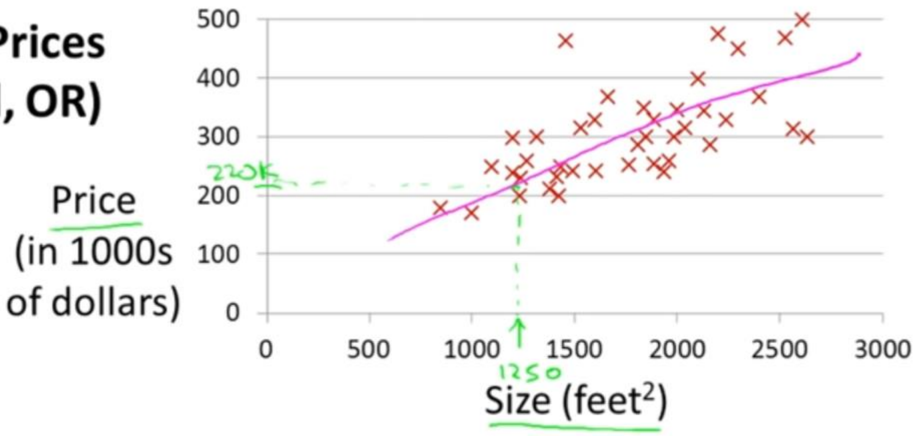
Linear Regression model Part-1

Housing Prices (Portland, OR)



Supervised Learning

Given the "right answer" for each example in the data.

Regression Problem

Predict real-valued output

Classification: Discrete-valued output

Andrew Ng

Terminologies

Training set of housing prices (Portland, OR)

Size in feet² (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...

m = 47

Notation:

- **m** = Number of training examples
- **x**'s = "input" variable / features
- **y**'s = "output" variable / "target" variable

(x, y) - one training example

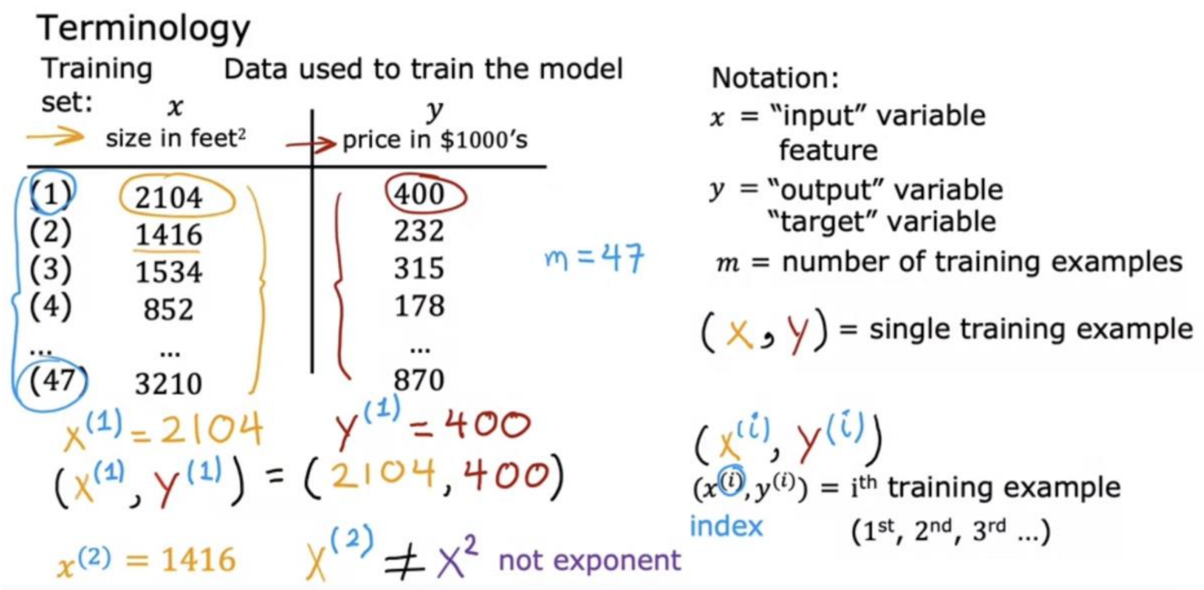
$(x^{(i)}, y^{(i)})$ - i^{th} training example

$x^{(1)} = 2104$

$x^{(2)} = 1416$

$y^{(1)} = 460$

Andrew Ng

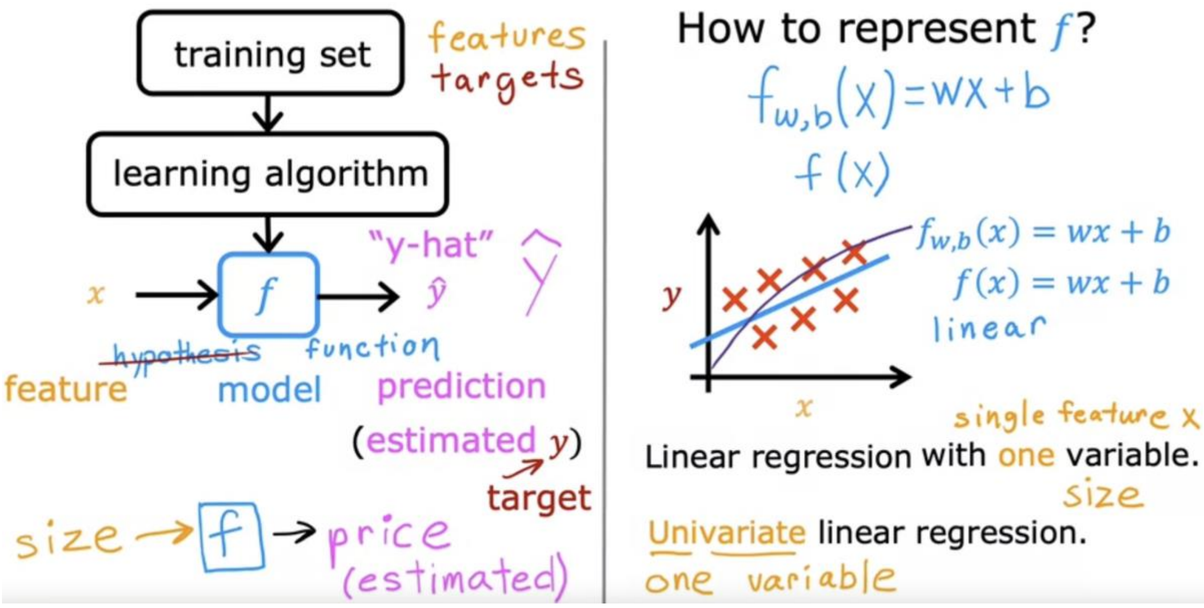


For linear regression, the model is represented by $f_{w,b}(x) = wx + b$. Which of the following is the output or "target" variable?

