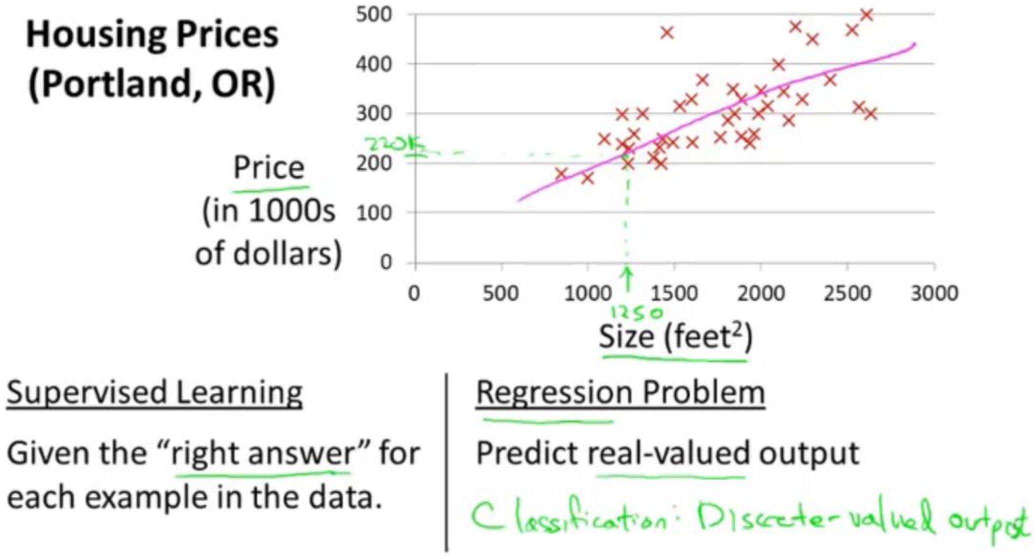
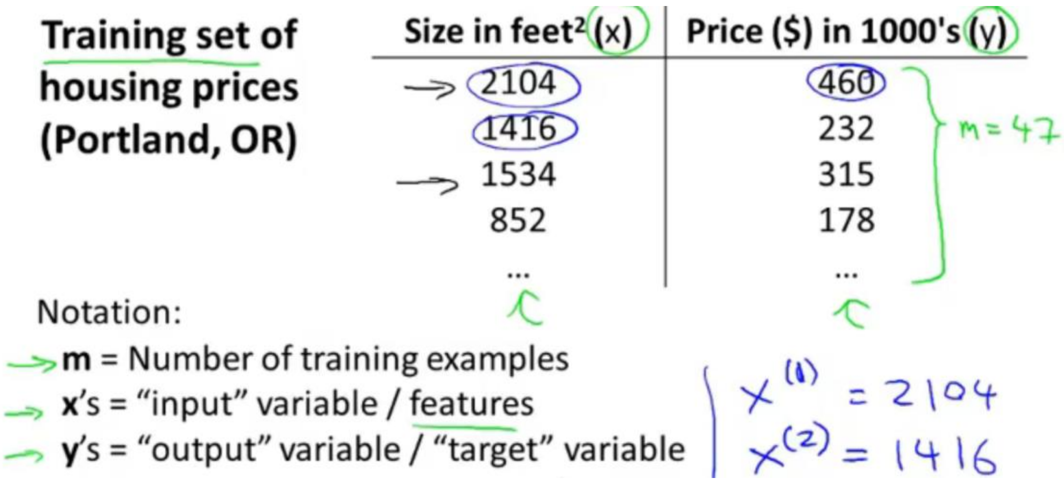
Linear regression with one variable

