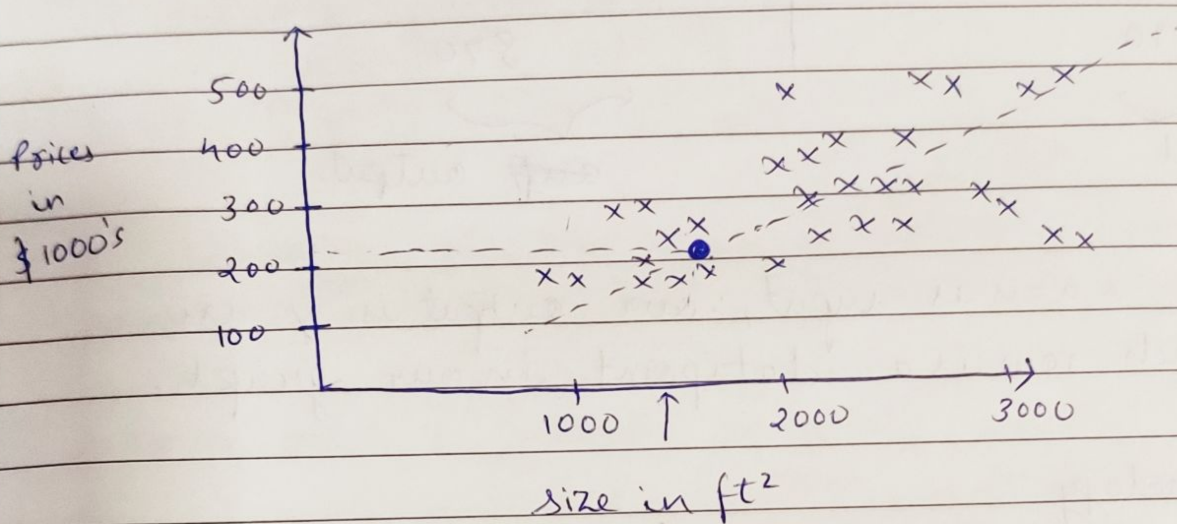
4. Linear Regression.

I. Linear Regression with one variable.

Example: House price prediction.



Each datapoint is the size and the price of that house.
For a house with 1250ft², what is the price?

If we draw a straight line, we get around 220K.

Supervised learning model. Data has "right answers".
Regression model Predicts numbers. We have infinitely many possible outputs.
Another supervised model is, Classification model that predicts categories. We only have small number of possible outputs.

Another way to look at the table is by using a

Data Table

size in ft^2	price in $1000's.
2104	400
1416	232
1534	315
852	178
....	...
3210	870

$\underbrace{\quad}_{\text{input}}$ $\underbrace{\quad}_{\text{output}}$

input ~~outp~~ output

So, our x axis is input, our output is y-axis.
Each row is a datapoint in our graph.

II. Terminology.

Data used to train the model is called a Training Set.
We use our training set to train our ML model.

Notation:

x = "input" variable.
it is also called a feature.

y = "output" variable.
"target" variable.

We can have a dataset representing each point with (x, y). = $(2104, 400)$.

Since, in our dataset, we have $m = 47$.
To denote a specific training set example, we represent them using this:

$$\left(x^{(i)}, y^{(i)}\right) = i^{th} \text{ training example.}$$
$(1^{st}, 2^{nd}, 3^{rd} \ldots)$

i donate a specific row in our table.

So, $\left(x^{(1)}, y^{(1)}\right) = (2104, 400)$.

i is actually not the expoment. The i is just the index in our training set.

VII. training set: {features, targets}.

We feed training set to learning algorithm.

Let our 'f' be called the function, the output of our learning algorithm.
It used to be called Hypothesis as well.

the job of 'f', is to take a new input 'x', and output a new output $\hat{y}$ (y-hat).

$$\underset{\text{(feature)}}{x} \longrightarrow \boxed{f} \longrightarrow \hat{y} \text{ (prediction)}.$$
(estimated value of y).

