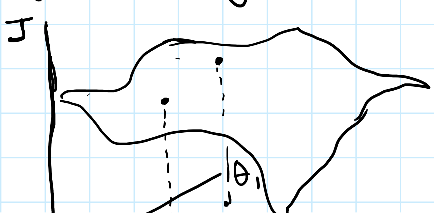
OBJECTIVE: $\min_{\theta_0, \theta_1} \underbrace{\frac{1}{2m} \sum_{i=1}^{m} \left(h(x^{(i)}) - y^{(i)}\right)^2}_{\text{COST FUNCTION } (J)}$

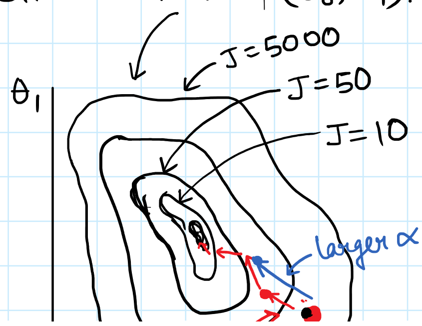
$$J = \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$$

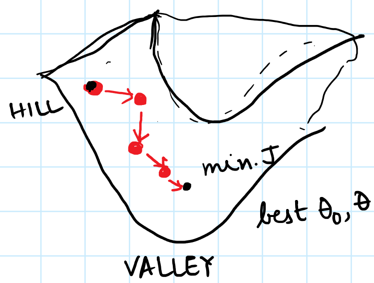
$J \rightarrow$ FUNCTION OF θ_0, θ_1

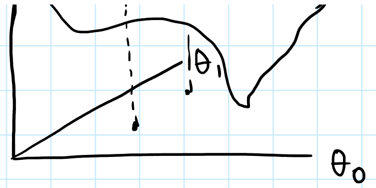
OBJECTIVE $\rightarrow$ min. $J(\theta_0, \theta_1)$.

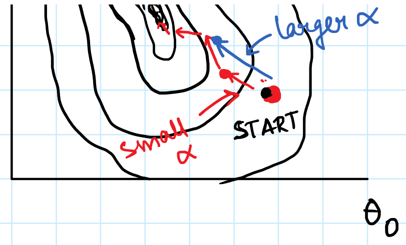
Physical meaning?

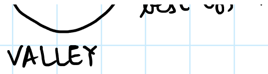












GRADIENT DESCENT

$\frac{\partial}{\partial \theta} J(\theta) \rightarrow$ steepness.

Repeat: $\theta = \theta - \alpha \frac{\partial}{\partial \theta} J(\theta)$.

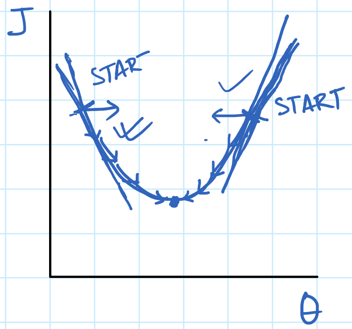
final, initial, factor

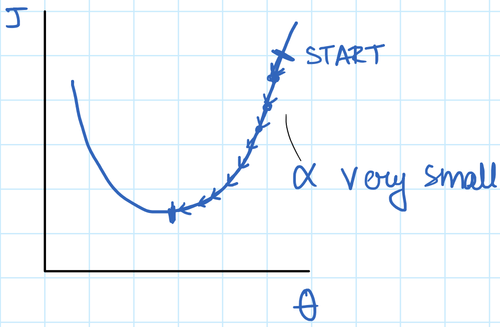
Repeat for 1000 times.

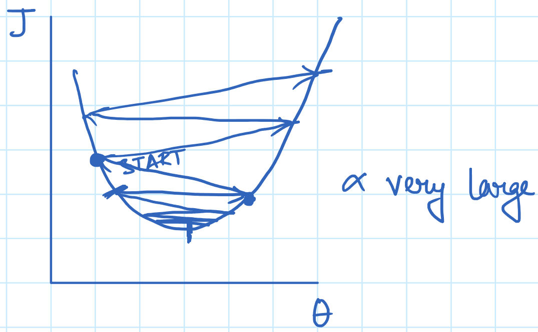
$$\theta_0 = \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

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

APPLIES ALL THE TIME







For linear regression, $h(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$ $\quad y^{(i)}$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h(x^{(i)}) - y^{(i)}\right)^2 = \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial}{\partial \theta_0} J = \frac{1}{m} \sum_{i=1}^{m} \left(\underbrace{\theta_0 + \theta_1 x^{(i)}}_{h(x^{(i)})} - y^{(i)}\right) \qquad \frac{\partial}{\partial \theta_1} J = \frac{1}{m} \sum_{i=1}^{m} x^{(i)} \left(h(x^{(i)}) - y\right)$$

Gradient descent:

$$\theta_0 = \theta_0 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \left(h(x^{(i)}) - y^{(i)}\right)$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} x^{(i)} \left(h(x^{(i)}) - y^{(i)}\right).$$

APPLIES ONLY TO LINEAR REGRESSION

WE WANT BEST-FIT LINE $h(x) = \theta_0 + \theta_1 x$.

if we minimize $J(\theta_0, \theta_1)$ then we get best line.

$\hookrightarrow$ gradient descent $\rightarrow$ final values of θ_0, θ_1 $\rightarrow$ $\boxed{h(x) = \theta_0 + \theta_1 x}$

COMPLETES LINEAR REGRESSION.

Things I need: initial θ_0, θ_1 (chosen randomly OR zero).

expression for J involving $\theta_0, \theta_1, y^{(i)}$.

need α, no. of steps.