

Week 1

Regression Model

[Mean squared error cost function formula \(Linear regression\)](#)

[Function overview](#)

[Simplified function view \(Without constant b\):](#)

[Visualising linear regression models](#)

[Gradient Descent](#)

[Update w and b simultaneously](#)

[What the derivative term is doing \(on J\(w\)\)](#)

[What if gradient is 0? \(Reached local minima\)](#)

[When w is close to the local minima](#)

[Gradient descent formula](#)

[For w](#)

[for b](#)

Regression Model

Mean squared error cost function formula (Linear regression)

Cost function: Squared error cost function

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^m \left(\underset{\substack{\uparrow \\ \text{error}}}{\hat{y}^{(i)}} - y^{(i)} \right)^2$$

m = number of training examples

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

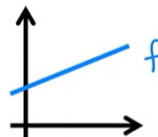
Function overview

model:

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

parameters:

$$w, b$$



cost function:

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

goal:

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

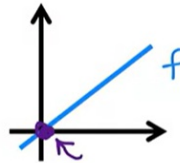
Simplified function view (Without constant b):

simplified

$$f_w(x) = wx$$

$$b = 0$$

w

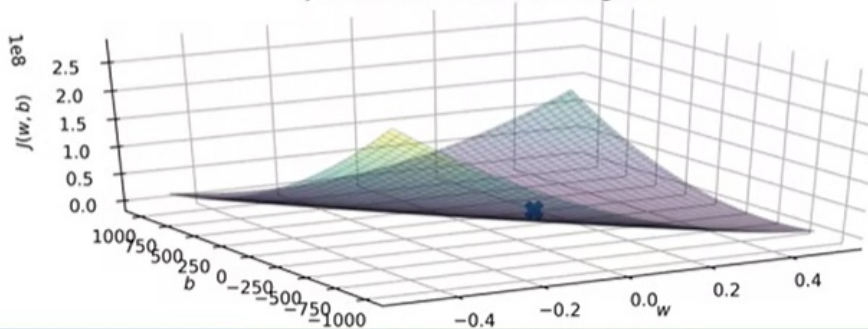
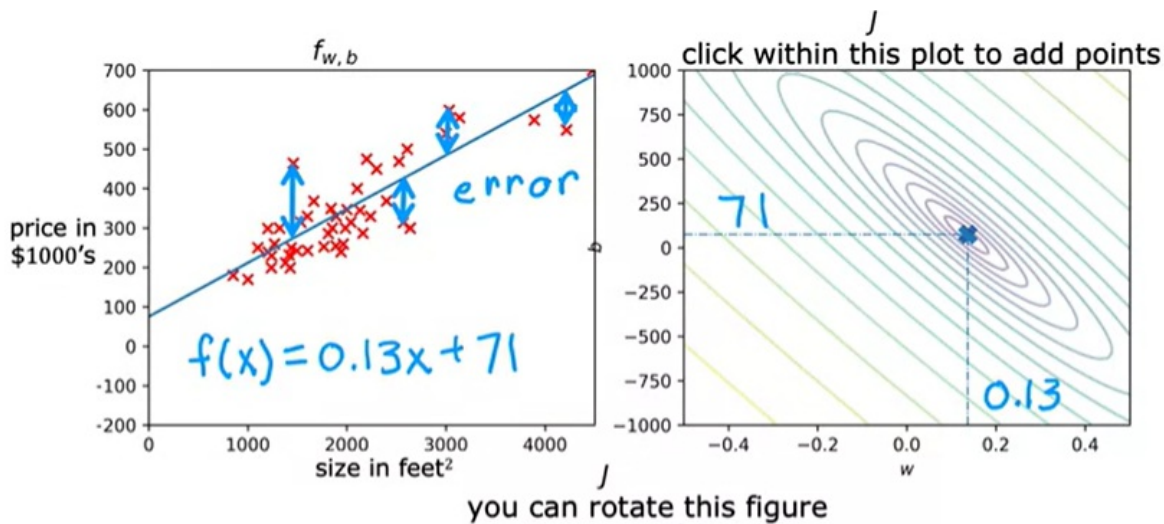


$$J(w) = \frac{1}{2m} \sum_{i=1}^m (f_w(x^{(i)}) - y^{(i)})^2$$

minimize $J(w)$

$$wx^{(i)}$$

Visualising linear regression models

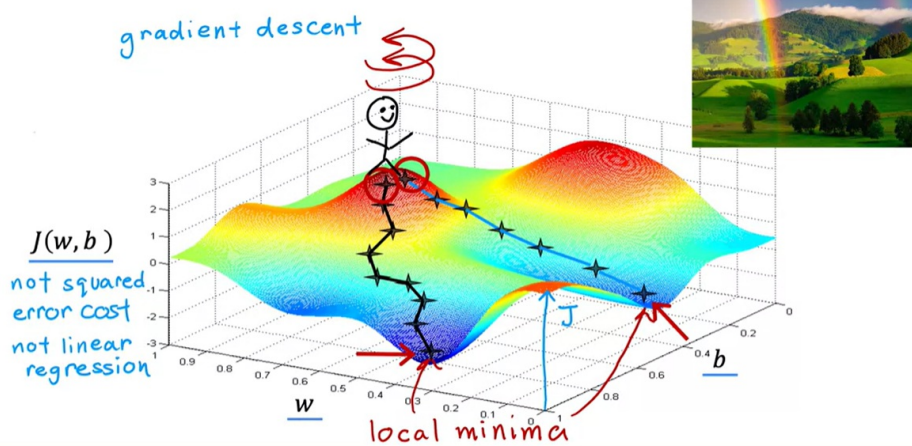


NB: Contours on the upper right represent the surface in which the costs are at the same height $\{J(w,b)\}$