$\hat{y}$ is the estimate or prediction of value y.

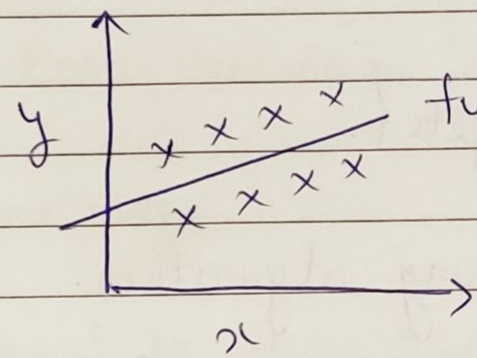
IV. How to represent f?

for now, let our f be a straight line.

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

here w and b are numbers. using the values of w and b, f will output some value with input as x.

An example:



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

$f(x) = wx + b$

Make prediction of y using w and b.

Here, we are just using linear values. We can use complex functions as well.

Here, our model is called a Linear Regression with one variable. [single input variable x].

It is also called <u>Univariate</u> linear regression.

I. Cost function formula.

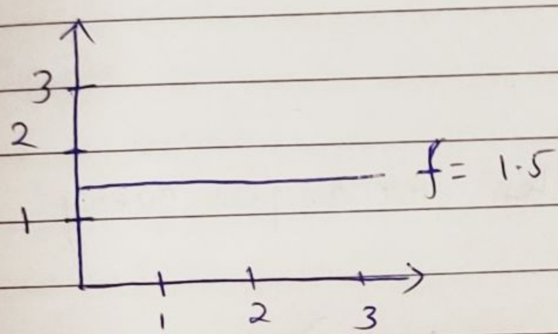
We have our training set with (x, y).

We use Model : $f_{w,b}(x) = wx + b$.

w, b are the parameters of the model.

{ coefficients or weights }.

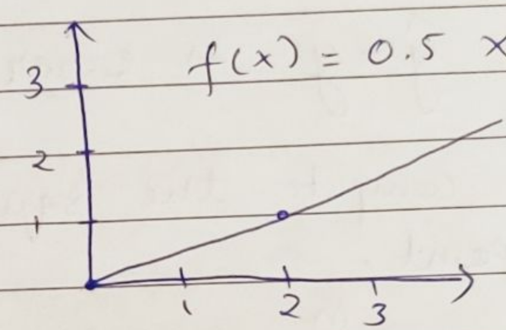
depending on the value of w, b we can have different plots.



$w = 0$

$b = 1.5$

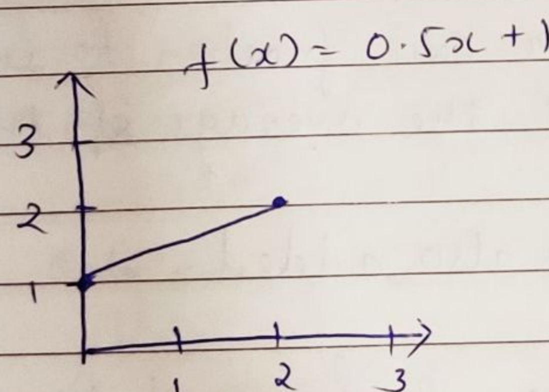
(b is called y-intercept).



$w = 0.5$

$b = 0$

the value of w is the slope of our line.



again, slope is 0.5, and y intercept is 1.

$w = 0.5$

$b = 1$

Now, predicted value for a $x^{(i)}$, is $\hat{y}^{(i)}$

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

our Model f is, $f_{w,b}(x^{(i)}) = w x^{(i)} + b$

Find w, b: such that $\hat{y}^{(i)}$ is close to $y^{(i)}$ for all $(x^{(i)}, y^{(i)})$.

VI. Cost function.: Squared error cost function.

$$\hat{y} - y \rightarrow \text{error}$$

we compute the square of the error. for every point.

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

m = number of training examples.