1. Model representation

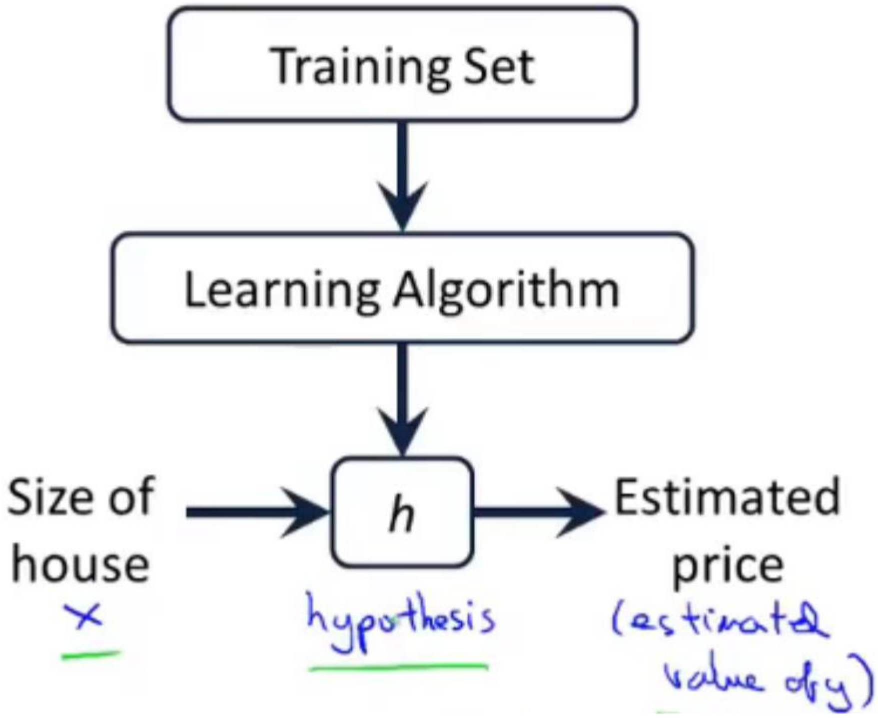


Training set: data set given

(x, y): one training example; $(x^{(i)}, y^{(i)})$: i^{th} training example;



h maps from x's to y's

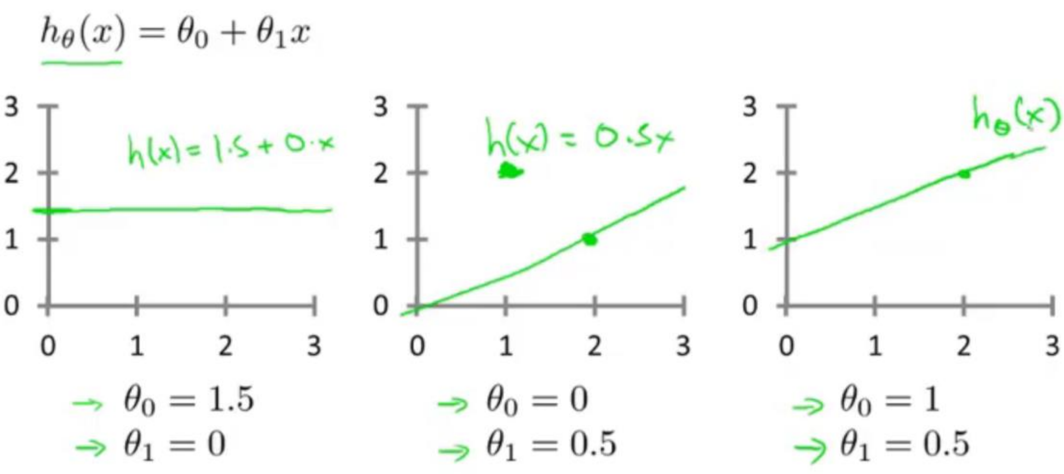


How do we represent h?

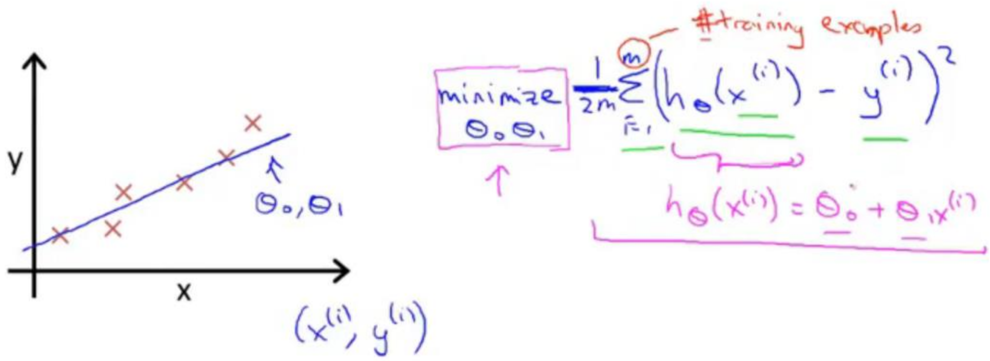
$$h_\theta(x) = \theta_0 + \theta_1 x$$
$$shorthand: h(x)$$