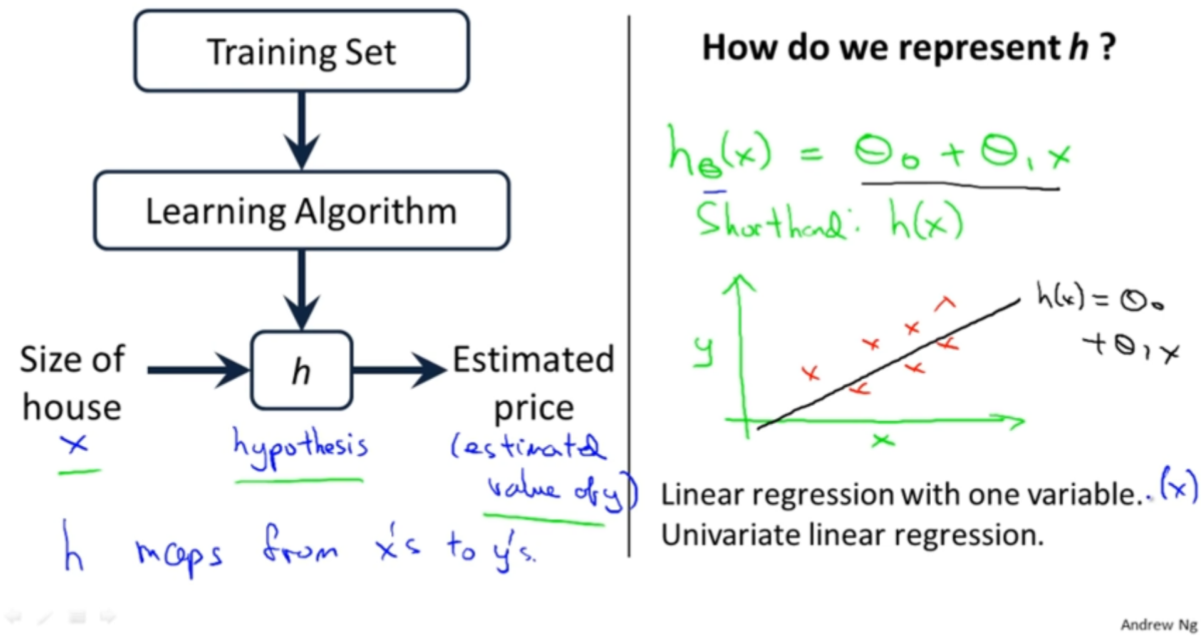
- ( ) $\hat{y}$.
- ( ) m
- (•) y
- ( ) x

Correct
y is the true value for that training example, referred to as the output variable, or "target".

Linear Regression model Part-2



The function f is also know as function/hypothesis/model.