Since, we don't want our cost function to increase as m increases, we calculate the average of the value.

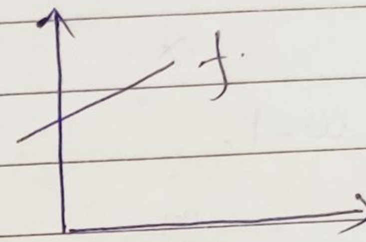
The extra division by 2 is also added.

We call this cost function, Squared Error Cost function.

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)} \right)^2$$

We need to find value of w, b which makes the cost function small.

VII. Cost function Intuition.



Recap:

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

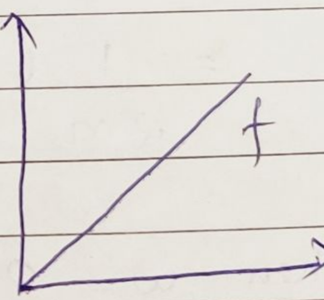
parameters = w, b

cost function:

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)} \right)^2$$

goal: $\underset{w,b}{\text{minimize}}\ J(w,b)$

Simplified: $f_w(x) = wx$



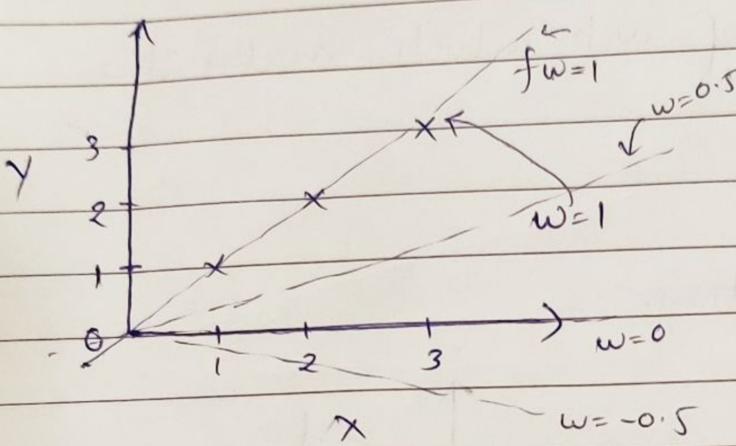
$$J(w) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_w(x^{(i)}) - y^{(i)} \right)^2.$$

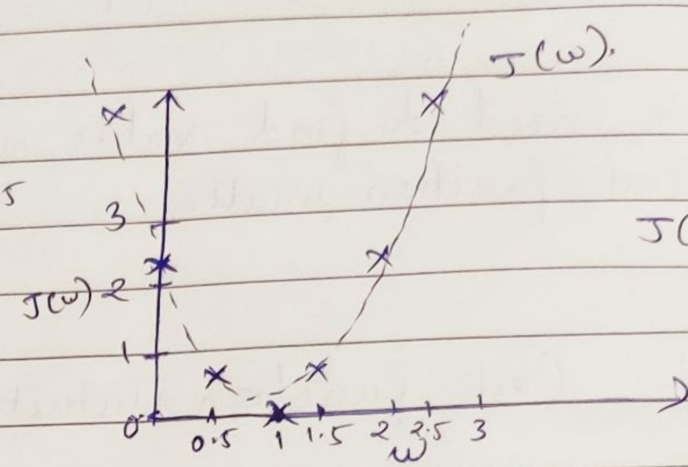
Only one parameter w.

$\underset{w}{\text{minimize}}\ J(w)$.

$f_w(x)$
(for fixed w, function of x)
$f_w(x)$ depends on input.

$J(w)$.
(function of w)
J depends on parameter w.





$J(1)=0$

when $w=1$,

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

$$= \frac{1}{2m}\sum_{i=1}^{m}\left(\cancel{w(x^{(i)})}\; w x^{(i)} - y^{(i)}\right)^2$$

for every point, our cost here is 0.

$$= \frac{1}{2m}\left(0^2+0^2+0^2\right) = 0$$

$m=3$ here.

when $w = 0.5$, $J(0.5) = \dfrac{(0.5-1)^2 + (1-2)^2 + (1.5-3)^2}{2\times 6}$

$$= \frac{3.5}{6} = 0.58$$

how to choose value of w?.
choose w to minimize $J(w)$.

