

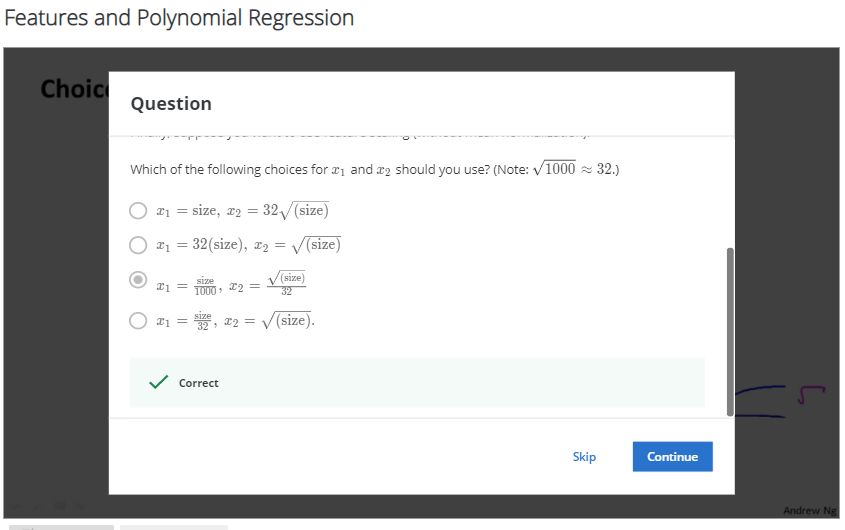
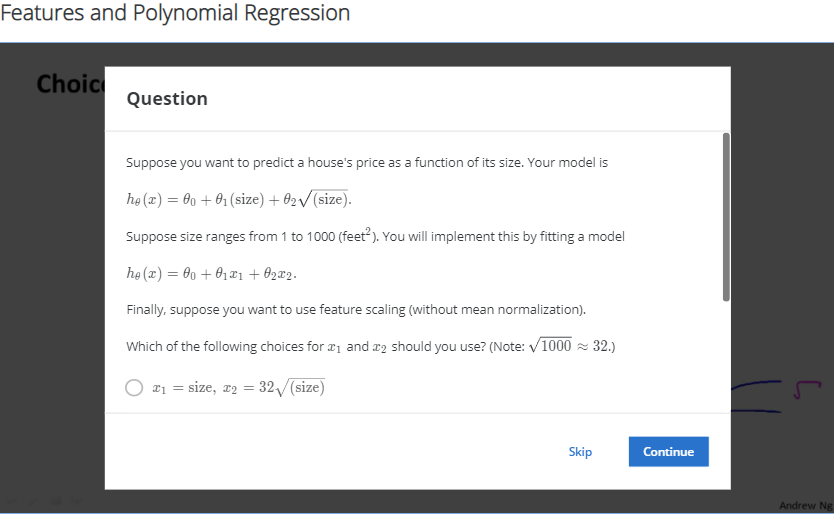
The learning rate 0.01 is probably too small, so 0.1 gives a better result (in diagram A).

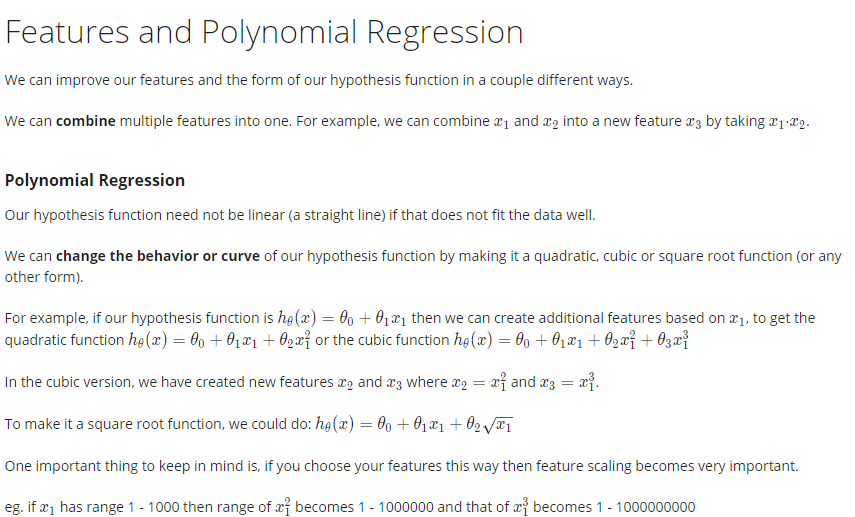
Gradient Descent in Practice II - Learning Rate

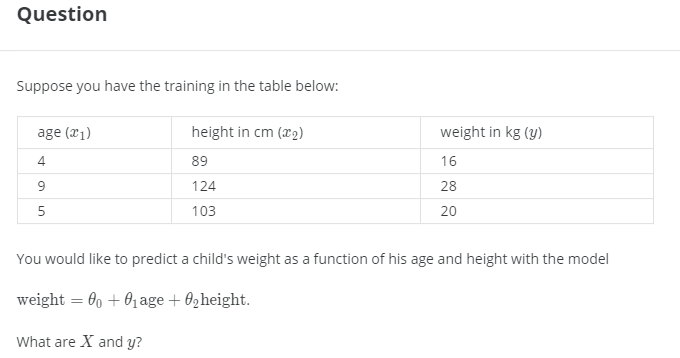
**Note:** [5:20 - the x -axis label in the right graph should be \theta*θ* rather than No. of iterations ]

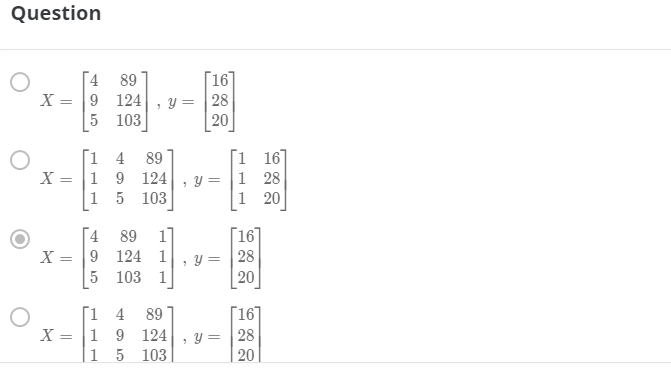
**Debugging gradient descent.** Make a plot with *number of iterations* on the x-axis. Now plot the cost function, J(θ) over the number of iterations of gradient descent. If J(θ) ever increases, then you probably need to decrease α.

**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}10−3. However in practice it's difficult to choose this threshold value.









This question was already set with an answer, but it’s not the one I would have chosen! Why not D?

The next page with the summary notes on Normal Equation seems to confirm this…

