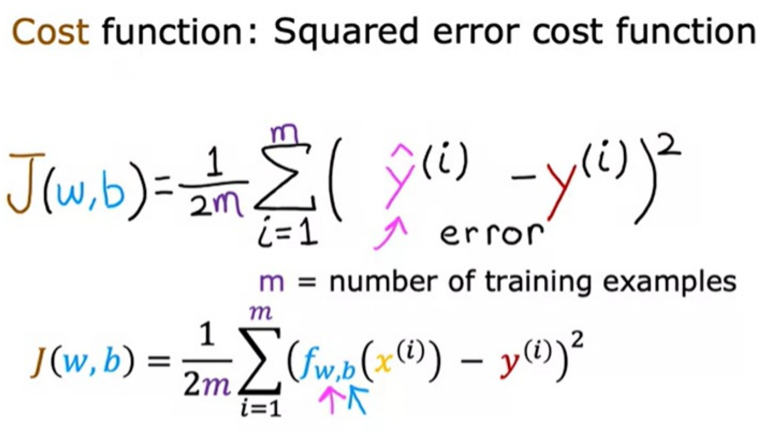
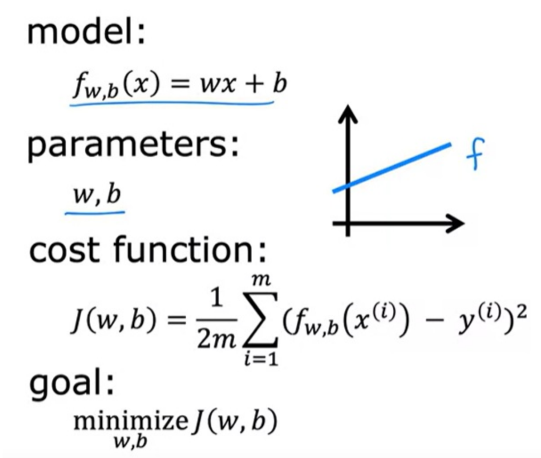
Week 1

Regression Model

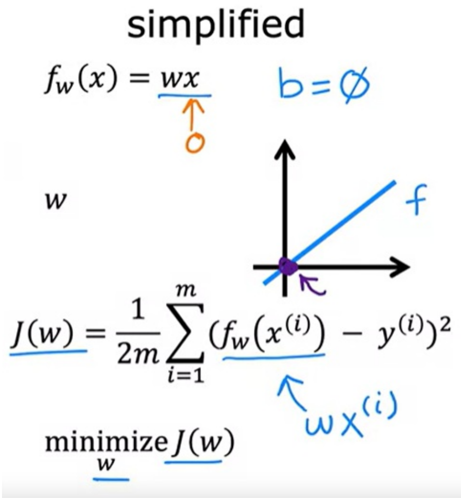
Mean squared error cost function formula (Linear regression)