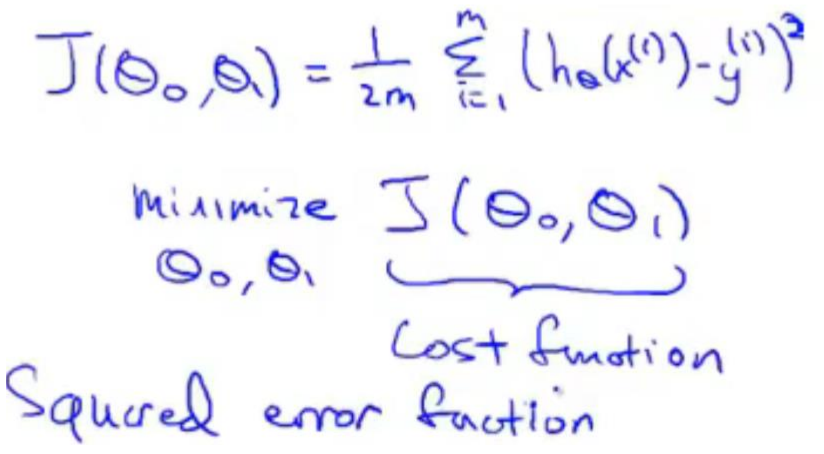
Linear regression with one variable.
Univariate linear regression.

2. Cost function

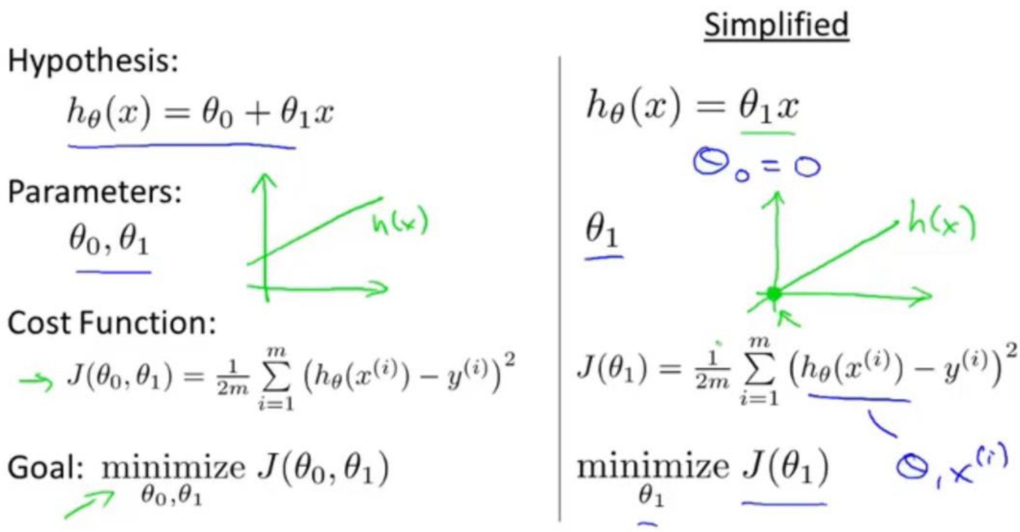


Idea: Choose theta 0, theta 1 so that h(x) is close to y for out training examples (x, y)

