deeplearning.ai

Hyperparameter tuning

Using an appropriate scale to pick hyperparameters

Picking hyperparameters at random

Sample at random was the Advice

↳ ⇒ you can start taking $\alpha = 100, \alpha = 10,000, \alpha = 10^6$

- But all of them will give similar values of model perf
- ⇒ Sample Intelligently to avoid waste of compute

But there are other hyperparams that can be sampled within a given Range

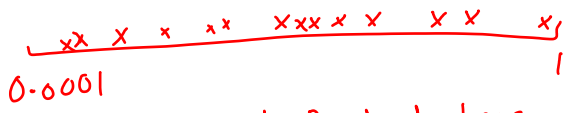
layers → 2 to 10

hidden units → $n^{[l]}$ → 50 to 100

Appropriate scale for hyperparameters

Say you are selecting best α

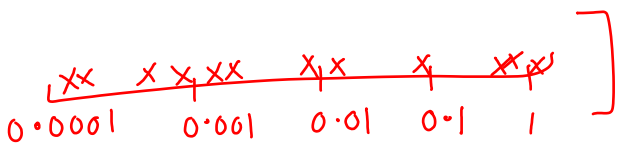
$\alpha = .0001 \ldots 1$



- If you sample Randomly here you'd end up picking more values b/w 0.1 & 1 compared to picking b/w (.0001 & .001)

$\Rightarrow$ you may end up choosing sub optimal α

$\Rightarrow$ use log scale in the Range you sample



Now If you sample, you're equally likely to choose from Any of the milestones

$\Rightarrow$ .0001 to .001 gets a fair shot at being chosen

$r = -4 * np.random.rand()$

$\Rightarrow r = \in [-4, 0]$

$\alpha = 10^r \Rightarrow \alpha = \in [10^{-4}, 10^0]$

$10^{-4} \rightarrow 0.0001$

$10^0 \rightarrow 1$

So you do $\log(0.0001) = (-4) = a$

$\log 1 = 0 = b$

Andrew Ng

Hyperparameters for exponentially weighted averages

Finding β to find V_{dw}, V_{db} etc in Adam/RMS prop

$\beta = 0.9$ to 0.999

$\frac{1}{1-\beta}$ "Last 10 days temp" $\frac{1}{1-0.9}$

$\frac{1}{1-0.999}$ "Last 1000 days temp"

$\Rightarrow$ Sampling uniformly wont help - same reason as last slide

$1-\beta = 0.1 \ldots 0.001$ $\Rightarrow$ same log sampling

Q Why is it such a big deal to consider diff samples, esp when $\beta \approx 1$

Because it will affect the model perf Alot!

How? Say $\beta = 0.999 \rightarrow \frac{1}{1-\beta} = 1000$ day temp

If $\beta = 0.9995 \rightarrow \frac{1}{1-\beta} = 2000$ day temp

$\Rightarrow$ small change in β changes moving Avg drastically, esp when β is close to 1

0.1 — 0.01 — 0.001

$a = 10^{-1}$ $b = 10^{-3}$

Same logic as before do a log scale & sample $\Rightarrow$ This allows equal sampling b/w (.1 & .01) v/s (.01, .001)

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

$\Rightarrow \beta = 1 - 10^{-r}$

Andrew Ng