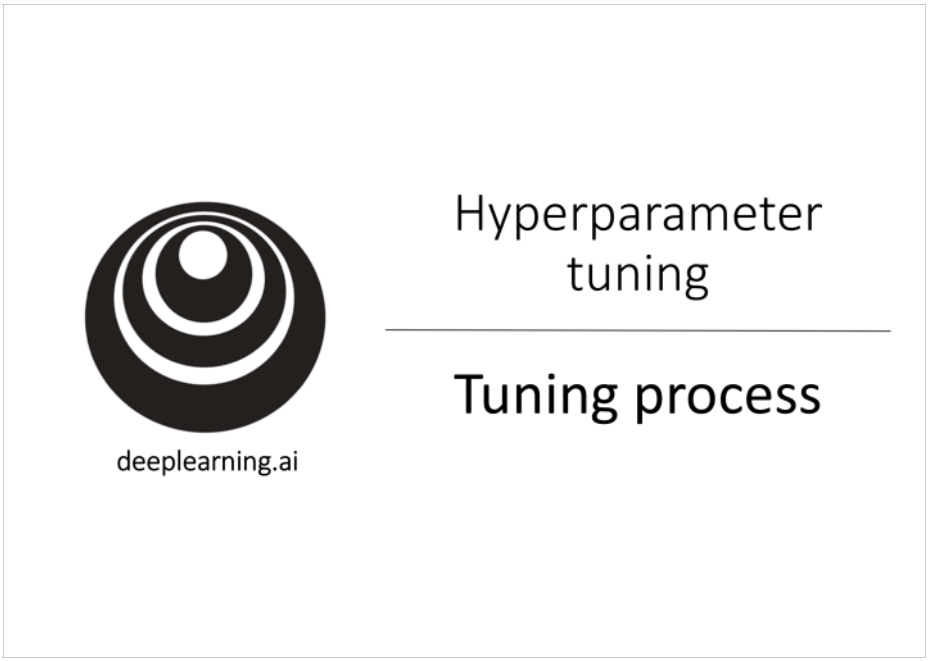
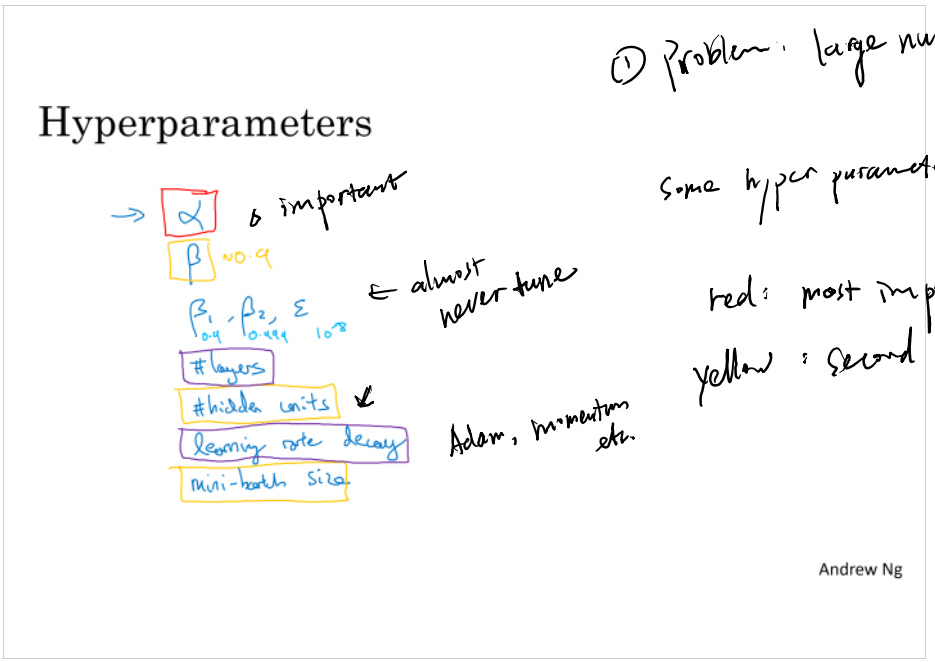
Tuning Hyper Parameters

