

tuning1

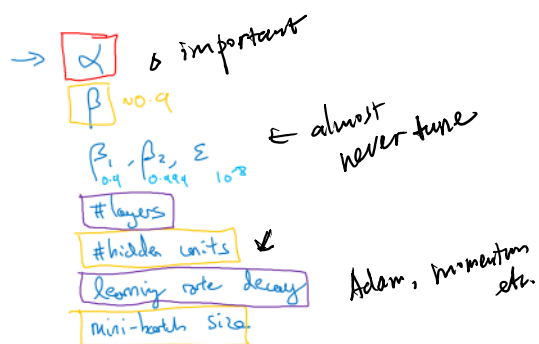


deeplearning.ai

Hyperparameter tuning

Tuning process

Hyperparameters



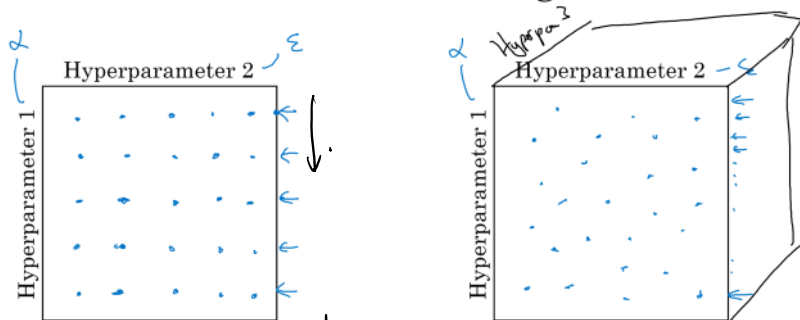
① Problem: large number of parameters.

Some hyper parameters are more important than others

red: most important
yellow: second important

Andrew Ng

Try random values: Don't use a grid



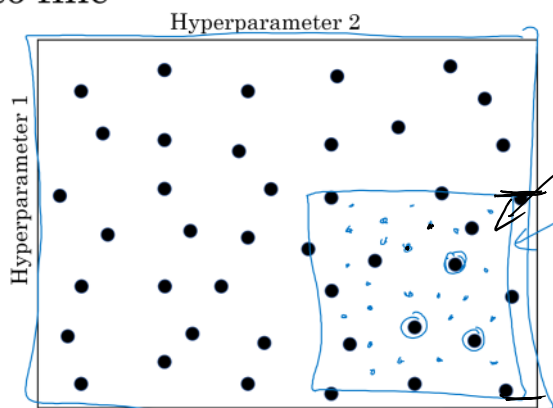
previous: sample from a grid
ex. 25 points \Rightarrow pick whichever
works best. - 4-4 排列
3次 \Rightarrow DL 这样不行. \Rightarrow 随机找. 重要的试试.

Δ DL choose the points
at random.

Andrew Ng

Sometimes it's hard to know
which hyperparameter is more important

Coarse to fine



Andrew Ng

search more densely
here

找 4-4 排列
地 5-找
得 dense 些.

Zoom in, sample more
densely in some area
you suspect to contain
the optimal hyperparam.

TAKE AWAY: ① Don't search by grid - use random search
② Coarsed define searching



deeplearning.ai

Hyperparameter tuning

Using an appropriate scale to pick hyperparameters

Not searching randomly uniformly
pick the scale!

Picking hyperparameters at random

→ $n^{\text{test}} = 50, \dots, 100$

$\begin{array}{c} \text{[x x x x x x x]} \\ 50 \qquad \qquad \qquad 100 \end{array}$

