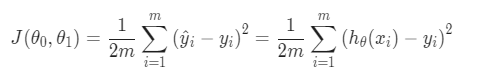
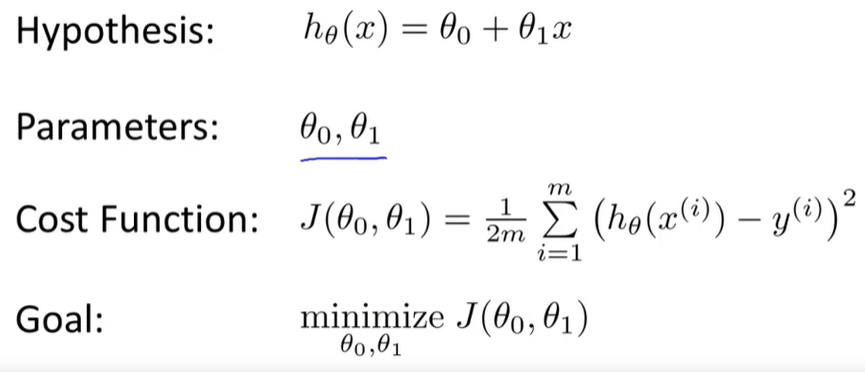
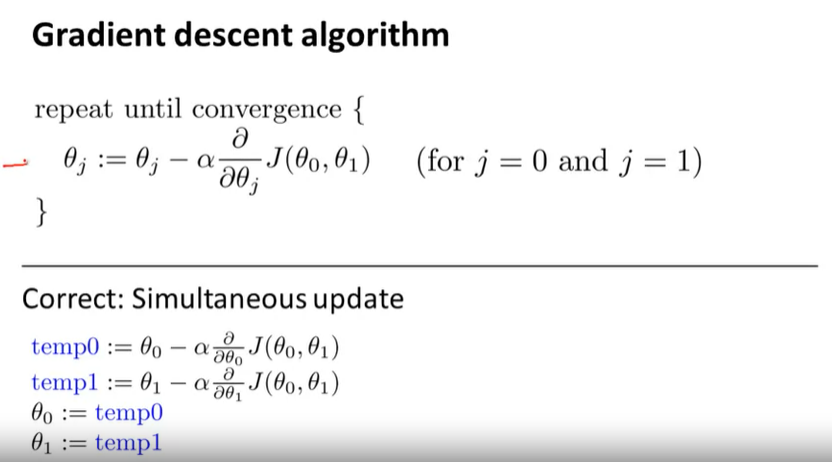
Cost Function:



This is called the “Squared error function” or Mean squared error.



Gradient descent algorithm:

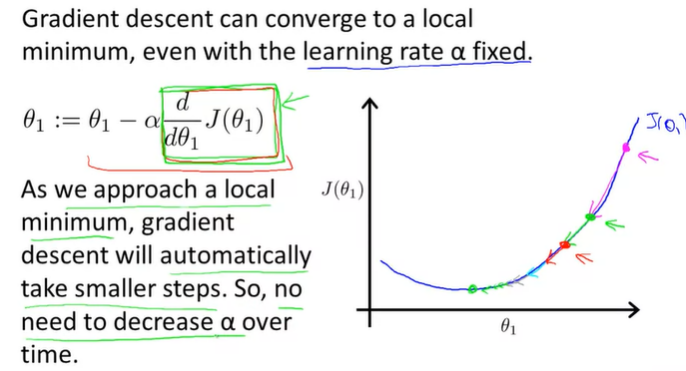


:= is assignment

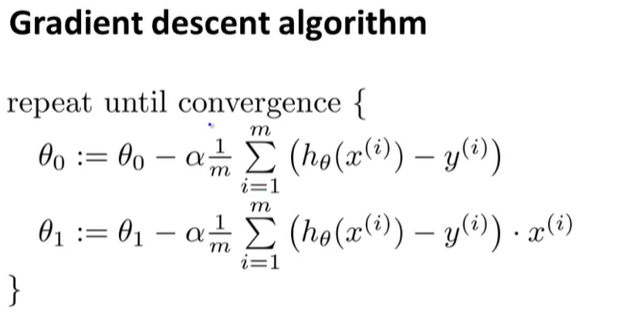
Alpha = learning convergence. If alpha is too large it might not converge even it can diverge.

Temp is needed to make sure simultaneous update.

As we approach local minima the derivative term get smaller and gradually it converges.



Gradient descent algorithm for linear regression:



“Batch” Gradient Descent.

Each step of gradient descent uses all the training example.

Feature scalling:

If the features are having different value range the contour can be narrow and convergence can be very slow. So normalizing by a range will help to converge faster.