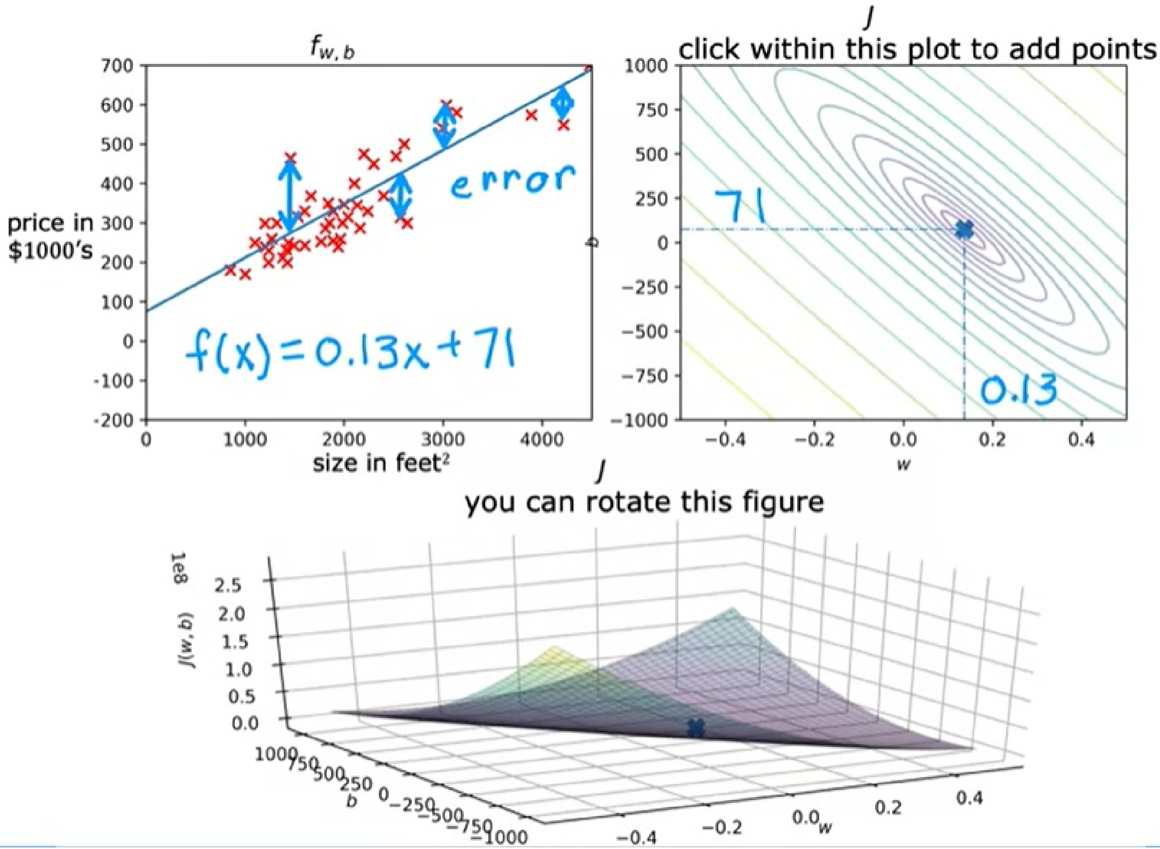
Function overview



Simplified function view (Without constant b):



Visualising linear regression models

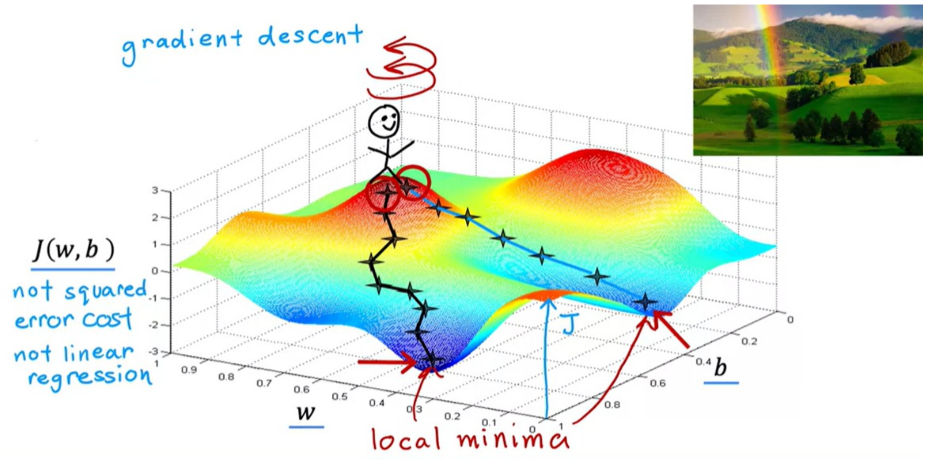


NB: Contours on the upper right represent the surface in which the costs are at the same height {J(w,b)}