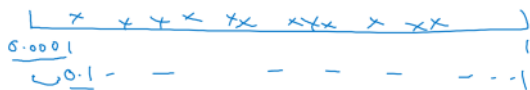
→ #layers $L: \frac{2-4}{2, 3, 4}$

Scale easier to choose
for these hyperparam

Andrew Ng

Appropriate scale for hyperparameters

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



10^4 \rightarrow $\frac{0.0001}{10^{-4}}$ \leftarrow 10^0 \leftarrow 10^6 $\frac{b = \log_{10} 1}{= 0}$
 $a = \log_{10}(0.0001) = -4$ $r = -4 * \text{np.random.rand}() \leftarrow r \in [-4, 0]$ \leftarrow $10^{-4} \dots 10^0$
 $\alpha = 10^r$ $r \in [a, b]$ $\alpha = 10^r$
 $10^a \dots 10^b$ $r \in [-4, 0]$

Andrew Ng

How to fix β bet. $\frac{\beta}{\beta+1}$. [train, X_train]
 $[\log X_{min}, \log X_{max}]$

\Rightarrow take random sample \Rightarrow take exp. of the sample

use 90% resample to search between 0.1 to 1.

Not good.

What should do: log scale. ?

4. If you think β (hyperparameter for momentum) is between 0.9 and 0.99, which of the following is the recommended way to sample a value for beta?

- ☐ 1 $r = \text{np.random.rand}()$
 - ☐ 2 $\beta = r * 0.99 + 0.9$
 - ☒ 3 $r = \text{np.random.rand}()$
 - ☐ 4 $\beta = 1 - 10^{r-1}$
- This should not be selected

Hyperparameters for exponentially weighted averages

$$\beta = 0.9 \dots 0.999$$

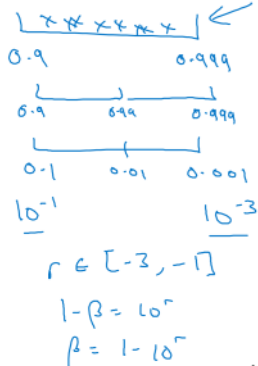
\downarrow \downarrow
 10 1000

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

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

$$\beta: 0.999 \rightarrow 0.9995 \} \sim 1000$$

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



$\beta \rightarrow 0.999$
 \Rightarrow fix $1-\beta$.

Why bad idea to sample linearly.

When beta close to 1.

e.g. $0.999 \rightarrow 0.9995$, the model is sensitive to the small change of β

\Rightarrow So. we want to sample densely for β close to 1, because in that area, small β makes big difference

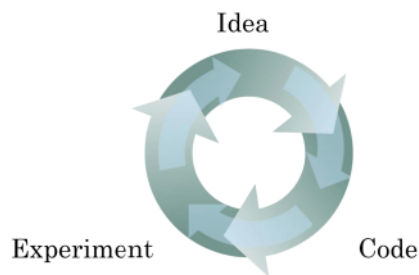


deeplearning.ai

Hyperparameters tuning

Hyperparameters tuning in practice: Pandas vs. Caviar

Re-test hyperparameters occasionally

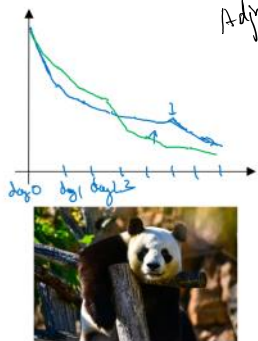


- NLP, Vision, Speech,
Ads, logistics,
- Intuitions do get stale.
Re-evaluate occasionally.

- People from different domains read works of those from other domains.
- Models developed in one domain work in other domains.

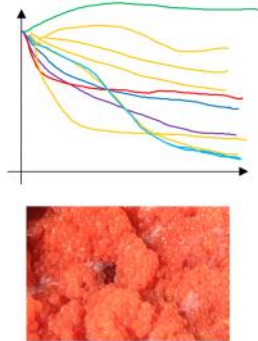
Andrew Ng

Babysitting one model



Panda

Training many models in parallel



Caviar

Andrew Ng

Adjust one model

when you do not have lots of computational power.

努力调一个模型
细致地调参。



同时调不同模型。

A function of computational resource

Fish ...
 - lots of computational power. Caviar
 - many data, limited computational power
Panda!
 e.g. online advertisement.

Another way to make your DL model more robust

