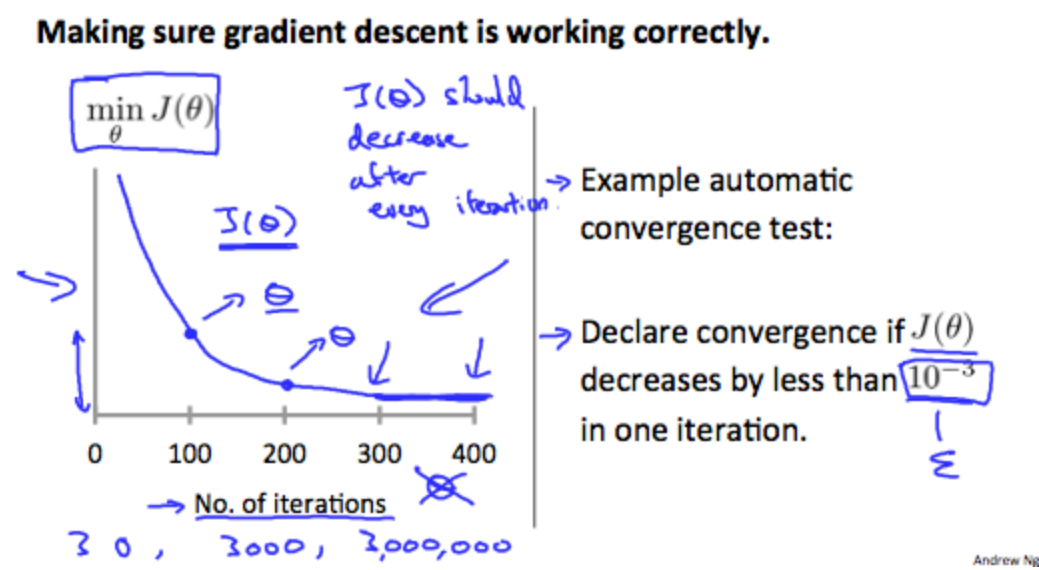


# Gradient Descent in Practice II - Learning Rate

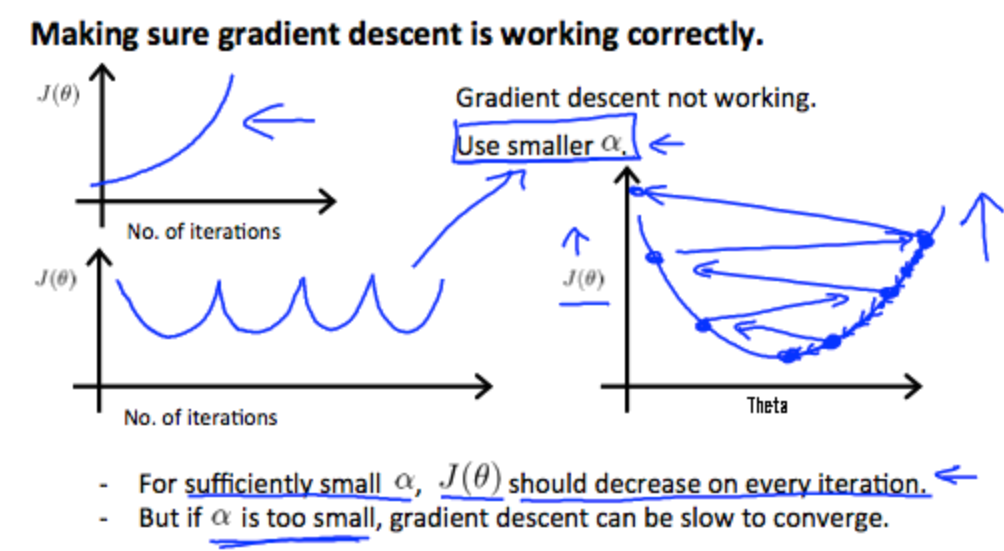
**Note:** [5:20 - the x-axis label in the right graph should be  $\theta$  rather than No. of iterations]

**Debugging gradient descent.** Make a plot with *number of iterations* on the x-axis. Now plot the cost function,  $J(\theta)$  over the number of iterations of gradient descent. If  $J(\theta)$  ever increases, then you probably need to decrease  $\alpha$ .

**Automatic convergence test.** Declare convergence if  $J(\theta)$  decreases by less than  $E$  in one iteration, where  $E$  is some small value such as  $10^{-3}$ . However in practice it's difficult to choose this threshold value.



It has been proven that if learning rate  $\alpha$  is sufficiently small, then  $J(\theta)$  will decrease on every iteration.



To summarize:

If  $\alpha$  is too small: slow convergence.

If  $\alpha$  is too large: may not decrease on every iteration and thus may not converge.

✓ Complete