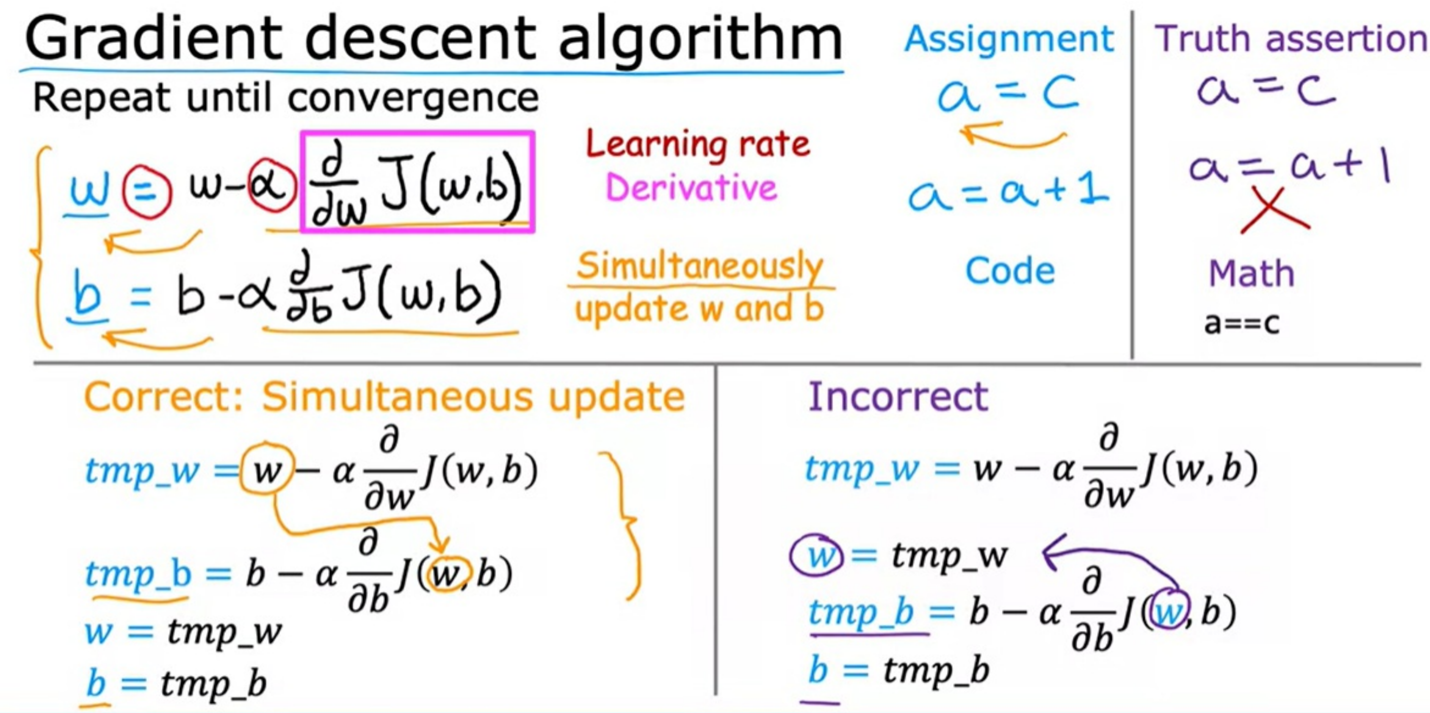
Gradient Descent

Visualisation:

