

## Week 3-1 Hyperparameter Tuning

笔记本: DL 2 - Deep NN Hyperparameter Tuning, Regularization & Optimization

创建时间: 2021/1/9 13:41

更新时间: 2021/1/9 13:56

---

learning rate

> momentum & mini-batch size &  
hidden units

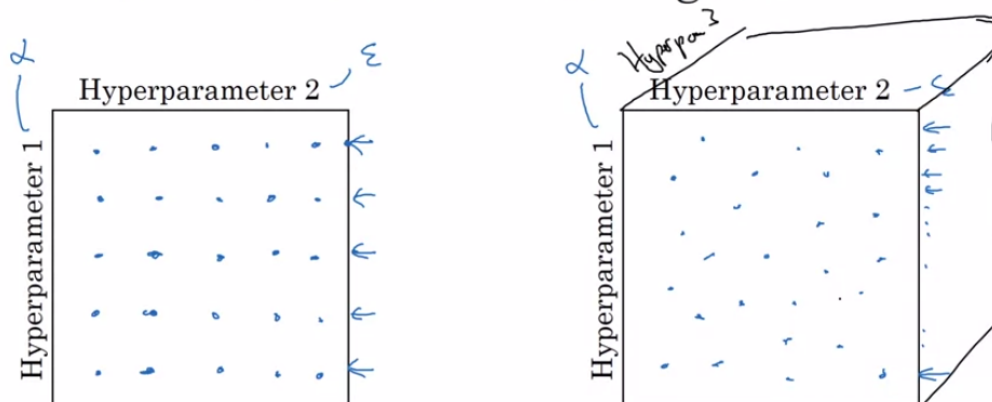
> layers & learning rate decay

seldom tunes Adam parameter

---

random -> able to try 25 different  
values of alpha

Try random values: Don't use a grid

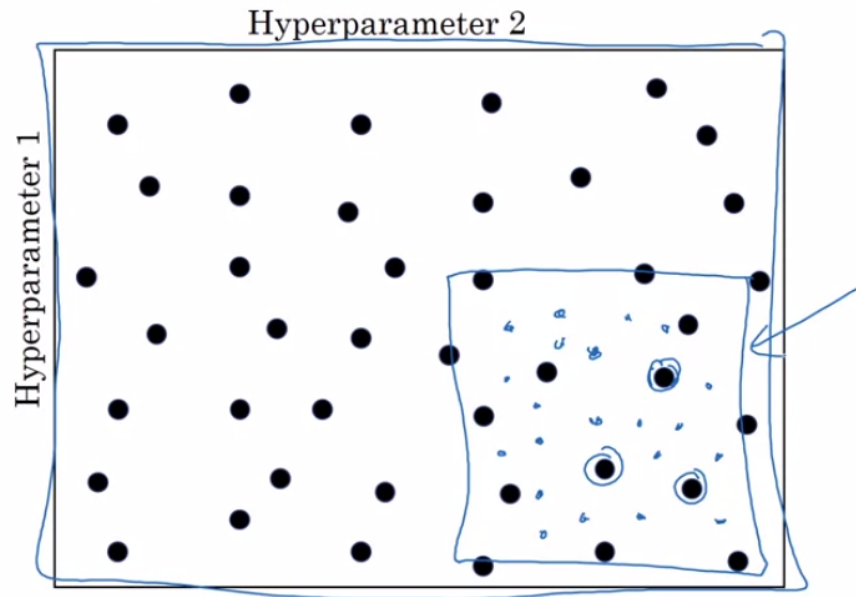


you are more richly exploring set of  
possible values for the most

important hyperparameters,  
whatever they turn out to be.

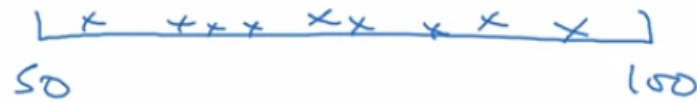
Coarse

Coarse to fine



appropriate scale for hyperpara

$$\rightarrow n^{[2]} = 50, \dots, 100$$



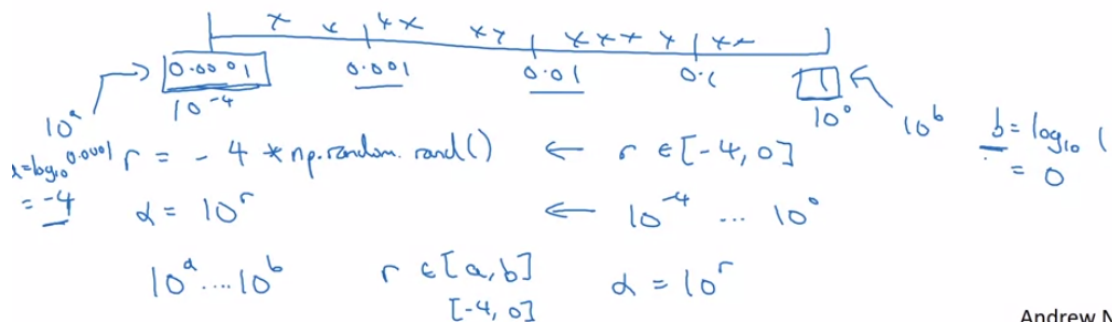
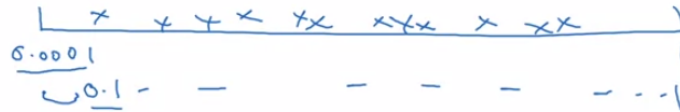
$$\rightarrow \# \text{ layers } L: 2 - 4$$

$$2, 3, 4$$

linear for these are fine

Appropriate scale for hyperparameters

$$\alpha = 0.0001, \dots, 1$$



Andrew Ng

linear scale will focus 90% on 0.1 - 1, use log scale

# hyperpara for exponentially weighted averages

Hyperparameters for exponentially weighted averages

$$\beta = 0.9 \quad \dots \quad 0.999$$

$\downarrow$                        $\downarrow$   
10                      1000

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

$$\beta: 0.999 \rightarrow 0.9995$$

$$\beta: 0.999 \rightarrow 0.9995$$

$\left| \begin{array}{c} \times \times \times \times \times \times \end{array} \right| \leftarrow$

0.9                      0.999

0.9                      0.999

0.1                      0.001

$10^{-1}$                        $10^{-3}$

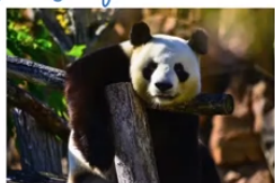
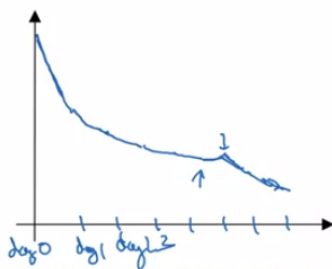
$$r \in [-3, -1]$$

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

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

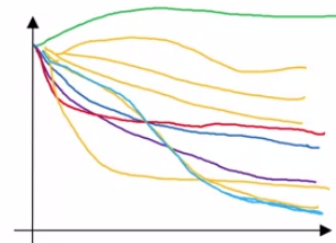
## Pandas vs. Caviar (computational resources)

Babysitting one  
model



Panda

Training many  
models in parallel



Caviar

Andrew Ng

