# [DL] [Neural Networks and Deep Learning](https://www.coursera.org/learn/neural-networks-deep-learning/home/welcome) [Week 3](https://www.coursera.org/learn/neural-networks-deep-learning/home/week/3)

## Why randomize weight initialization?

So that the hidden layers don’t compute the same things. (Symmetric)

## Why do we initialize the weights to a small number?

i.e. multiplying weights by 0.01

Because big weights will slow learning.

Look at the tanh function. The farther you are, both positive and negative, the smaller the gradient. Small gradient = small updates = slow learning.