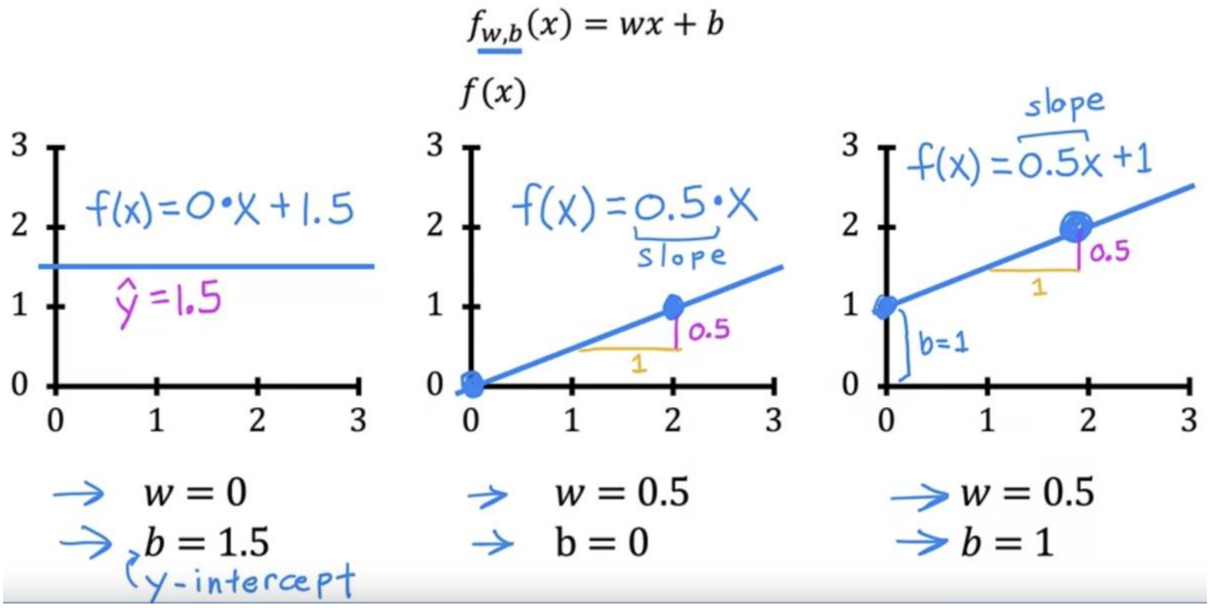


Model Representation

Refer 3_Model_Representation_Solution.ipynb file

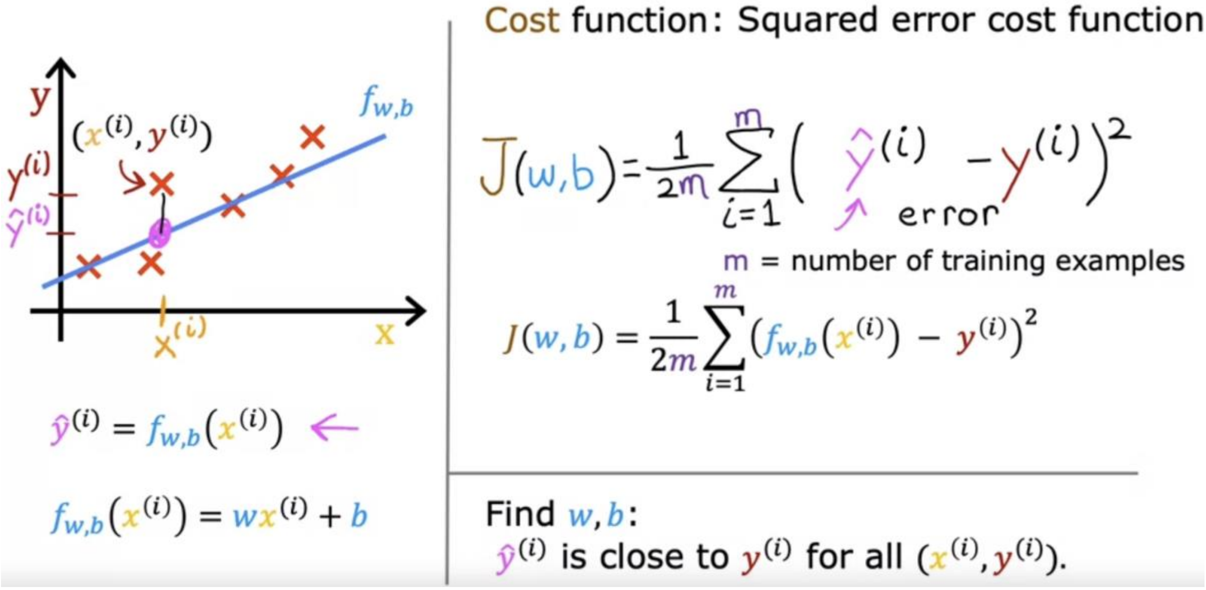
Cost Function

Lets see the effect of slope and constant



Cost function/Squared error function/Residual Sum of Squares

Average square of the distance between the observed and the predicted value



The cost function used for linear regression is

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

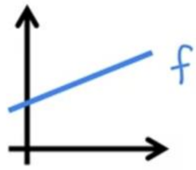
Which of these are the parameters of the model that can be adjusted?

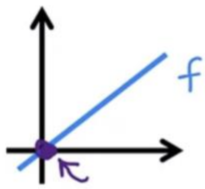
- (●) w and b
- ( ) $f_{w,b}(x^{(i)})$
- ( ) w only, because we should choose b=0
- ( ) $\hat{y}$

✓ **Correct**

Cost function Intuition

model:	simplified
$f_{w,b}(x) = wx + b$	$f_w(x) = wx$ $b = \varnothing$
parameters: w, b	w
cost function: $J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})^2$	$J(w) = \frac{1}{2m}\sum_{i=1}^{m}(\underbrace{f_w(x^{(i)})}_{wx^{(i)}} - y^{(i)})^2$
goal: $\underset{w,b}{\text{minimize}}\, J(w,b)$	$\underset{w}{\text{minimize}}\, J(w)$

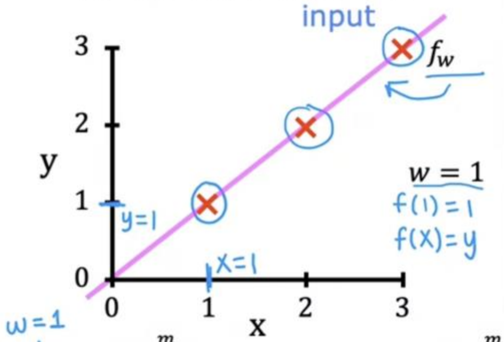




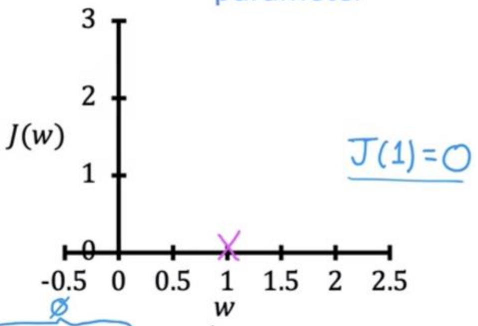
Currently put b=0

w=1

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



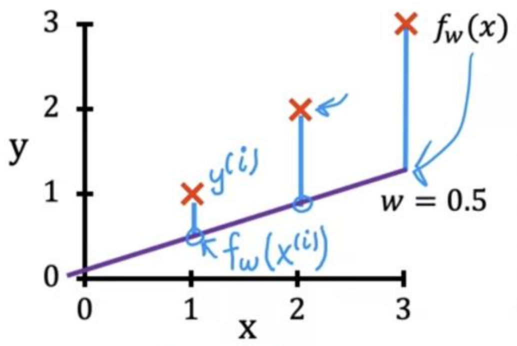
$J(w)$ (function of w) parameter



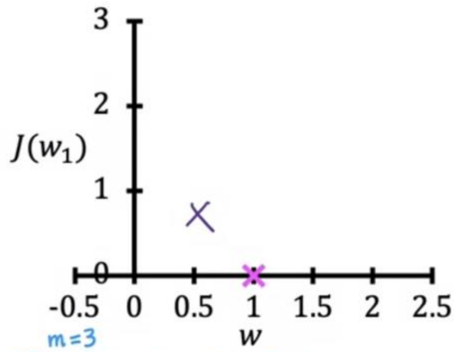
$$\underset{w=1}{J(w)} = \frac{1}{2m}\sum_{i=1}^{m}(\underbrace{f_w(x^{(i)})}_{wx^{(i)}} - y^{(i)})^2 = \frac{1}{2m}\sum_{i=1}^{m}(\overbrace{wx^{(i)} - y^{(i)}}^{\varnothing})^2 = \frac{1}{2m}(0^2 + 0^2 + 0^2) = 0$$

w=0.5

$f_w(x)$ (function of x)