Monday, June 4, 2018 9:08 PM

tuning1





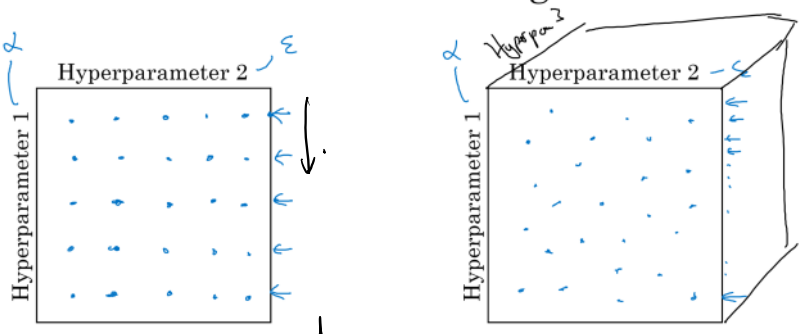
① Problem: large number of parameters.

Some hyper parameters are more important than others

red: most important

yellow: second important.

Try random values: Don't use a grid

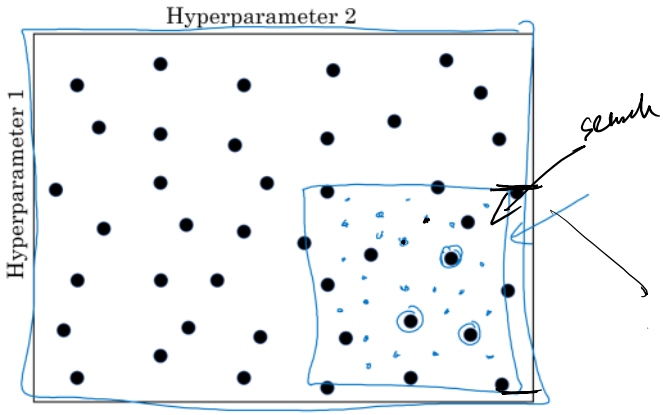


previous: sample from a grid
ex. 25 points ⇒ pick whichever works best.

Δ DL choose the points at random.

Andrew Ng

Coarse to fine



Andrew Ng

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

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

tuning2

deeplearning.ai

Hyperparameter tuning

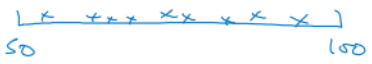
Using an appropriate scale to pick hyperparameters

Not searching randomly uniformly

pick the scale!

Picking hyperparameters at random

→ $n^{[l]}$ = 50, ..., 100



→ #layers L: 2 – 4

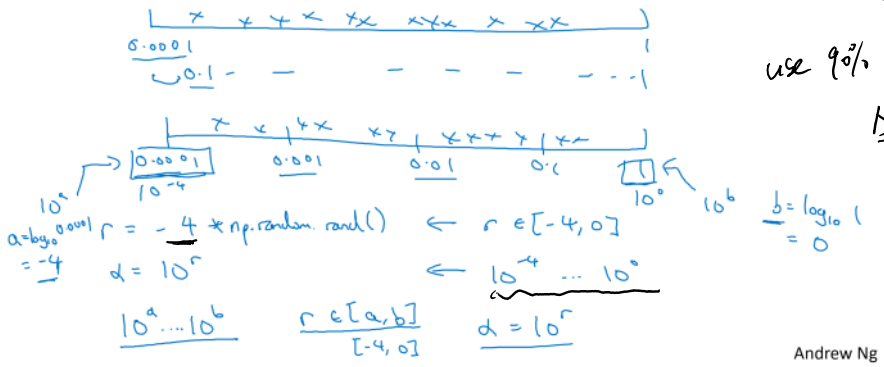
2, 3, 4

Scale easier to choose for these hyperparam

Andrew Ng

Appropriate scale for hyperparameters

$\alpha = 0.0001 \ldots, 1$



Andrew Ng

How to 取值? 在 [Xmin, Xmax]

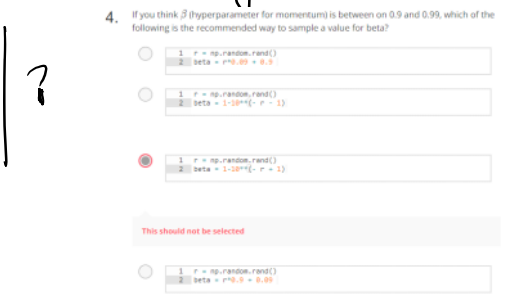
[log Xmin, log Xmax]

⇒ take random sample ⇒ take exp of the sample

use 90% resource to search between 0.1 to 1.