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

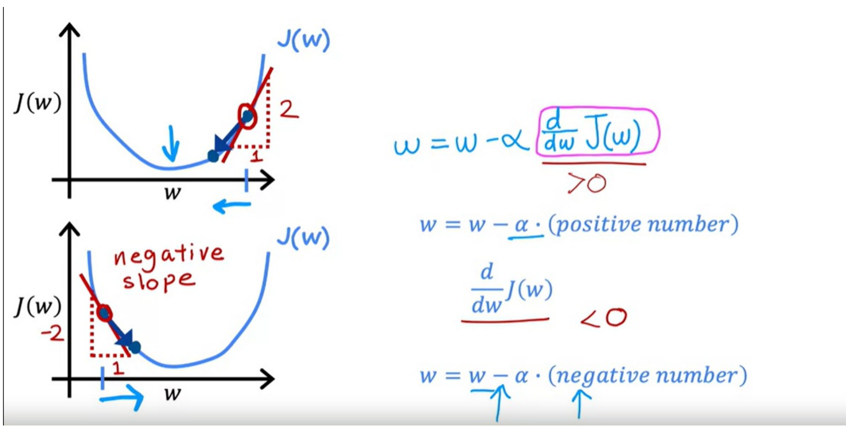


Update w and b simultaneously

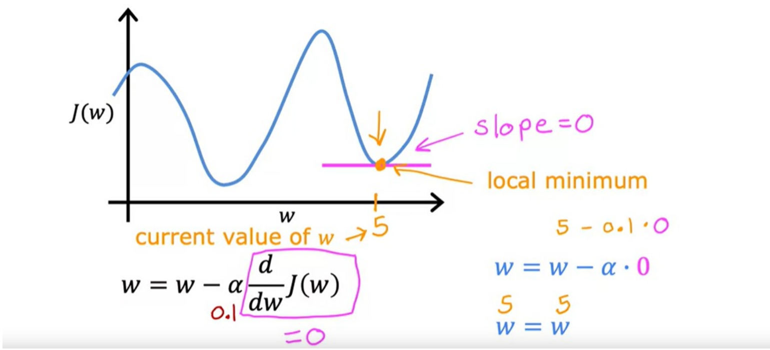


- α → learning rate. Controls how big of a step you take downhill (Usually small)
- Derivative determines the rate
- Cannot simply use w = w-α d/dw(J(w,b)), as you need the old w value for tmp_b (see left)

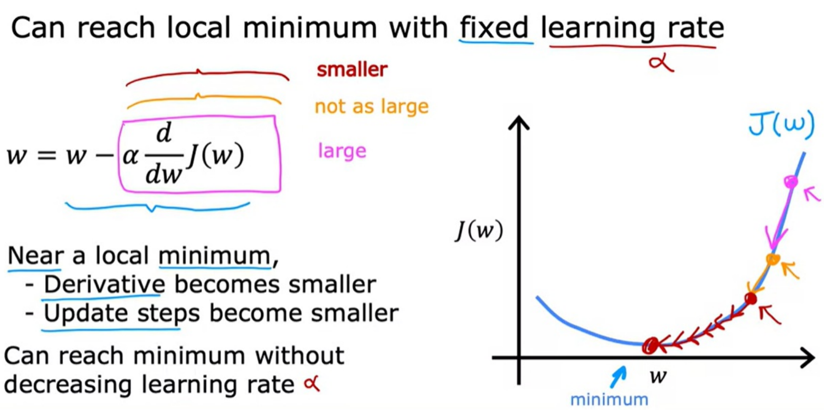
What the derivative term is doing (on J(w))



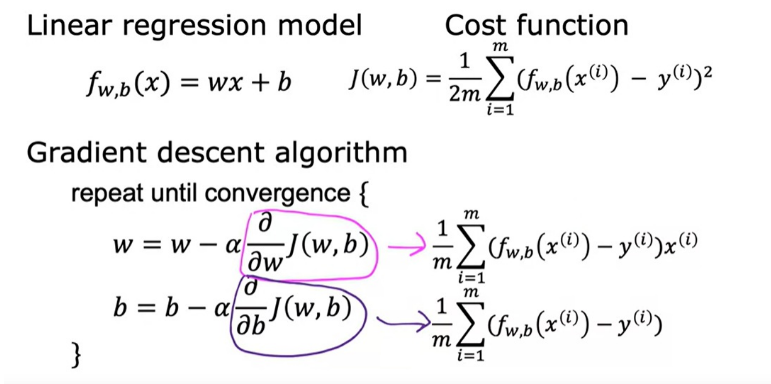
What if gradient is 0? (Reached local minima)



When w is close to the local minima



Gradient descent formula



For w

(Optional)

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

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

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

for b

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

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

no $x^{(i)}$