

η ← most important

$\beta_1, \beta_2, \epsilon = 10^{-8}$

layers

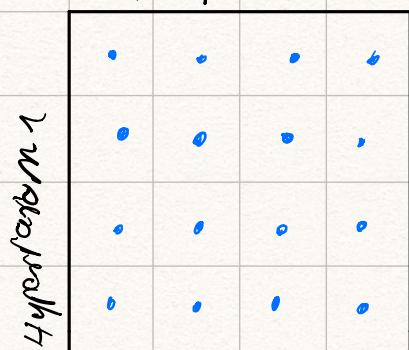
hidden units

learning rate decay

mini batch size

How to select set of values to explore

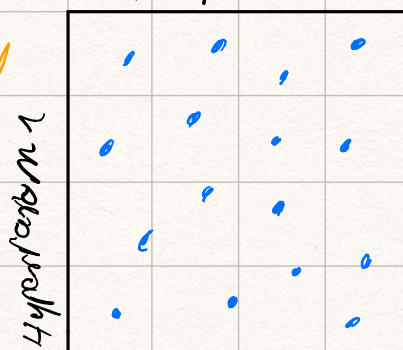
Hyperparam 2



Recommended



Hyperparam 2



Grid sample

Random select

$4 \times 4 = 16$ models but only 4 values of Hyperparam 1 and 4 values of Hyperparam 2

Random select \Rightarrow 16 values of both Hyperparameters.

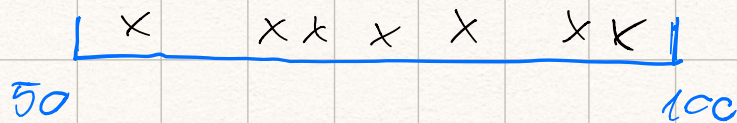
Coarse to fine sampling



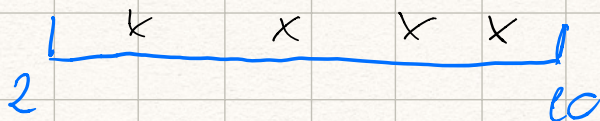
worked the best
zoom in and sample randomly
again

Picking Hyperparameters at random

of hidden units $\approx 50, \dots, 100$



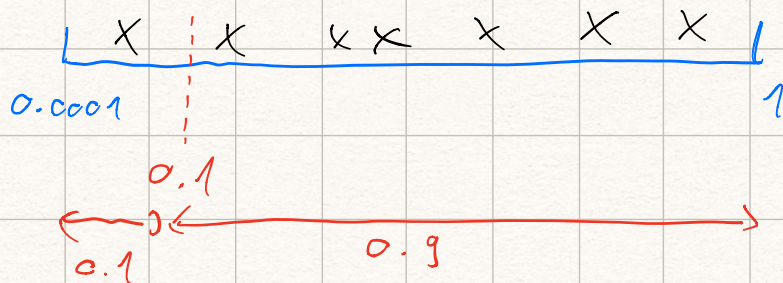
of layers $\approx 2, \dots, 10$



sampling uniformly random is a reasonable thing by some hyper params, but it is not true for all of them

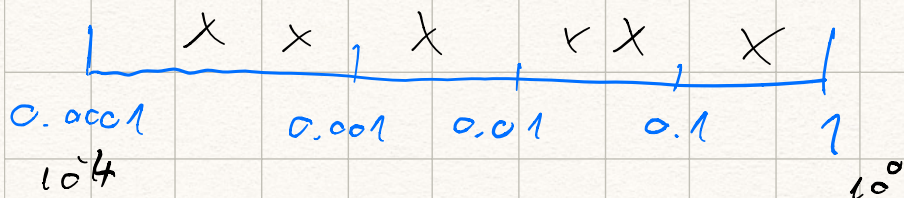
Appropriate scale for hyperparameters

$\eta = 0.0001, \dots, 1$



only 10% of the values between 0.0001 - 0.1

instead of use logarithmic sampling



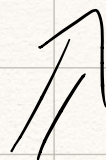
sampling $\eta \in [a, b]$
and $\eta \in [10^a, 10^b]$

$$\tau = -4 \cdot \text{np.random}()$$

$$\leftarrow \tau \in [-4, 0]$$

$$\eta = 10^\tau$$

$$\leftarrow \eta \in [10^{-4}, 10^0]$$

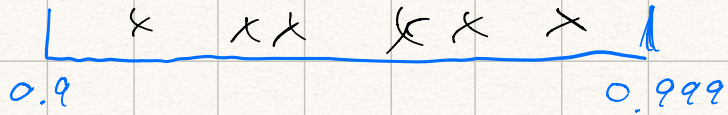


Hyperparameters for exponentially weighted averages

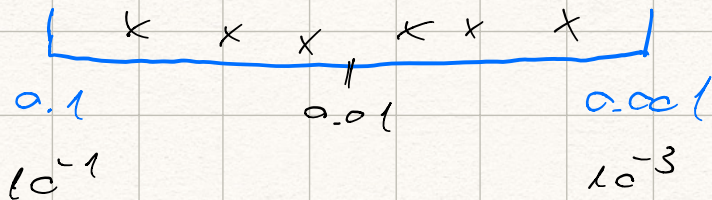
$$\beta = 0.9 \dots 0.999$$

~ 10
average

~ 1000
average



$$1 - \beta = 0.1 \dots 0.0001$$



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

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

$$\beta : 0.9000 \rightarrow 0.9005 \quad \} \sim \text{average of } 10$$

$$\beta : 0.999 \rightarrow 0.9995$$

~ 1000 ~ 2000

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

small changes on $\beta \Rightarrow$

\Rightarrow high effect