Not good.

What should do: log scale ?



Hyperparameters for exponentially weighted averages

$\beta = 0.9 \ldots 0.999$

10 1000

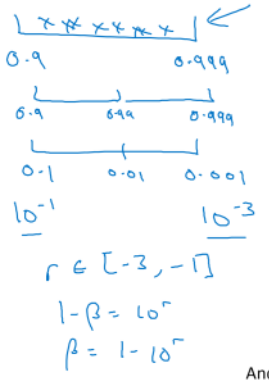
$1-\beta = 0.1 \ldots 0.001$

β: 0.9000 → 0.9005 } ≈10

β: 0.999 → 0.9995

≈1000 ≈2000

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



Andrew Ng

β 越接近 1 sensitive 越高 β → 0.999.

⇒ 取 1-β.

Why bad idea to sample linearly.

when beta close to 1.

e.g. 0.999 → 0.9995, the model is sensitive to the small change of β

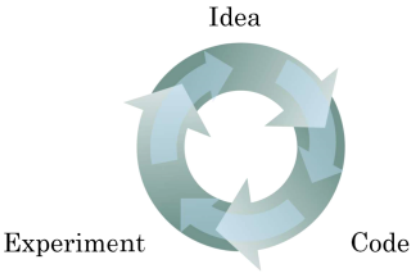
⇒ So. we want to sample densely for β close to 1, because in that area, small β change makes big difference

deeplearning.ai

Hyperparameters tuning

Hyperparameters tuning in practice: Pandas vs. Caviar

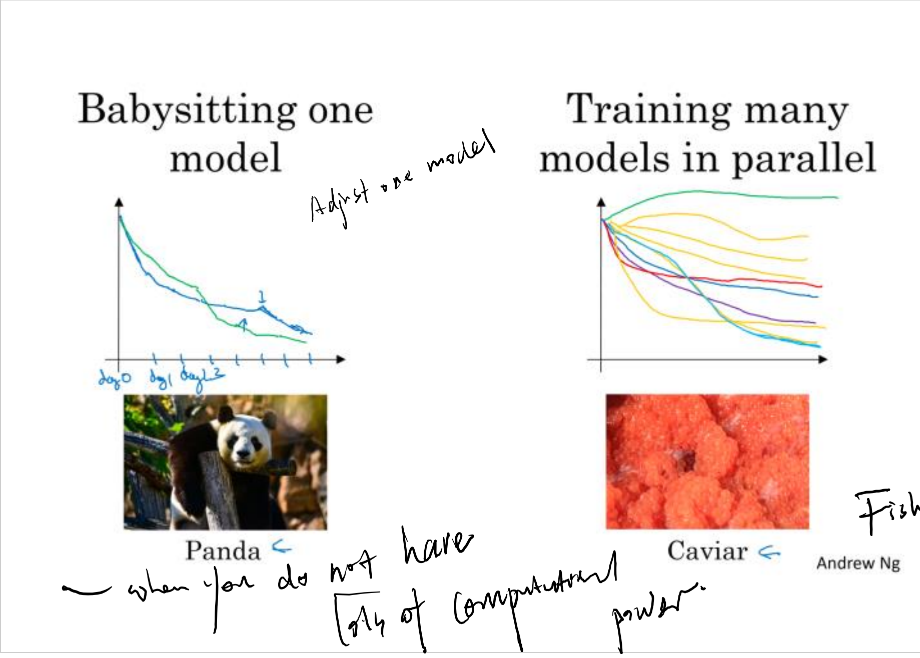
Re-test hyperparameters occasionally



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

Andrew Ng

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



努力调一个模型
细致地调参.

⇕

同时调不同模型.

A function of computational resources

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

⇓.