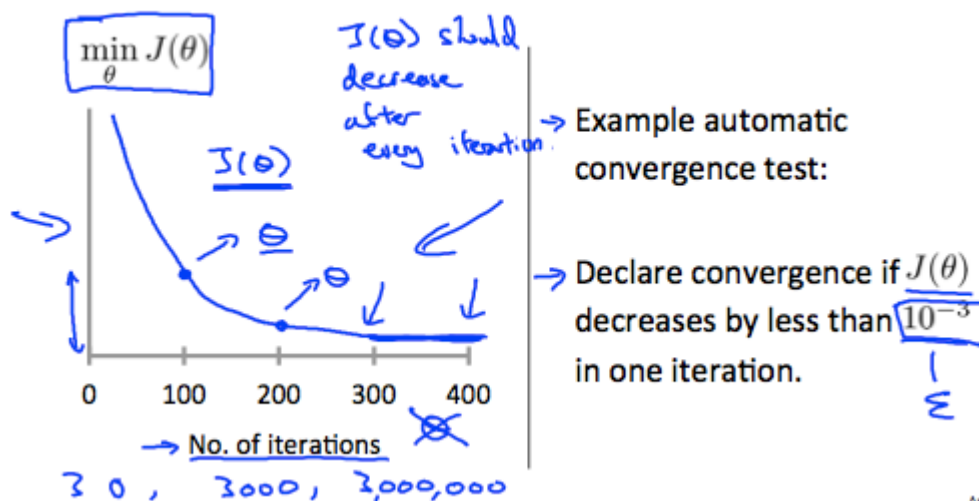# Gradient Descent in Practice II - Learning Rate

**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(θ) over the number of iterations of gradient descent. If J(θ) ever increases, then you probably need to decrease α.

**Automatic convergence test.** Declare convergence if J(θ) 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.

**Making sure gradient descent is working correctly.**



It has been proven that if learning rate α is sufficiently small, then J(θ) will decrease on every iteration.

## Making sure gradient descent is working correctly.

Gradient descent not working.

Use smaller $\alpha$.
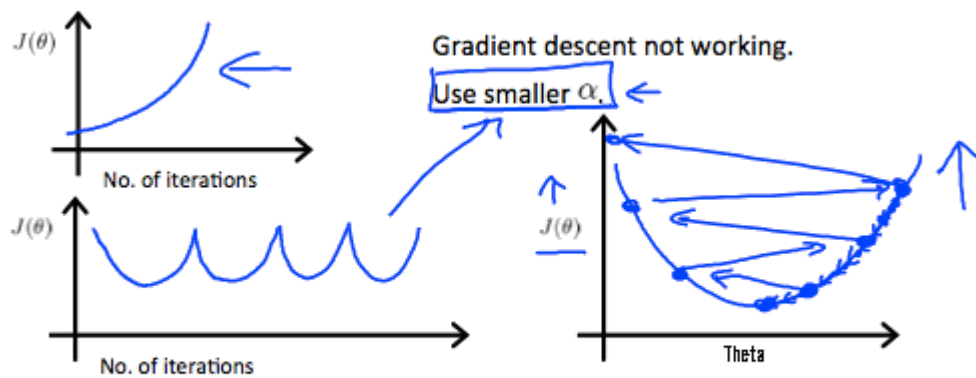
- For sufficiently small $\alpha$, $J(\theta)$ should decrease on every iteration.
- But if $\alpha$ is too small, gradient descent can be slow to converge.

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.

Mark as completed