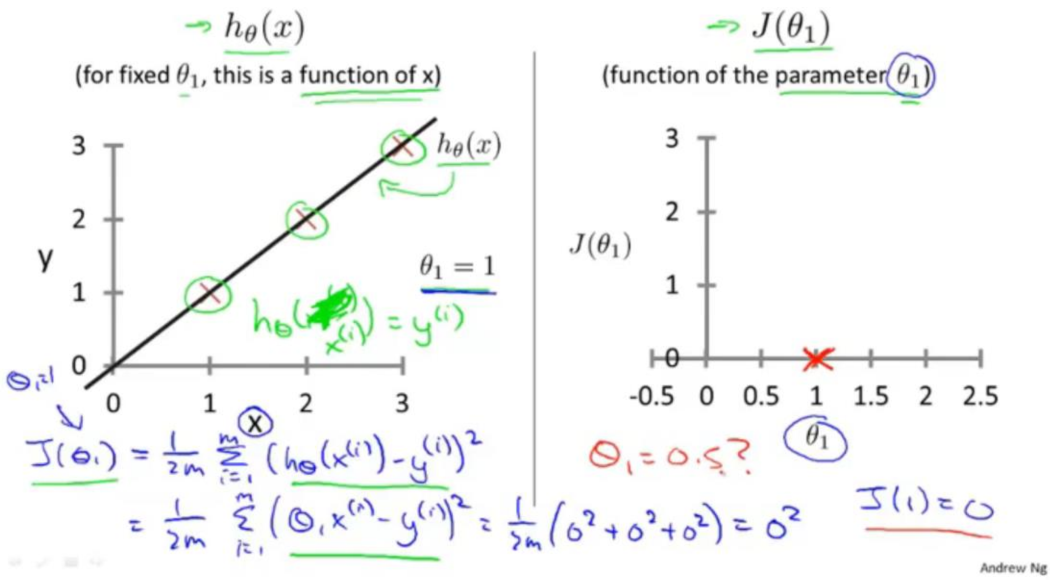




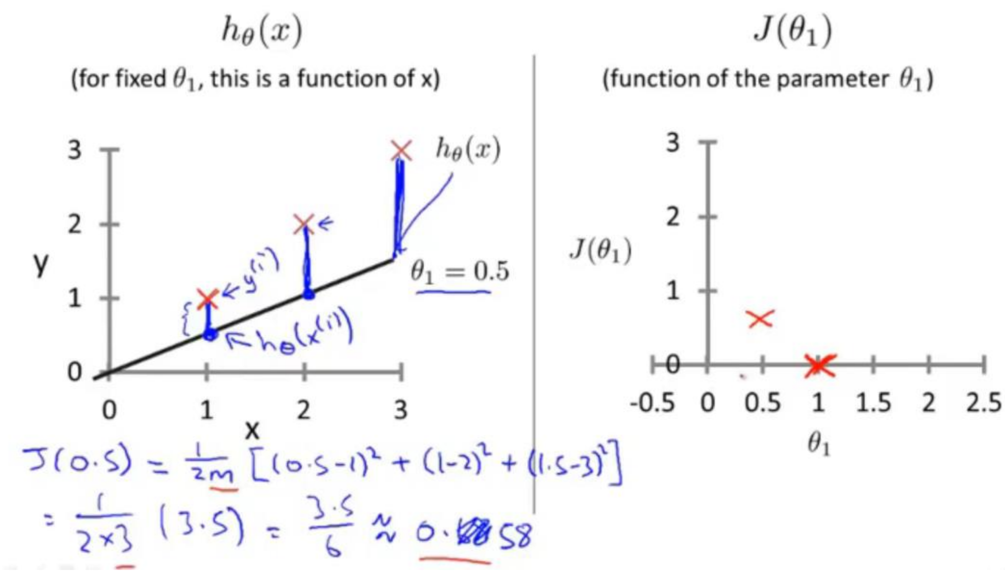
3. Cost function intuition I



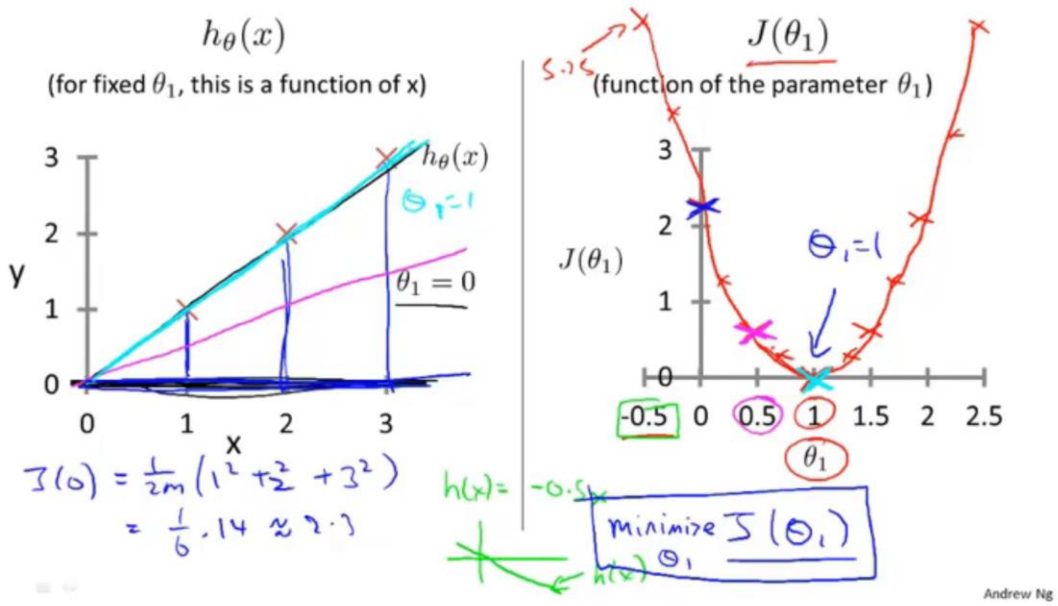
theta 1 = 1:



theta 1 = 0.5:

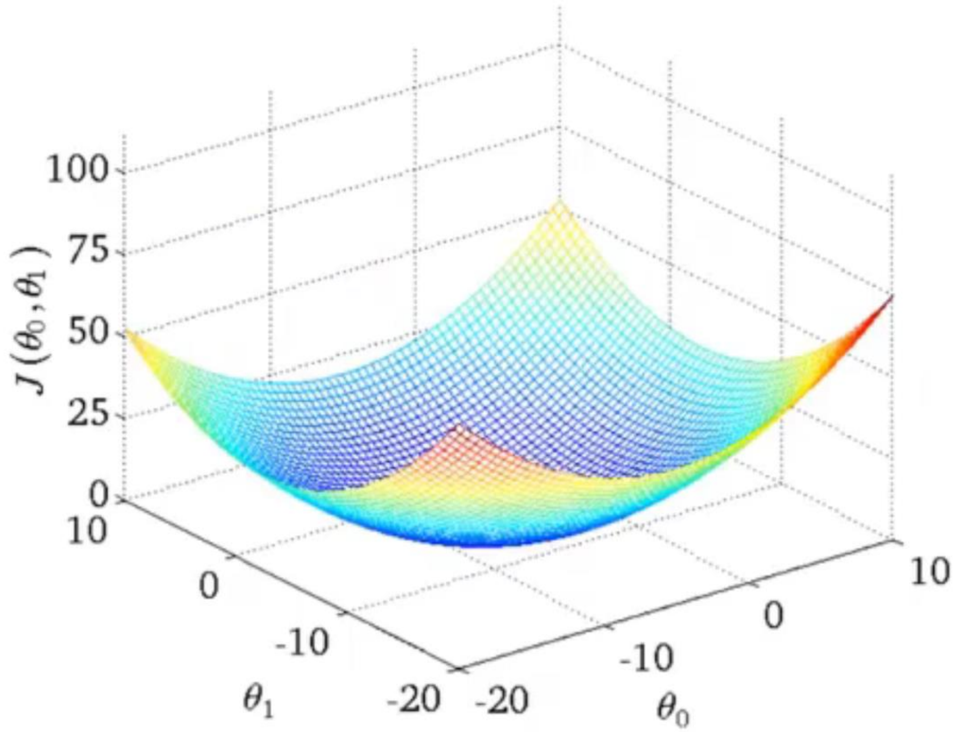


theta 1 = 0:

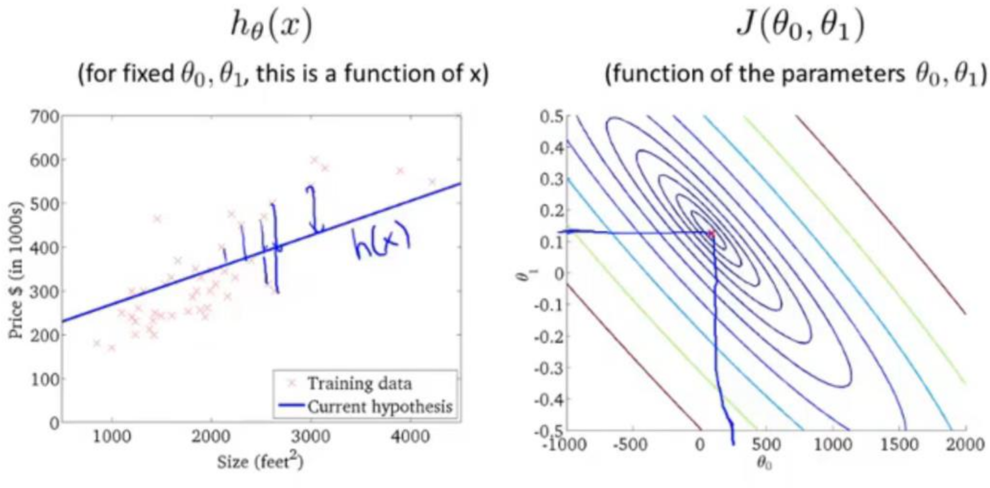


4. Cost function intuition II

bowl shape(3-D surface plot)