The smallest possible value of $J(w)$ here is 0, so we use the value of I here.

goal of linear regression:

$$\underset{w}{\text{minimize}}\ J(w)$$

General case,

$$\underset{w,b}{\text{minimize}}\ J(w,b)$$

VIII. Visualizing Cost function.

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

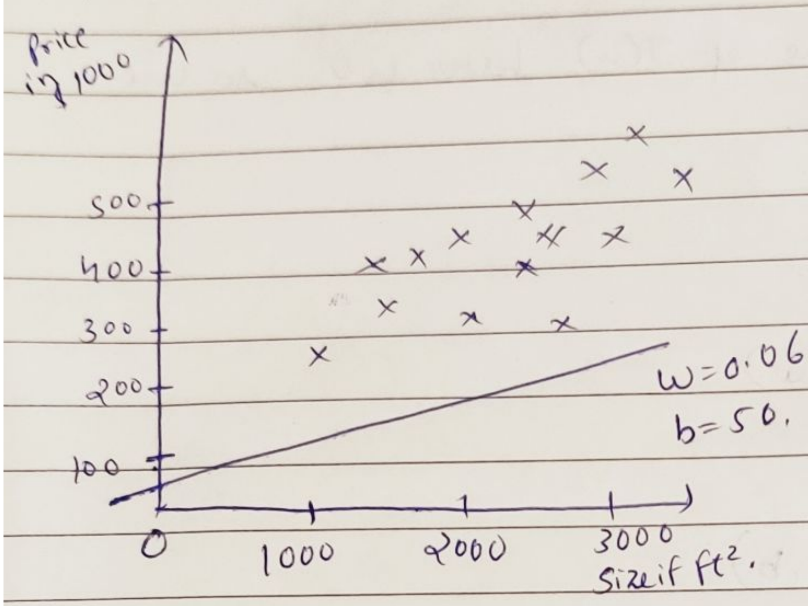
Parameters: w, b.

Cost Function: $J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} [f_{w,b}(x^{(i)}) - y^{(i)})^2$

Objective $\underset{w,b}{\text{minimize}}\ J(w,b)$

$f_{w,b}$.
function of x.