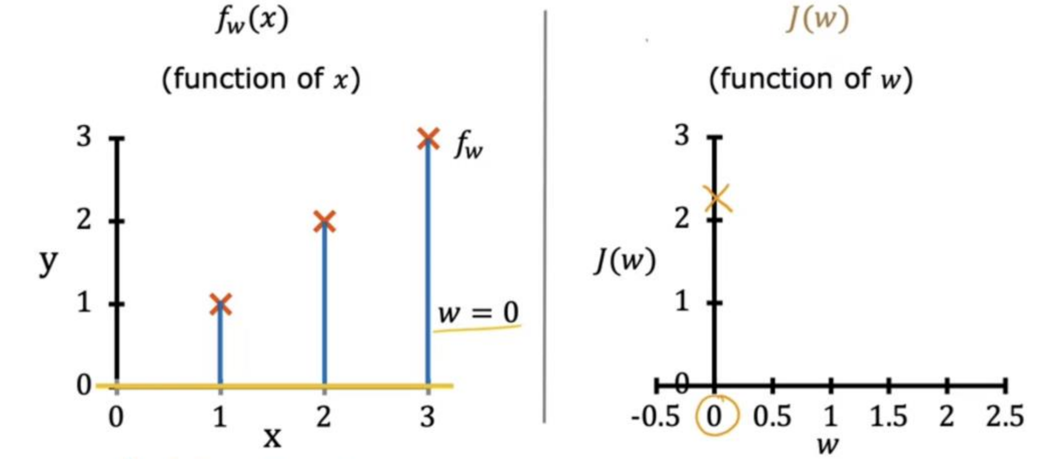


$J(w)$ (function of w)



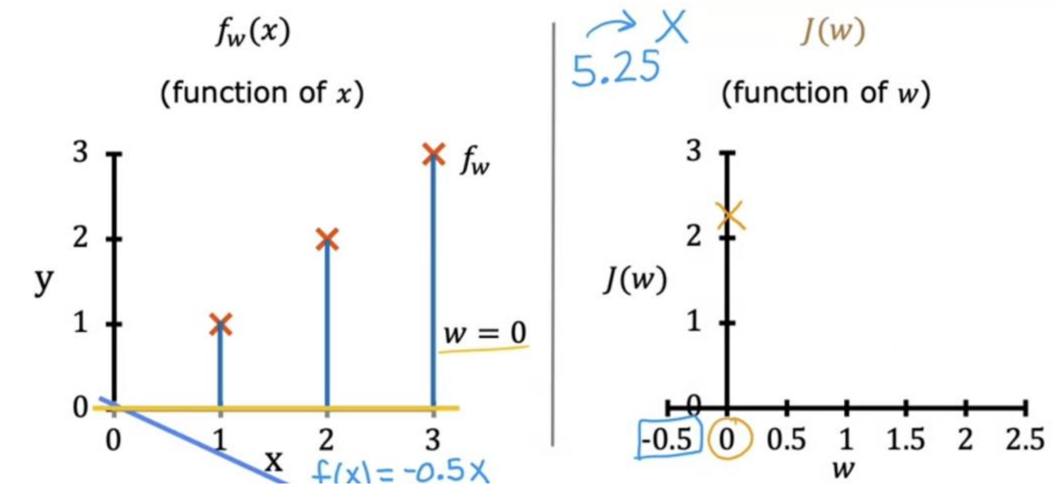
$$J(0.5) = \frac{1}{2m}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right] = \frac{1}{2\times 3}[3.5] = \frac{3.5}{6} \approx 0.58 \quad (m=3)$$

w=0

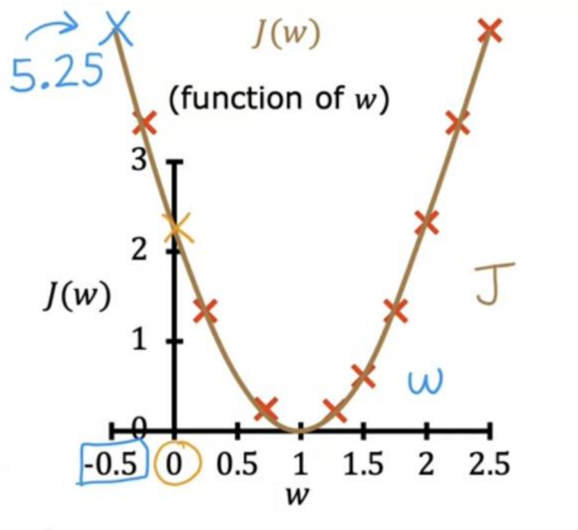


$$J(0) = \frac{1}{2m}\left(1^2 + 2^2 + 3^2\right) = \frac{1}{6}[14] \approx 2.3$$

