(Week 3)

→ Hyper parameter tuning

→ tuning process

• tips to organize hyper parameter tuning process

ex: hyper parameters ① α learning rate   1 ↝ most important to ture

                ② β momentum 2 ↝ 0.9 is good default

   (epilon)   ③ $\beta_1, \beta_2, \beta_3$ Param for Adam

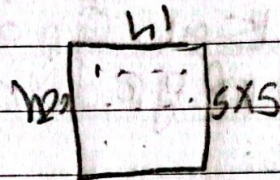   ④ # layers   ↝ 3

                ⑤ # hidden units   2 ↝

                ⑥ learning rate decay   ↝ 3

                ⑦ mini batch size   3 ↝

• to chose hyper parameters

   ① tryout hyper parameters randomly chosen h^1 ⬚ 5×5

   set of points

   ② using Coarse to fine when we found that the point

   work best & maybe a few points around that works well

   so we zoom in to smaller region of hyper para——

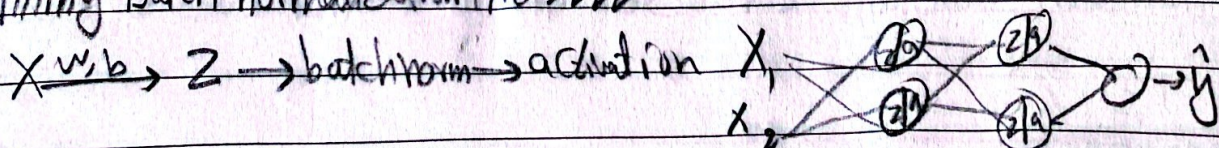→ using an appropriate scale to pick hyper parameters ✱

→ Baby sitting , train models in parallel

→ Batch normalization

   → Normalizing activation function

   • enable to train more easily even very deep networks &

   choice of hyperparam is a much bigger range of hyperparam

→ Fitting batch normalization int NN

   $X \xrightarrow{w,b} Z \rightarrow$ batchnorm $\rightarrow$ activation X_1 X_2 $\rightarrow \hat{y}$

* **Multi-Class classification**
  → softmax Regression
  ○ what if we have multiple possible classes? not binary
  $C =$ # of classes

  softmax @ Class
  @ activation

  → **Deep learning frameworks**
  - Caffe/Caffe2
  - CNTK
  - DL4J
  - keras
  - Lasagne
  - mxnet

  - paddlepaddle
  - tensorflow
  - Theano
  - Torch