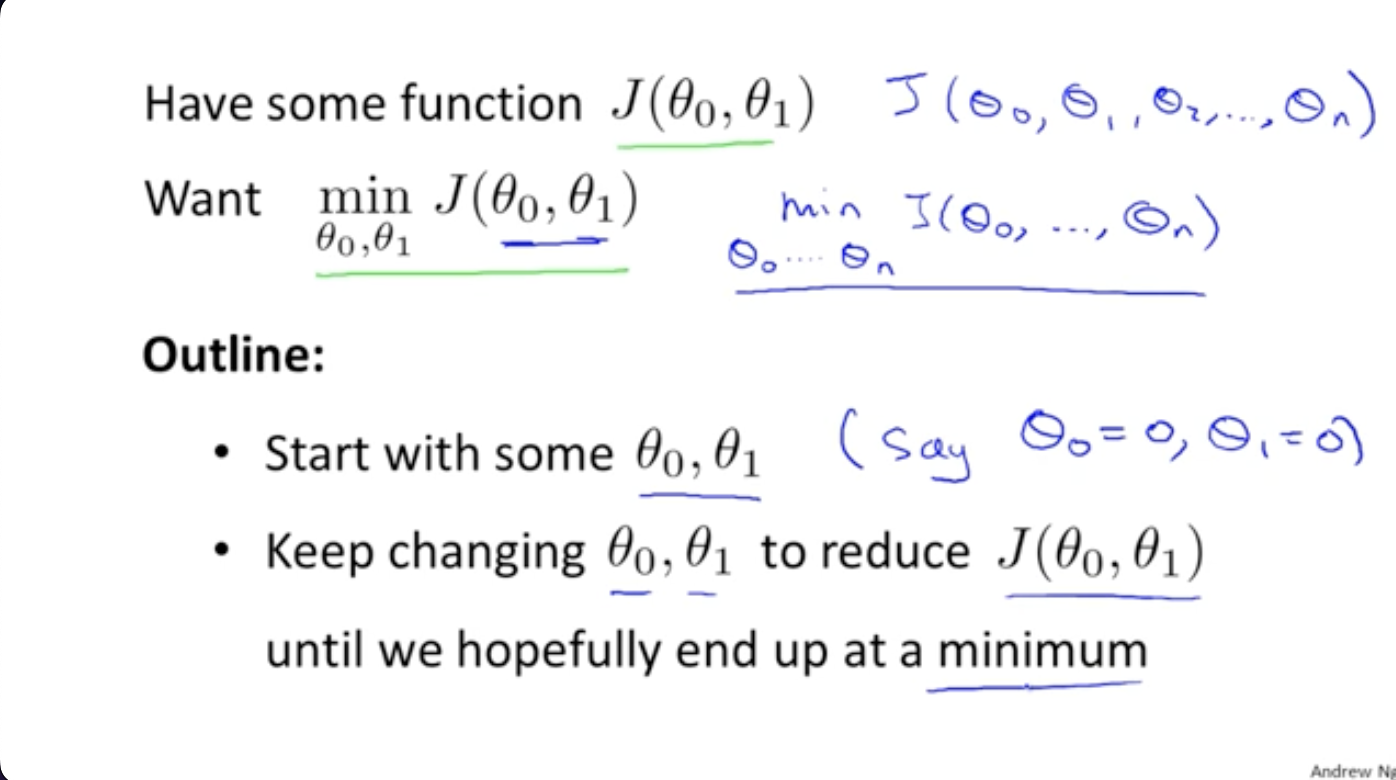
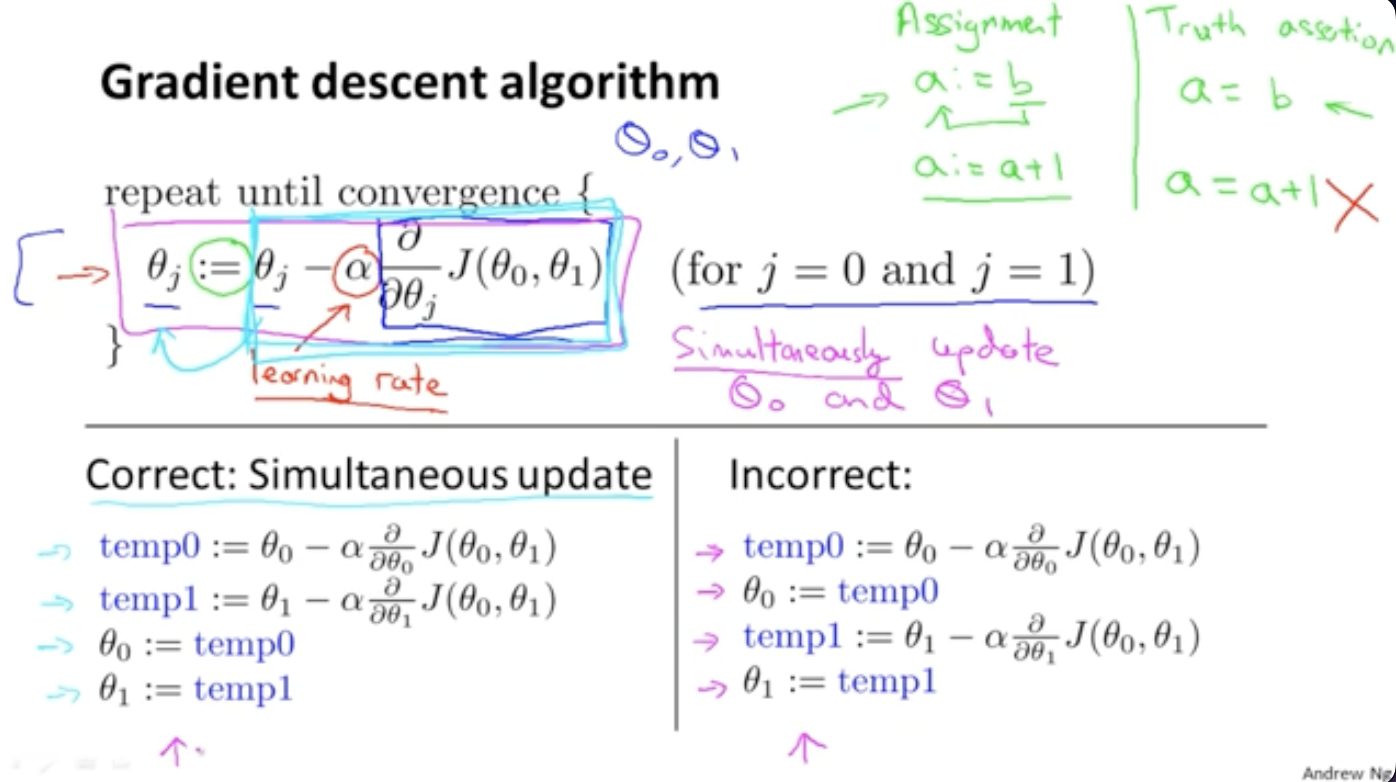
* The cost function J can be minimized using the gradient descent algorithm, which is applicable to more general functions.
  + Gradient descent are used for all sort of cost functions, including linear regression
* Gradient descent starts with initial parameter guesses and iteratively adjusts them to decrease J, aiming for a minimum or local minimum.
* 
* In each iteration, gradient descent calculates the direction of steepest descent by finding the negative gradient of J and taking a step in that direction.
* The step size is controlled by the learning rate alpha, which affects the aggressiveness of the descent.
* Gradient descent simultaneously updates both parameters (theta 0 and theta 1) to avoid using outdated values when computing the derivative.
* The derivative term used in gradient descent updates is related to the partial derivatives of J with respect to the parameters.
* 
* Although understanding calculus is helpful, the next video will provide the necessary intuition and explanations to calculate the derivative term without calculus knowledge.