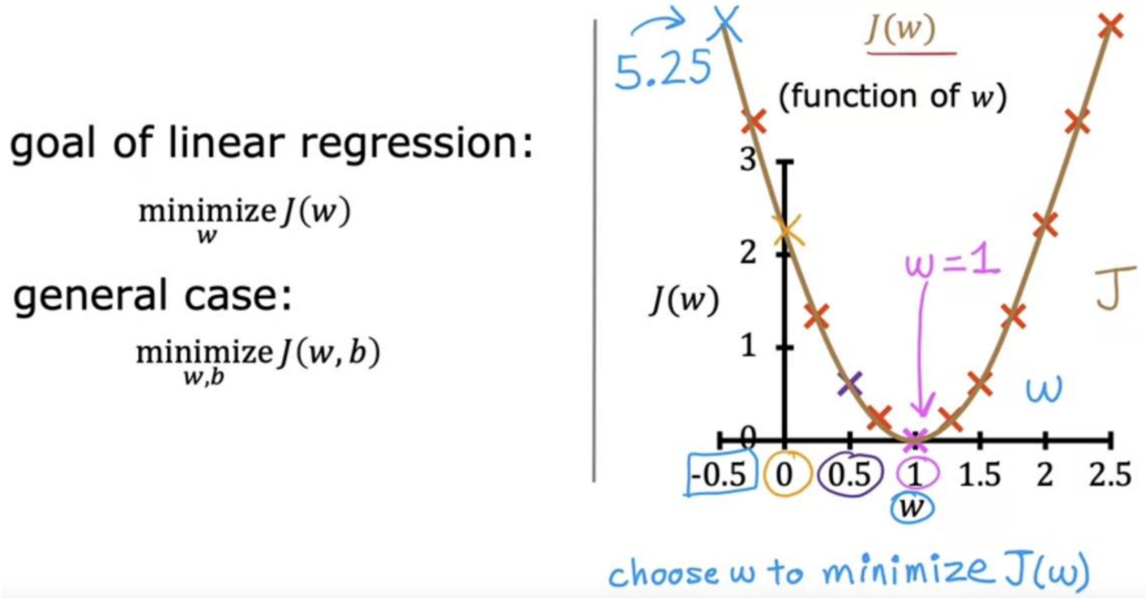
w=-0.5



As we move on with different values of w, and the corresponding J value we can obtain this curve



So we want to choose w, which minimises J(w)



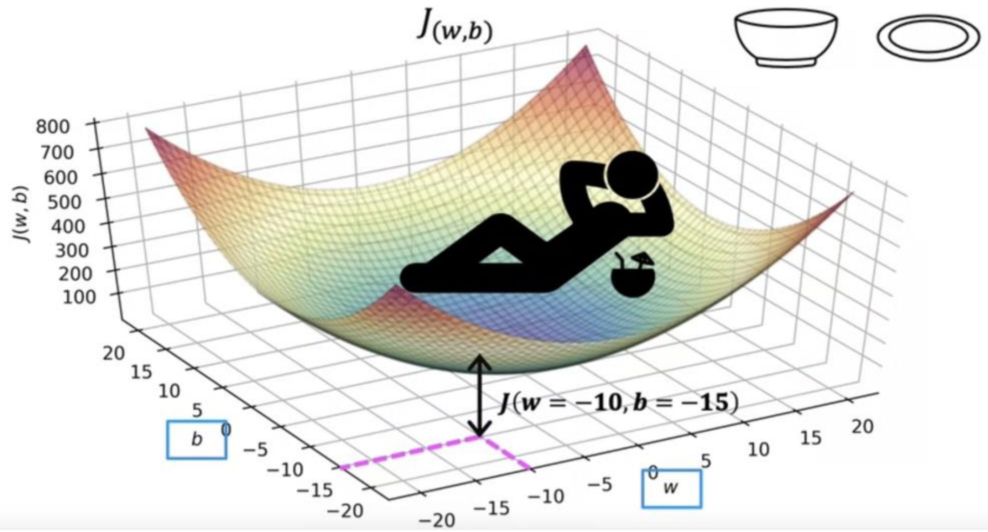
When does the model fit the data relatively well, compared to other choices for parameter w?

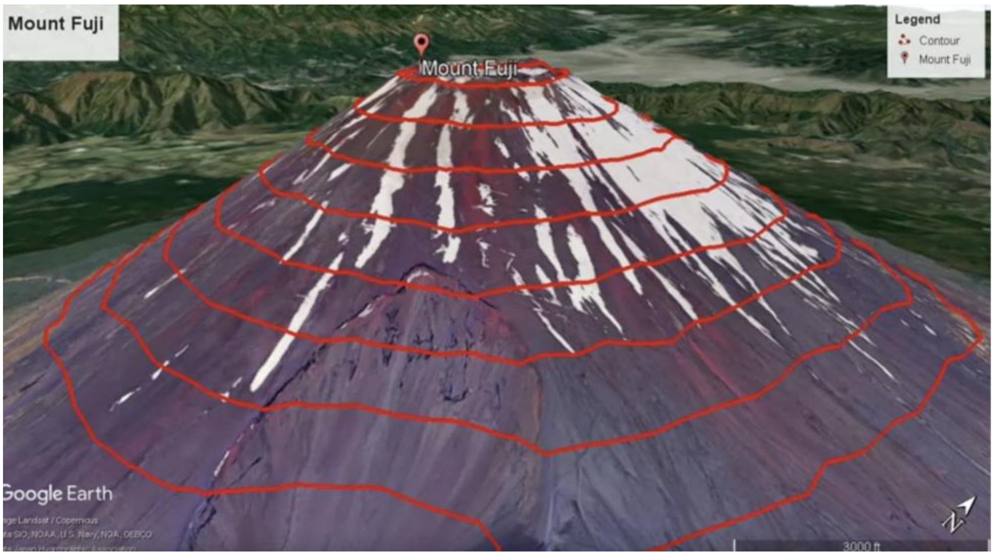
- ○ When $f_w(x)$ is at or near a minimum for all the values of x in the training set.
- ○ When w is close to zero.
- ◉ When the cost J is at or near a minimum.
- ○ When x is at or near a minimum.

