Supervised Learning

Data has "Right Answers"

Regression

- predicts Numbers
- Infinitely many possible outputs

(like prices, temp, Quantity, ----)

Classification

- predicts Categories
- Small Numbers of possible outputs

(like cat_vs_dog, disease 10 classes)

Model Representation — Linear Regression with one Variable

- Housing prices

Size (ft^2) (x)	Price (10^3 \$) (y)
2104	460
1416	232
--	--
--	--

$m = 47$

x = Input / features
y = output / target
m = No. of training examples
$(x^{(i)}, y^{(i)}) \rightarrow i^{th}$ training example

$x^{(1)} = 2104$ $\quad y^{(1)} = 460$
$x^{(2)} = 1416$ $\quad y^{(2)} = 232$

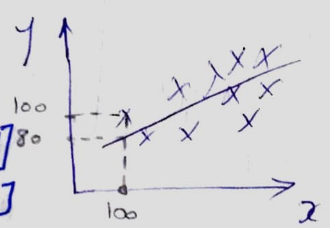
Training set
↓
Learning Algorithm
↓
$x \longrightarrow \boxed{f} \longrightarrow \hat{y}$ (prediction) (estimated y)

hypothesis (The model)
[Mapping function] f

$f(x)$ maps from x's to y's

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

Shorthand : $h(x)$ [old]
$f(x)$ [New]



→ We use Linear function only for simplification. Also, this is the basic function block that we can build on it.

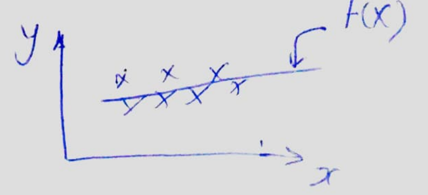
→ Recall that in regression problems, we are taking inputs variables to fit the output onto a continuous expected result function.

→ Linear Regression with one variable is also known as:

" Univariate linear Regression "

Single Input (x) — Single output $h(x)$

$f_{(w,b)}(x) = wx + b$



Cost Function

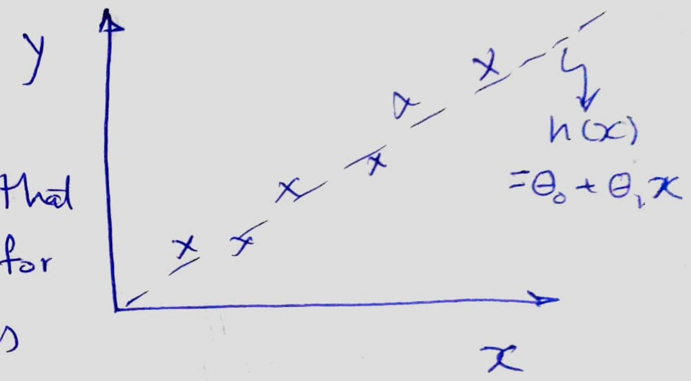
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(\overbrace{h_\theta(x^i)}^{\boxed{\hat{y}}} - y^i \right)^2$$

$$= \frac{1}{2} \overline{\text{error}^2}$$

Mean of squares of $\underbrace{(h(x) - y)}$

a Convenience for the computation of the gradient descent, as the derivative term of the square function will Cancel out the $\left(\frac{1}{2}\right)$ term

the difference between the predicted value ($h(x)$) & the actual value (y)



<u>Idea</u> : Choose θ_0, θ_1 so that $h_\theta(x)$ is close to (y) for our training examples (x, y)

by Minimizing the mean Squared error $J(\theta_0, \theta_1)$

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \underbrace{(\hat{y}^{(i)} - y^{(i)})^2}_{\text{error}}$$

$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

Find w, b so $\hat{y}$ close to y

- Cost Function

$$\text{Hypothesis:} \quad h_\theta(x) = \underset{b}{\theta_0} + \underset{w}{\theta_1}\, x$$

b w (New Terminology)

- Coefficients
- Weights

θ_i's : parameters → How to choose ??

$\boxed{\text{In ML: the variables you can adjust during training to improve the model}}$

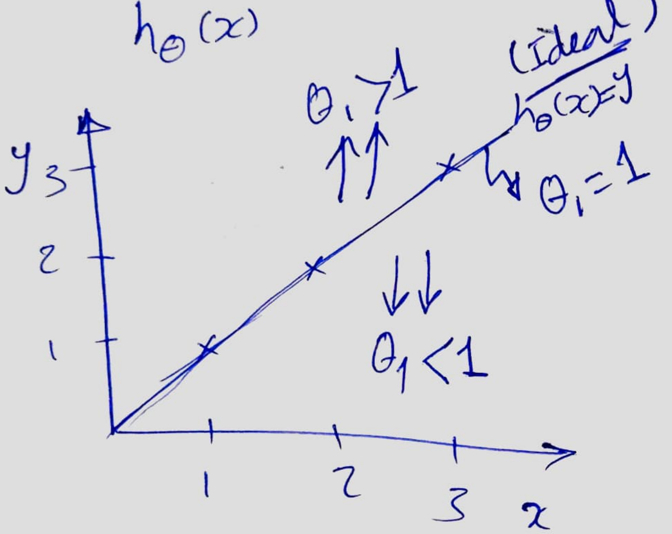
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 \Rightarrow \text{Cost function}$$