J.
function of w, b.



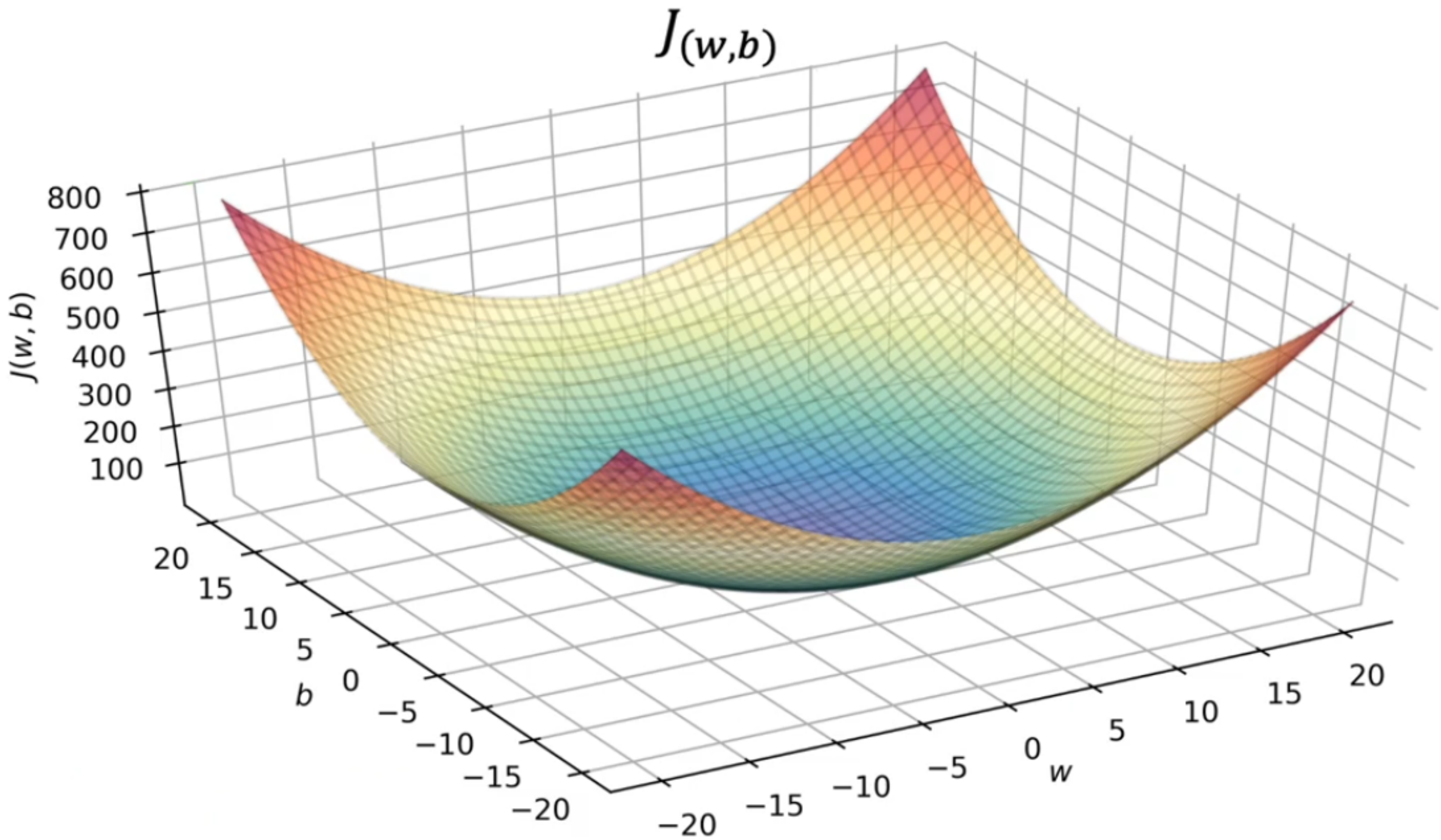
3D. Plot
look at the image.