✓ **Correct**

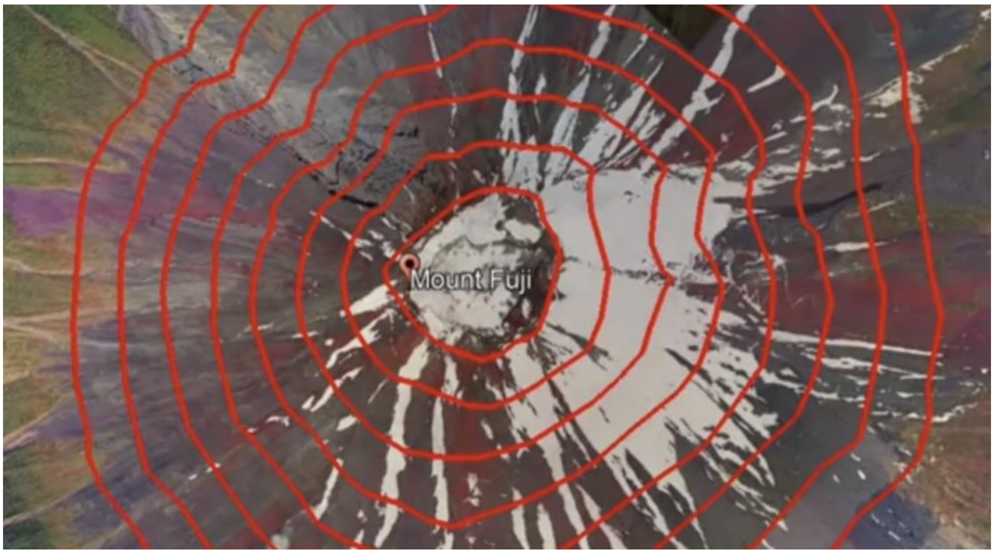
When the cost is relatively small, closer to zero, it means the model fits the data better compared to other choices for w and b.

Visualize the Cost function

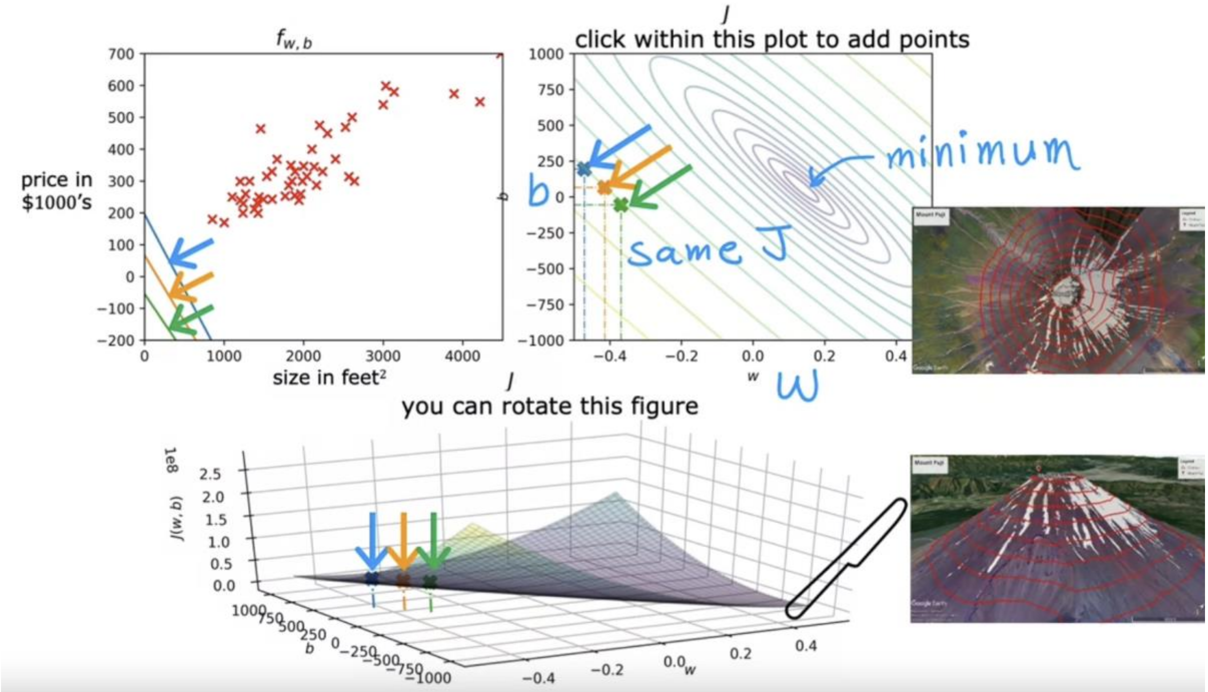




3D bowl shaped surface plot → 2D contour plots



View directly above the mountain.



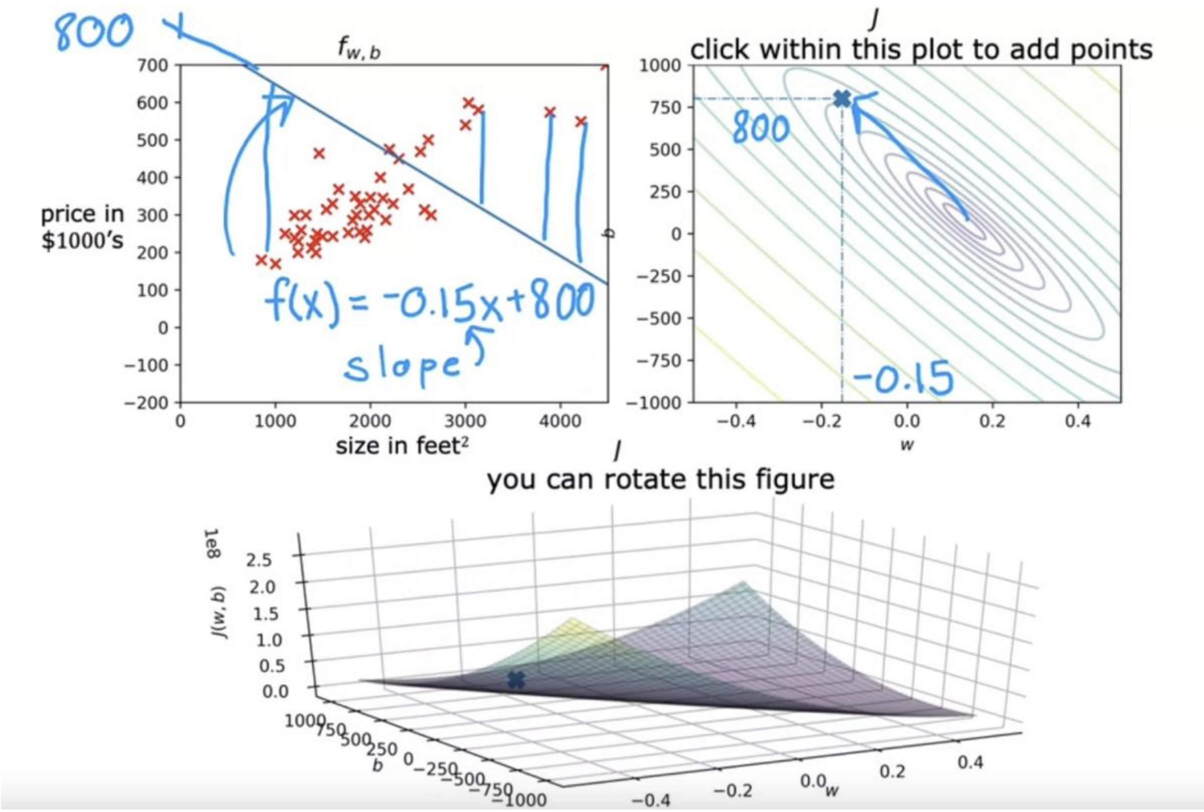
All of these three points have the same value for the cost function J, even though they have difference values for w and b(bottom figure)