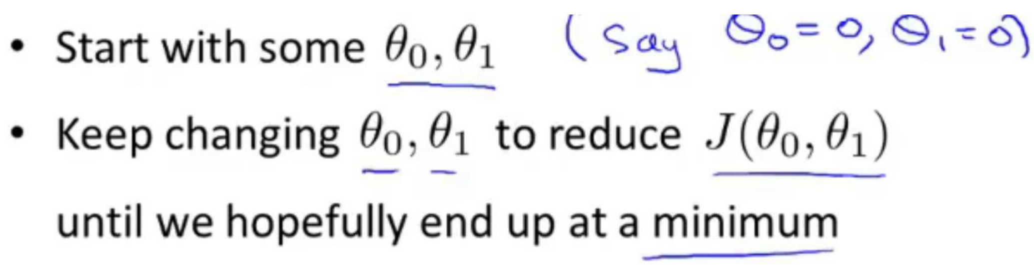


The minimum, the bottom of the bowl is this point right there, this middle of these concentric ellipses.

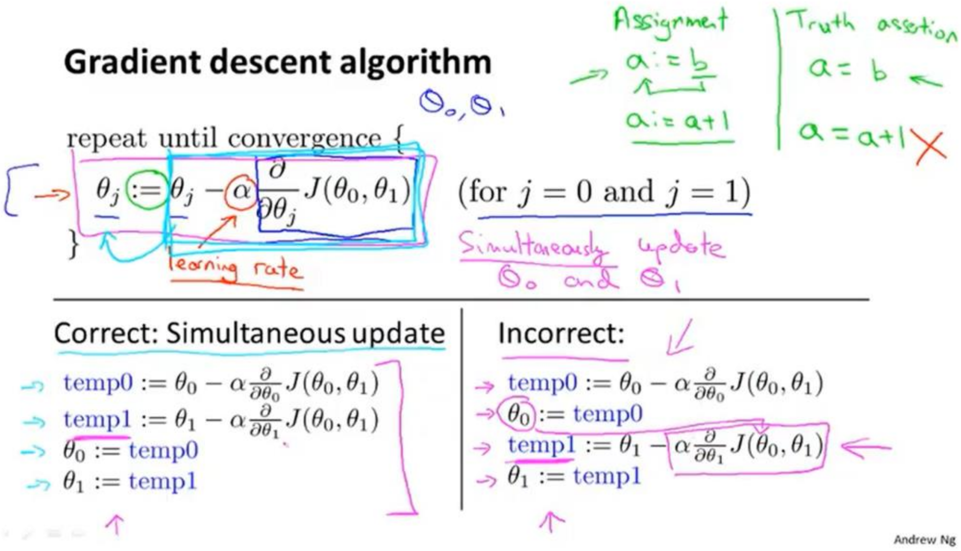


5. Gradient descent

Outline:



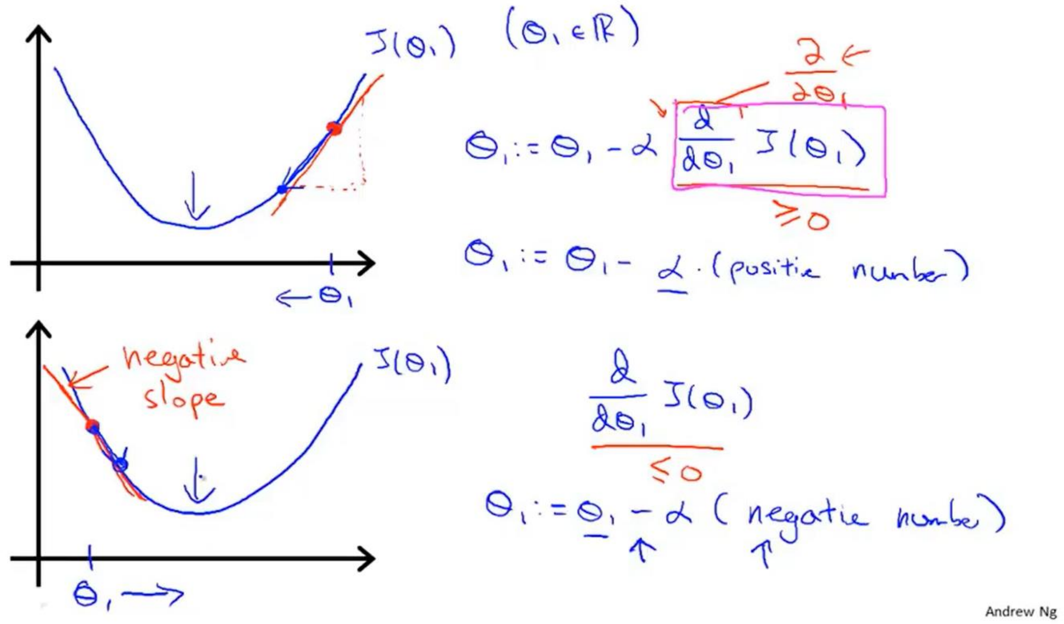
Simultaneous update!



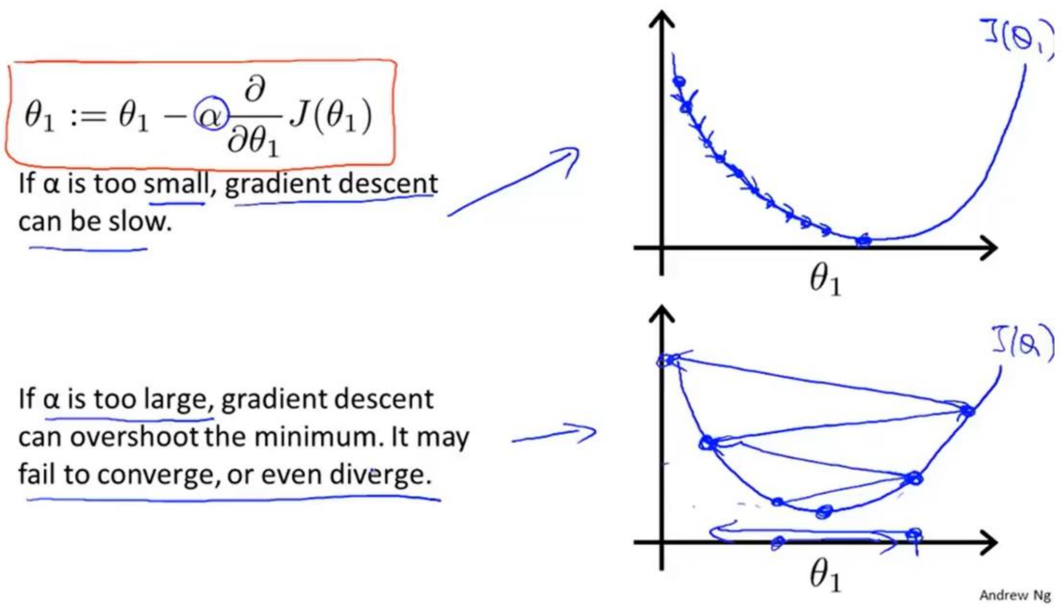
learning rate: how big a step we take downhill with gradient descent.