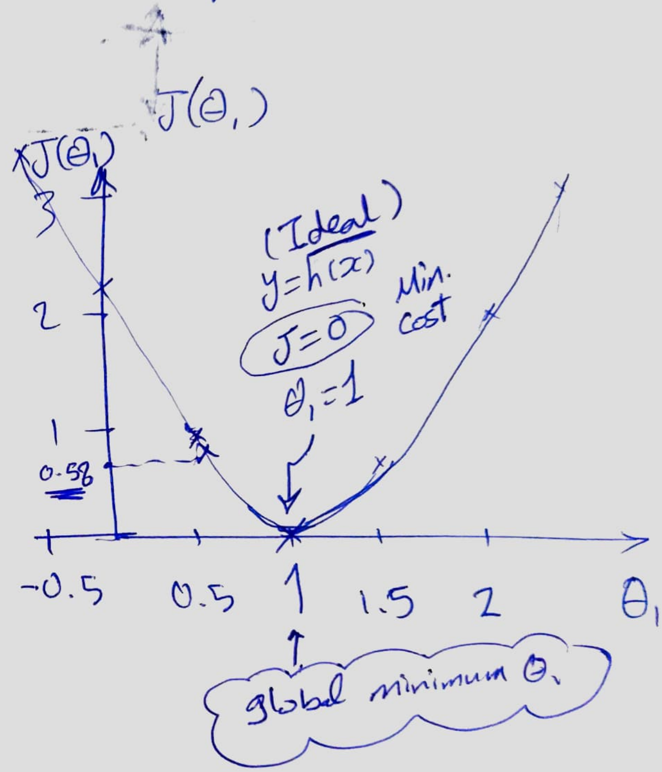
$\hat{y}$

To measure the accuracy of our Hypothesis

(Squared_Error function)

→ minimize $J(\theta_0, \theta_1)$ is our goal



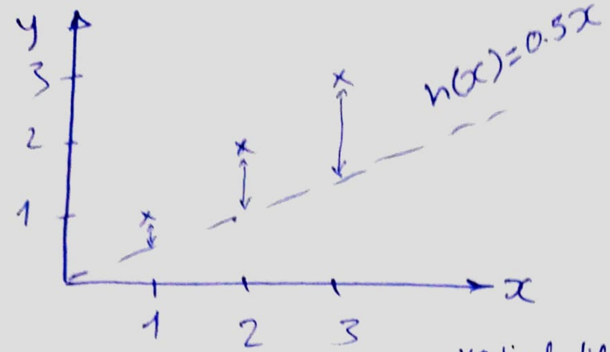


$h(x) = \theta_1 x$ (for simplicity) $\boxed{\theta_0 = 0}$

→ When using both θ_0 & θ_1 to represent the Cost function, we use 3D plotting or Contour plots to represent it.

→ The final target is to make a software to deal with such complicated plots to get the minimum Cost function which corresponds to the best fit hypothesis to our algorithm to predict the house pricing.

* $\underline{\theta_1 = 0.5}$

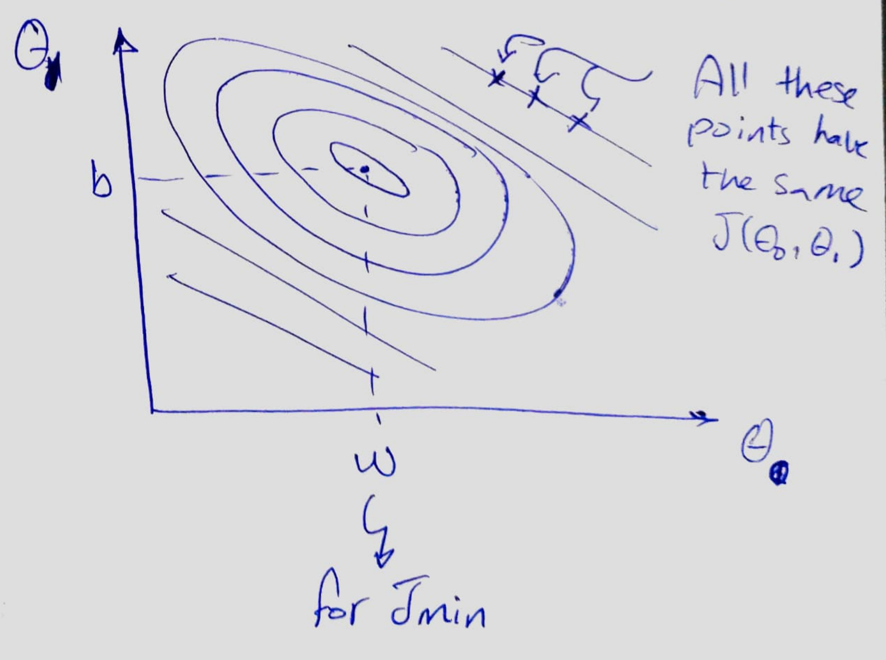


$$J(0.5) = \frac{1}{2 \times \underset{\underset{m}{\uparrow}}{3}} \Big[\overbrace{(0.5-1)^2 + (1-2)^2 + (1.5-3)^2}^{\text{vertical differences}} \Big]$$

$$= \frac{1}{6} \left[0.25 + 1 + 2.25 \right] \simeq 0.58$$

* Contour plot :

A two variable function has a constant value at all points of the same line



* 3D