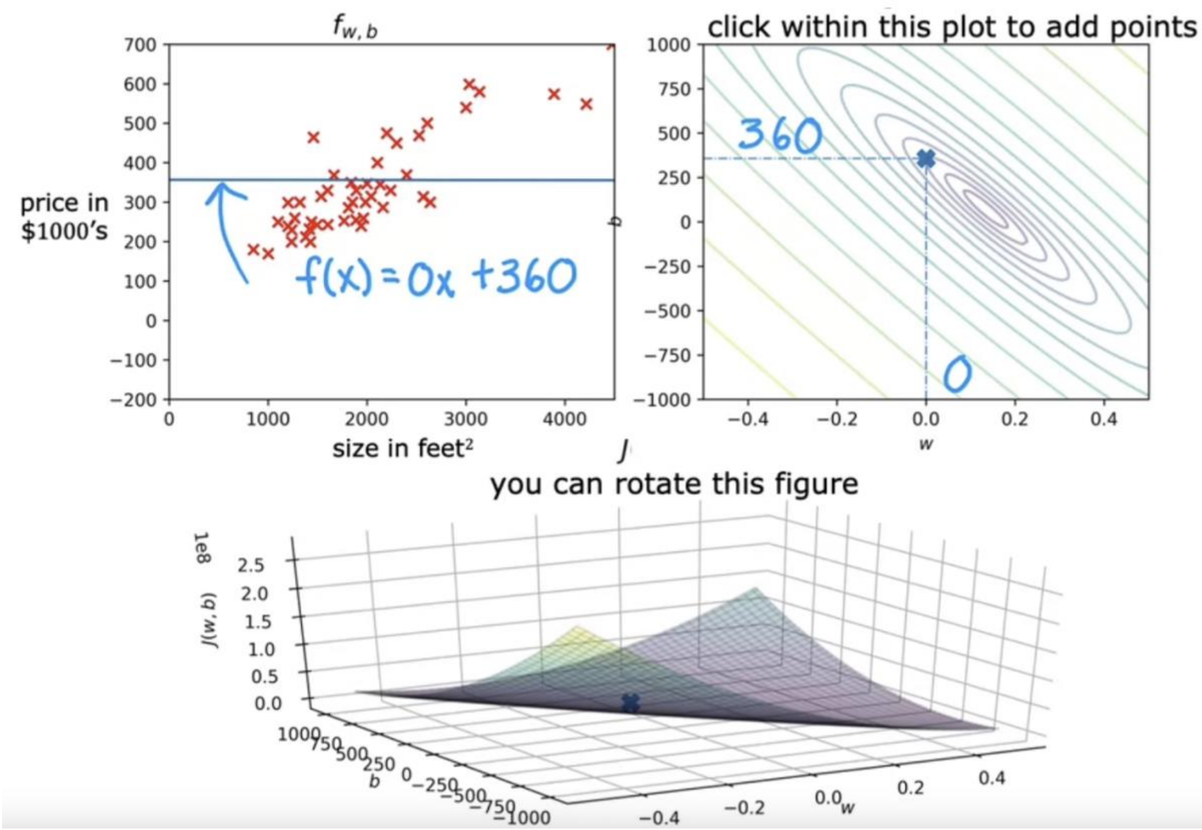
All these three points corresponds to a different function (upper left figure)

Contour plots are convenient way to visualize the 3D cost function J.

