deeplearning.ai

Hyperparameter tuning

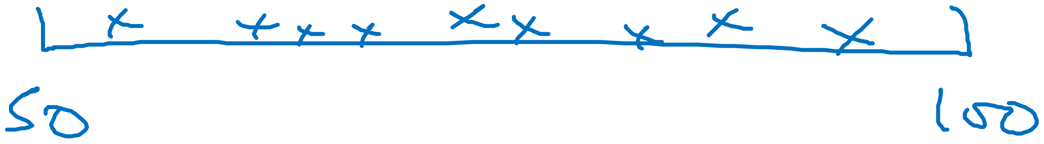
Using an appropriate scale to pick hyperparameters

But it turns out that sampling at random doesn't mean sampling uniformly at random over the range of valid values. Instead, it's important to pick up the appropriate scale on which to explore the hyperparameters

Picking hyperparameters at random

Maybe you think the total number of layers should be somewhere between 2 to 4. Then sampling uniformly at random along 2,3 and 4, might be reasonable or even using a grid search, where you explicitly evaluate the values 2, 3 and 4 might be reasonably.

$\rightarrow n^{[l]} = 50, \ldots, 100$



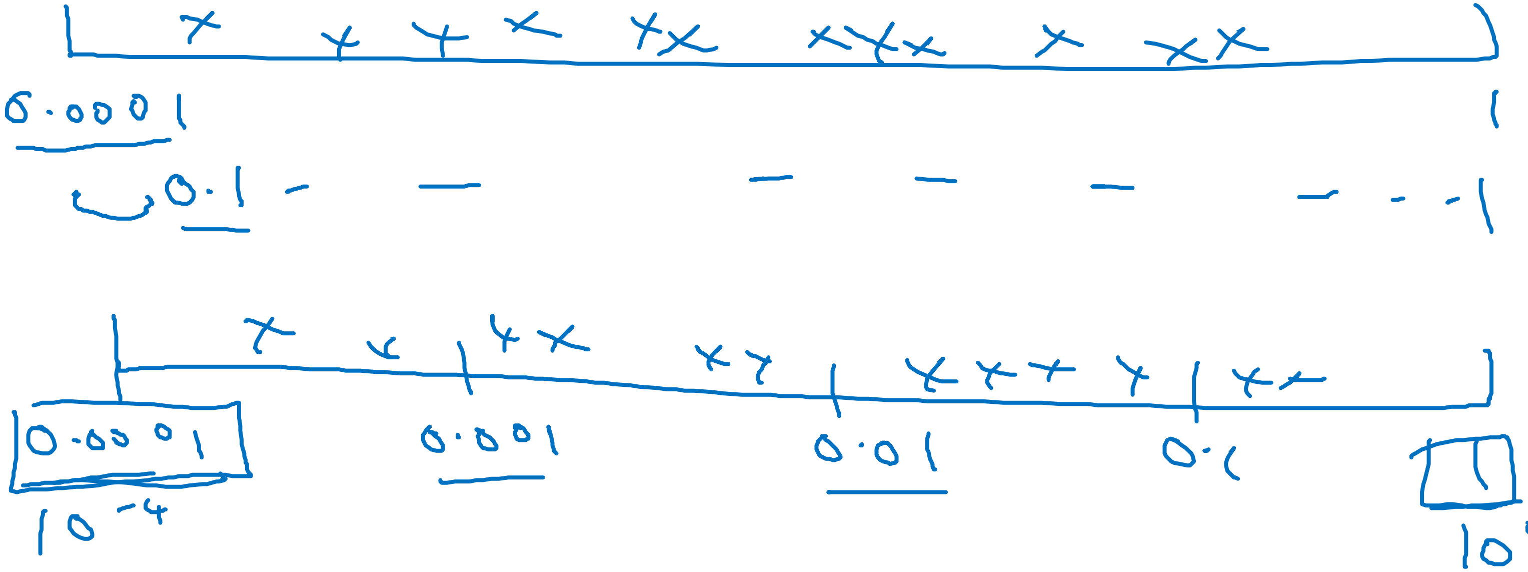
$\rightarrow$ #layers $L: 2 - 4$

$2, 3, 4$

Andrew Ng

Appropriate scale for hyperparameters

$\alpha = 0.0001, \ldots, 1$



So what you do, is then sample r uniformly, at random, between a and b.

10^a $\quad$ 10^b

$a = \log_{10} 0.0001 = -4$ $\quad$ $r = -4 * \text{np.random.rand}()$ $\leftarrow$ $r \in [-4, 0]$ $\quad$ $b = \log_{10} 1 = 0$

$\alpha = 10^r$ $\leftarrow$ $10^{-4} \ldots 10^{0}$

$10^a \ldots 10^b$ $\quad$ $r \in [a, b]$ $[-4, 0]$ $\quad$ $\alpha = 10^r$

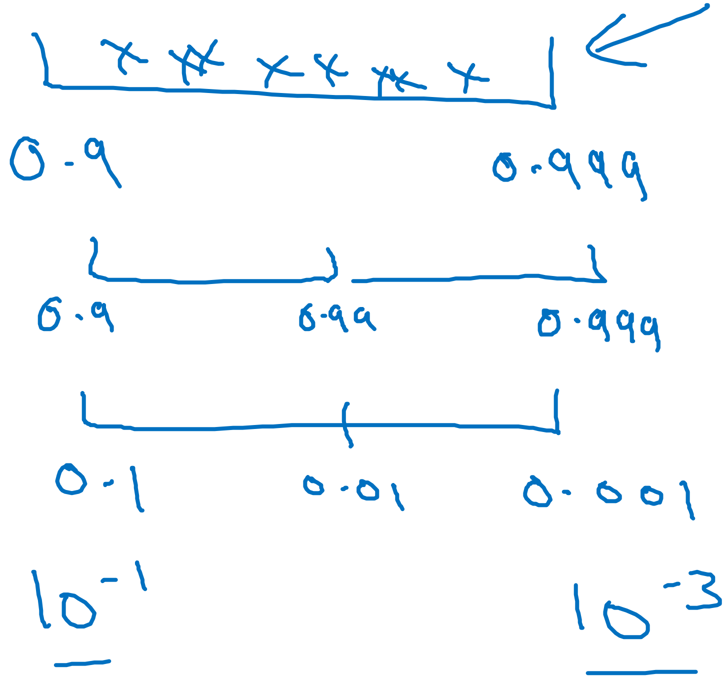
Andrew Ng

Hyperparameters for exponentially weighted averages

$$\beta = 0.9 \quad \ldots \quad 0.999$$

$$\downarrow 10 \qquad \downarrow 1000$$

$$1-\beta = 0.1 \quad \ldots \quad 0.001$$



$$\beta: 0.9000 \to 0.9005 \ \} \sim 10$$

$$\beta: \underset{\sim 1000}{0.999} \to \underset{\sim 2000}{0.9995}$$

$$\frac{1}{1-\beta}$$

$$r \in [-3, -1]$$

$$1-\beta = 10^r$$

$$\beta = 1 - 10^r$$

This is very sensitive to small changes in beta, when beta is close to 1.
So what this whole sampling process does is it causes you to sample more densely in the region of when beta is close to 1.

Andrew Ng