# Journal 3

### Weight Initialization

* Large n means smaller weights
* Initialize weights with variance of 1/n
* Variance of 2/n works better with ReLu
* Xavier initialization for tanh activation function
* Could potentially set the variance of your weight initialization in your hyperparameters in your neural network

### Regularization/Cost Function

* Intuition: penalize the higher-order terms in your model to avoid overfitting
* Regularization “simplifies” the hypothesis. This seems to aligned with the “keep it simple stupid” approach talked about when doing data science
* I’m curious as to what the intuition is behind setting a value for lambda. Maybe there are some rules of thumb for this. Too small there isn’t enough regularization and too high it will over-simplify the model

### Adam Optimization Algorithm

* Adam optimization “stands out” for neural networks
* Takes momentum & RMS prop to combine them
* This video seems out of context… perhaps there should be another recommended video before this. He jumps right into a proof that he doesn’t explain the context for beforehand
* Need to tune your optimizer’s hyperparameters
* Adam (Adaptive Moment Estimation)