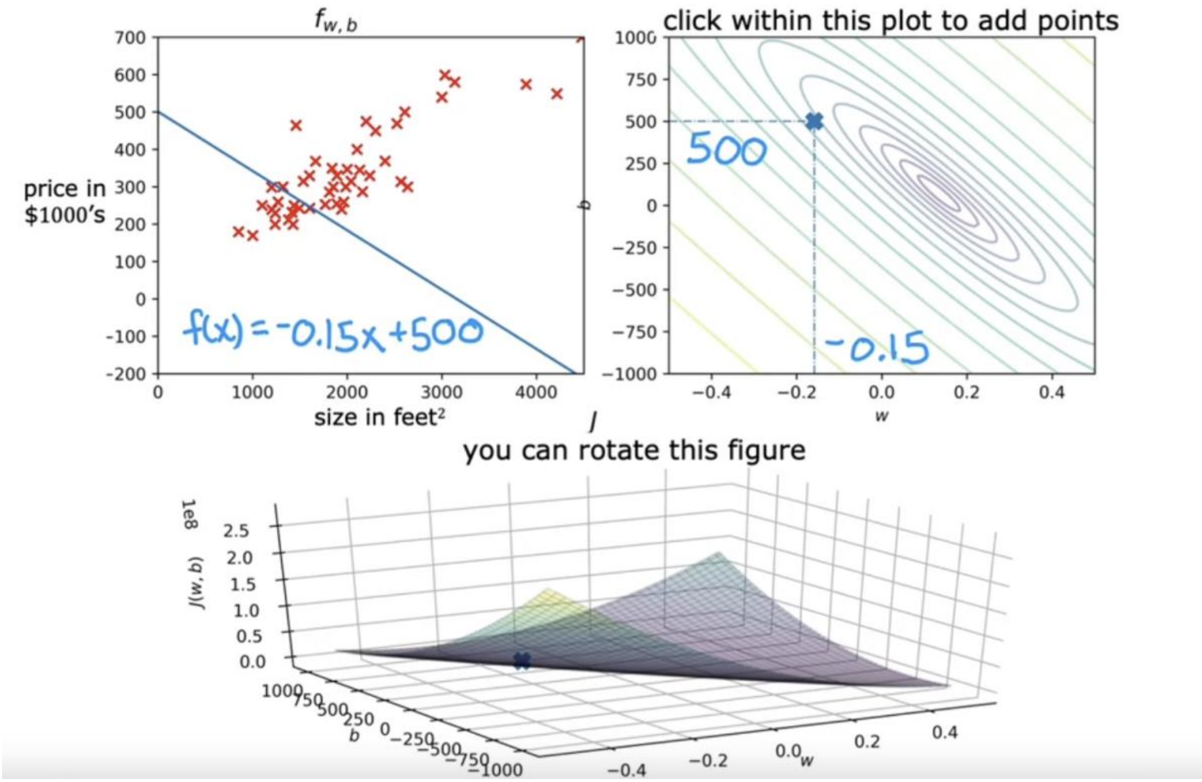
Minimum is the centre of the smallest eclipse.

Visualization example

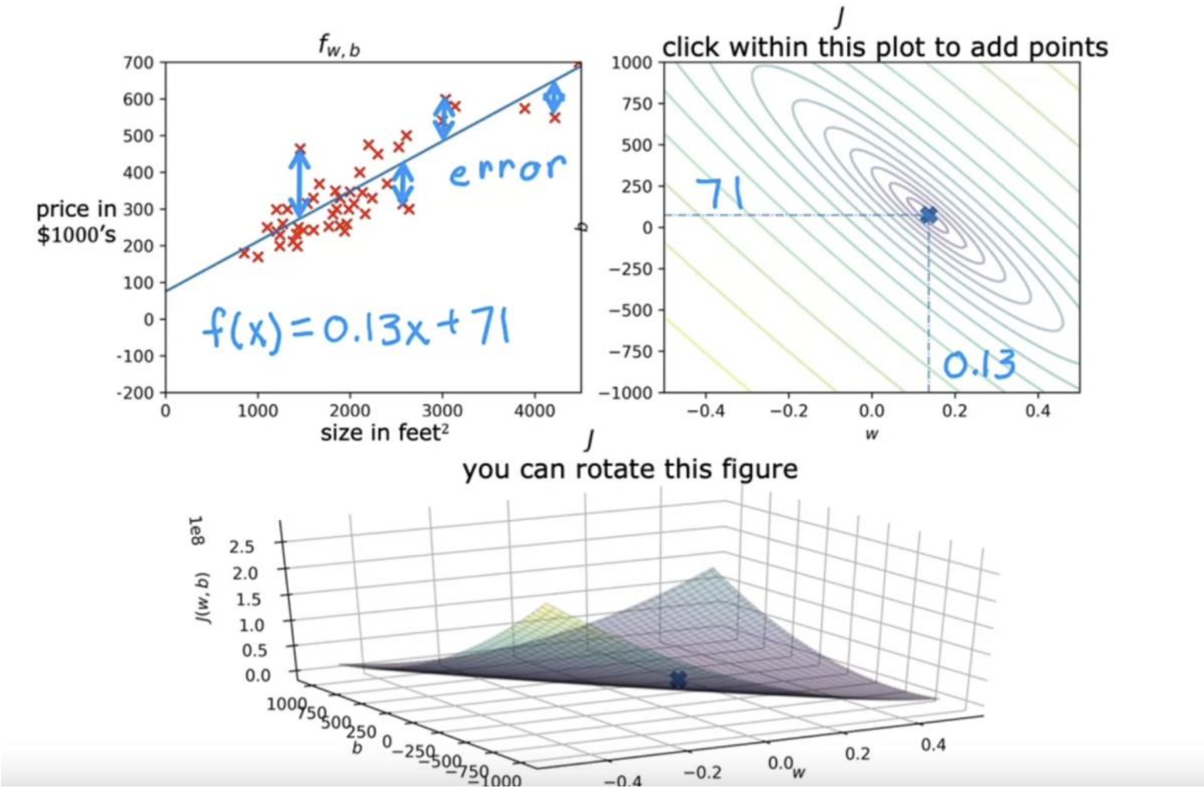




$f_{w,b}$
price in $1000's
size in feet²
f(x)=0x +360
click within this plot to add points
360
0
w
b
J
you can rotate this figure
1e8
J(w, b)
b
w



$f_{w,b}$
price in $1000's
size in feet²
f(x)=-0.15x+500
click within this plot to add points
500
-0.15
w
b
J
you can rotate this figure
1e8
J(w, b)
b
w



This is pretty close to the centre of the small eclipse.
How better the fit lines corresponds to the points on the graph of j that are closer to the minimum possible cost for the cost function J(w,b)

Quiz

1. Which are the two common types of supervised learning? (Choose two) — 1 / 1 point

- [x] Classification

  Correct
  Classification predicts from among a limited set of categories (also called classes). These could be a limited set of numbers or labels such as "cat" or "dog".

- [ ] Clustering
- [x] Regression

  Correct
  Regression predicts a number among potentially infinitely possible numbers.

2. — 1 / 1 point

Which of these is a type of unsupervised learning?

- ○ Regression
- ○ Classification
- ◉ Clustering

  Correct
  Clustering groups data into groups or clusters based on how similar each item (such as a hospital patient or shopping customer) are to each other.

References

[Linear regression model part 1 | Coursera]