Let $f_{w,b}(x) = 0.06x + 50$

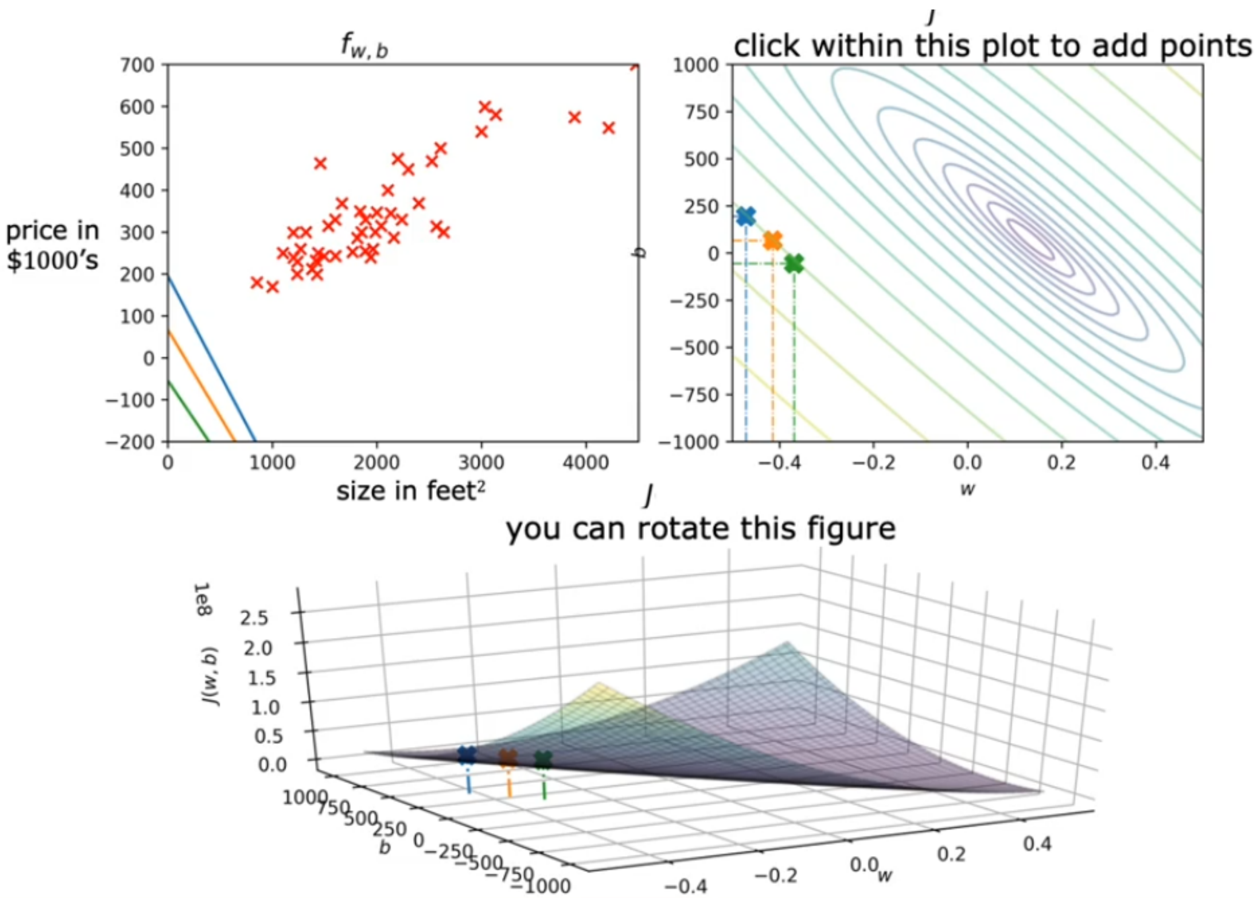
For J, we have two parameters, w, b. see screenshot for how cost function looks like for different value of w and b.
This needs visualization ~~on~~ in 3D.
This looks like hammock (soup bowl or curved plate).

As we vary w, b we get different values of J.
Every point in $J(w,b)$ ~~pa~~ plot represents the value of choice of w and b.

Contour plot is a horizontal slice of z axis (J values) for different values of w, b we have same J values.



$J_{(w,b)}$
$J(w, b)$
800
700
600
500
400
300
200
100
20
15
10
5
0
−5
−10
−15
−20
b
−20
−15
−10
−5
0
5
10
15
20
w



$f_{w,b}$
price in $1000's
size in feet²
J
click within this plot to add points
b
w
J
you can rotate this figure
1e8
J(w, b)
b
w

IX. Visualization Examples.

for $f(x) = -0.15x + 800$ } a bad function for
slope $= -0.15$, $b = 800$. } our use case.

$f(x) = 0x + 360 \rightarrow$

for $f(x) = 0.13x + 71$ } we have almost minimum value of J.

X. Quiz.

1. For a linear regression, the mode is $f_{w,b}(x) = wx + b$.

Which of the following are the ~~parts~~ inputs, or features, that are fed into the model and with which the model is expected to make a prediction?
a) w and b b) (x, y) c) m d) x.
I think answer is d) x.

2. For linear regression, if you find parameters w and b so that $J(w,b)$ is very close to zero, what can you conclude?
a) The selected values of the parameters w and b cause the algorithm to fit the training set really poorly.
b). —"— training set really well.