Gradient Descent

Visualisation:

Idea → Taking the biggest small steps to go down the fastest (Drive errors down)



Update w and b simultaneously

Gradient descent algorithm

Repeat until convergence

$$\begin{cases} \underline{w} = w - \alpha \frac{\partial}{\partial w} J(w, b) \\ \underline{b} = b - \alpha \frac{\partial}{\partial b} J(w, b) \end{cases}$$

Learning rate
Derivative

Simultaneously
update w and b

Assignment

$$a = c$$

$$a = a + 1$$

Code

Truth assertion

$$a = c$$

$$a = a + 1$$

Math

$$a == c$$

Correct: Simultaneous update

$$\text{tmp_w} = w - \alpha \frac{\partial}{\partial w} J(w, b)$$

$$\text{tmp_b} = b - \alpha \frac{\partial}{\partial b} J(\text{tmp_w}, b)$$

$$w = \text{tmp_w}$$

$$b = \text{tmp_b}$$

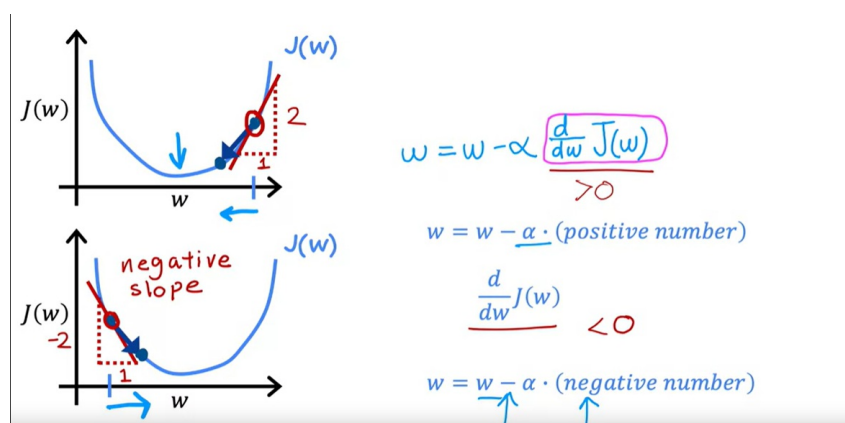
Incorrect

$$\text{tmp_w} = w - \alpha \frac{\partial}{\partial w} J(w, b)$$

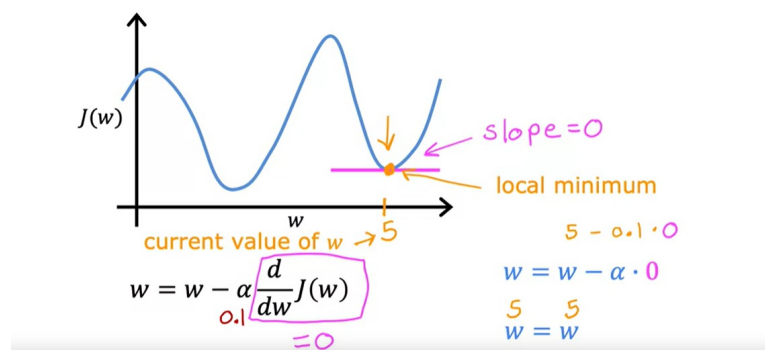
$$\begin{aligned} \underline{w} &= \text{tmp_w} \\ \text{tmp_b} &= b - \alpha \frac{\partial}{\partial b} J(\underline{w}, b) \\ \underline{b} &= \text{tmp_b} \end{aligned}$$

- $\alpha \rightarrow$ learning rate. Controls how big of a step you take downhill (Usually small)
- Derivative determines the rate
- Cannot simply use $w = w - \alpha \frac{d}{dw} J(w, b)$, as you need the old w value for tmp_b (see left)

What the derivative term is doing (on $J(w)$)

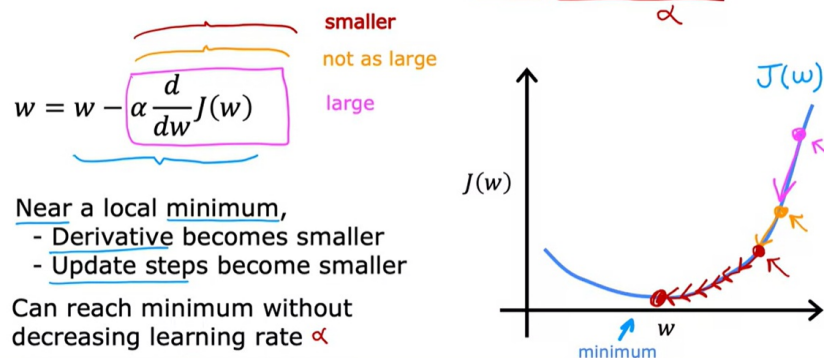


What if gradient is 0? (Reached local minima)



When w is close to the local minima

Can reach local minimum with fixed learning rate α



Gradient descent formula

Linear regression model

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

Cost function

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

Gradient descent algorithm

repeat until convergence {

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

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

}

For w

(Optional)

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

Notice that the 2 (from $1/2m$) is used to make the equation neater (from the derivative of the squared term).

for b

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

no $x^{(i)}$