6. Gradient descent intuition

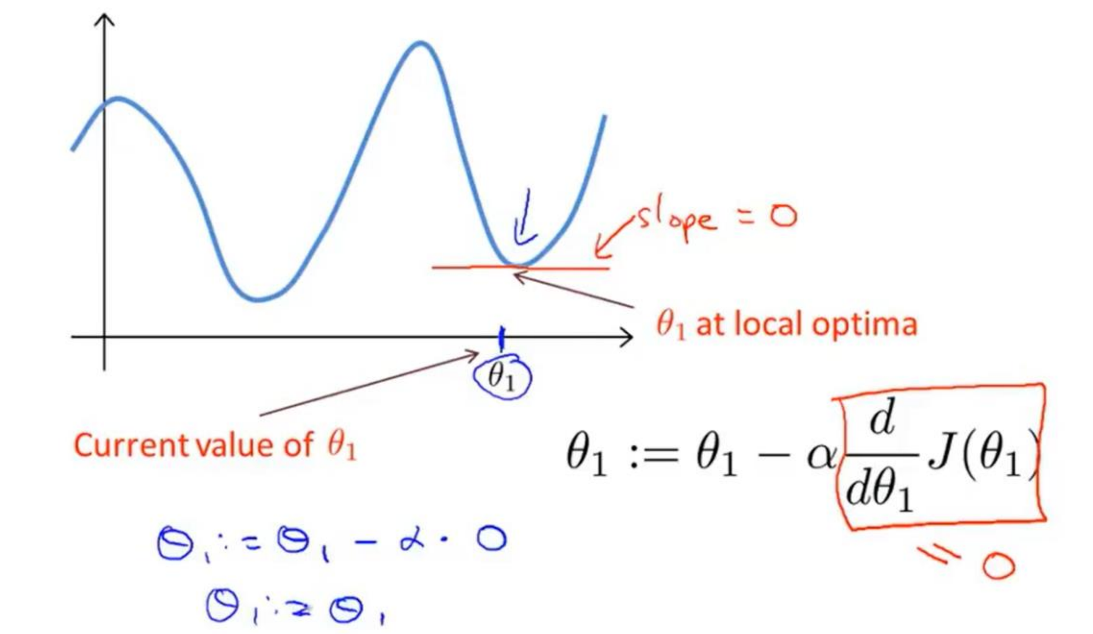
derivative term:

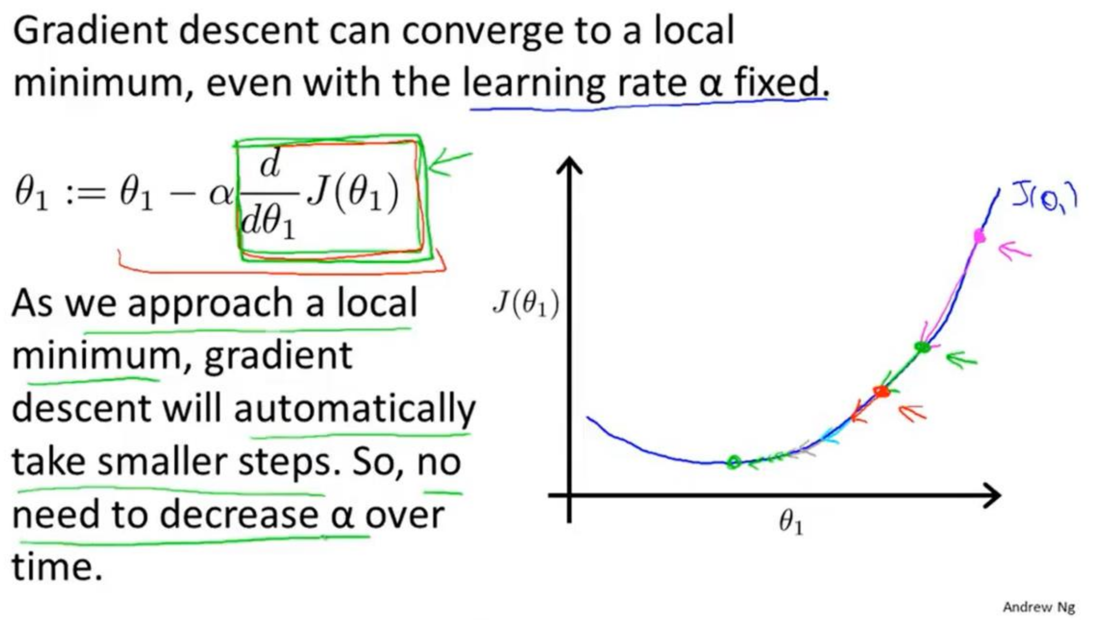


learning rate alpha:



local optimum:





7. Gradient descent for linear regression

repeat until convergence

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

$$= \frac{\partial}{\partial\theta_j}\frac{1}{2m}\sum_{i=1}^{m}\left(\theta_0+\theta_1 x^{(i)} - y^{(i)}\right)^2$$

$$\theta_0 \quad j=0: \frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)$$

$$\theta_1 \quad j=1: \frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)\cdot x^{(i)}$$

"Batch" Gradient Descent

"Batch": Each step of gradient descent uses all the training examples.