Deep Learning

Dr. Mehran Safayani

safayani@iut.ac.ir

safayani.iut.ac.ir

https://www.aparat.com/mehran.safayani

https://github.com/safayani/deep_learning_course

Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

Hyperparameter tuning

Tuning process

Hyper parameters

① α

     momentum $\beta = 0.9$

     Adam $\underbrace{\beta_1}_{0.9}$, $\underbrace{\beta_2}_{0.999}$, $\underbrace{\varepsilon}_{10^{-8}}$
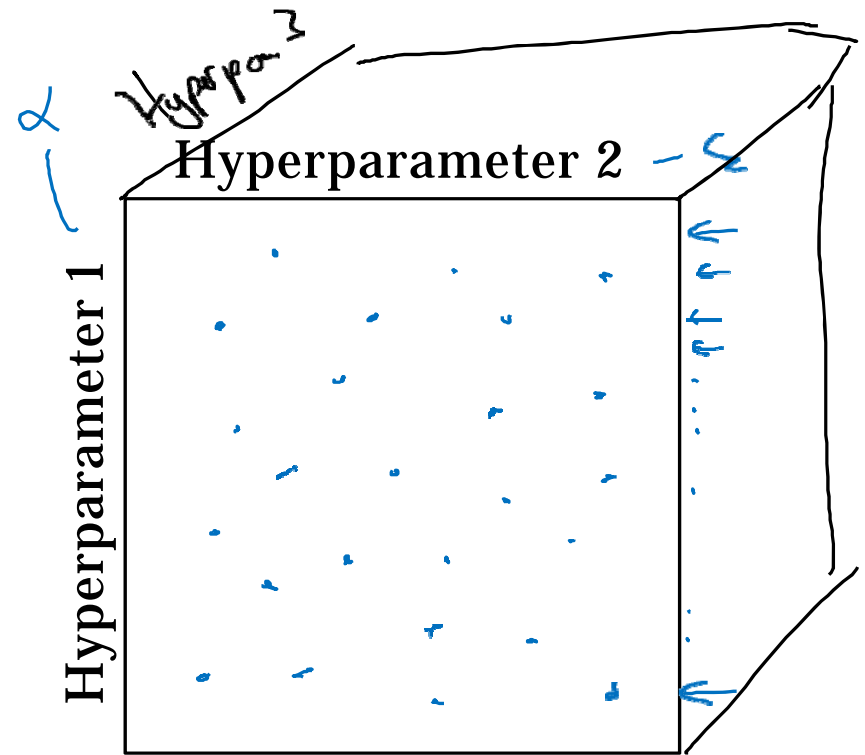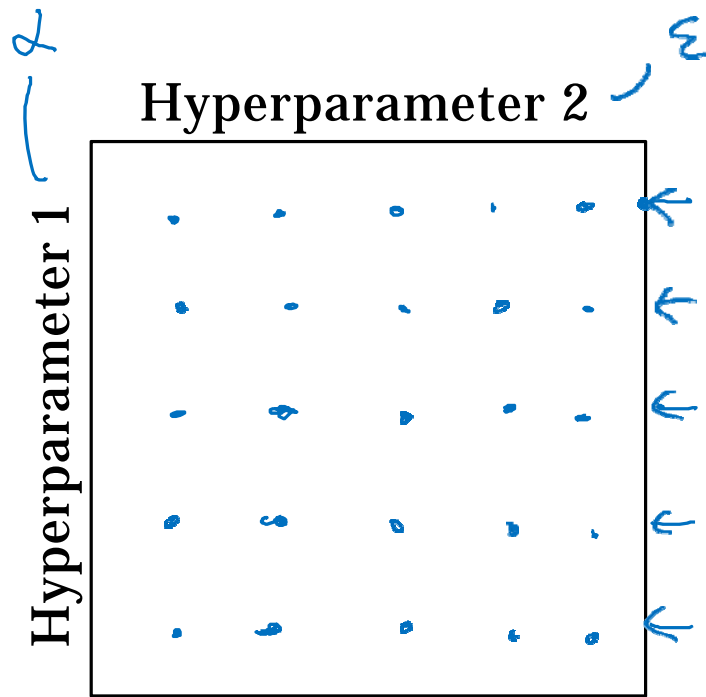
③   #layers

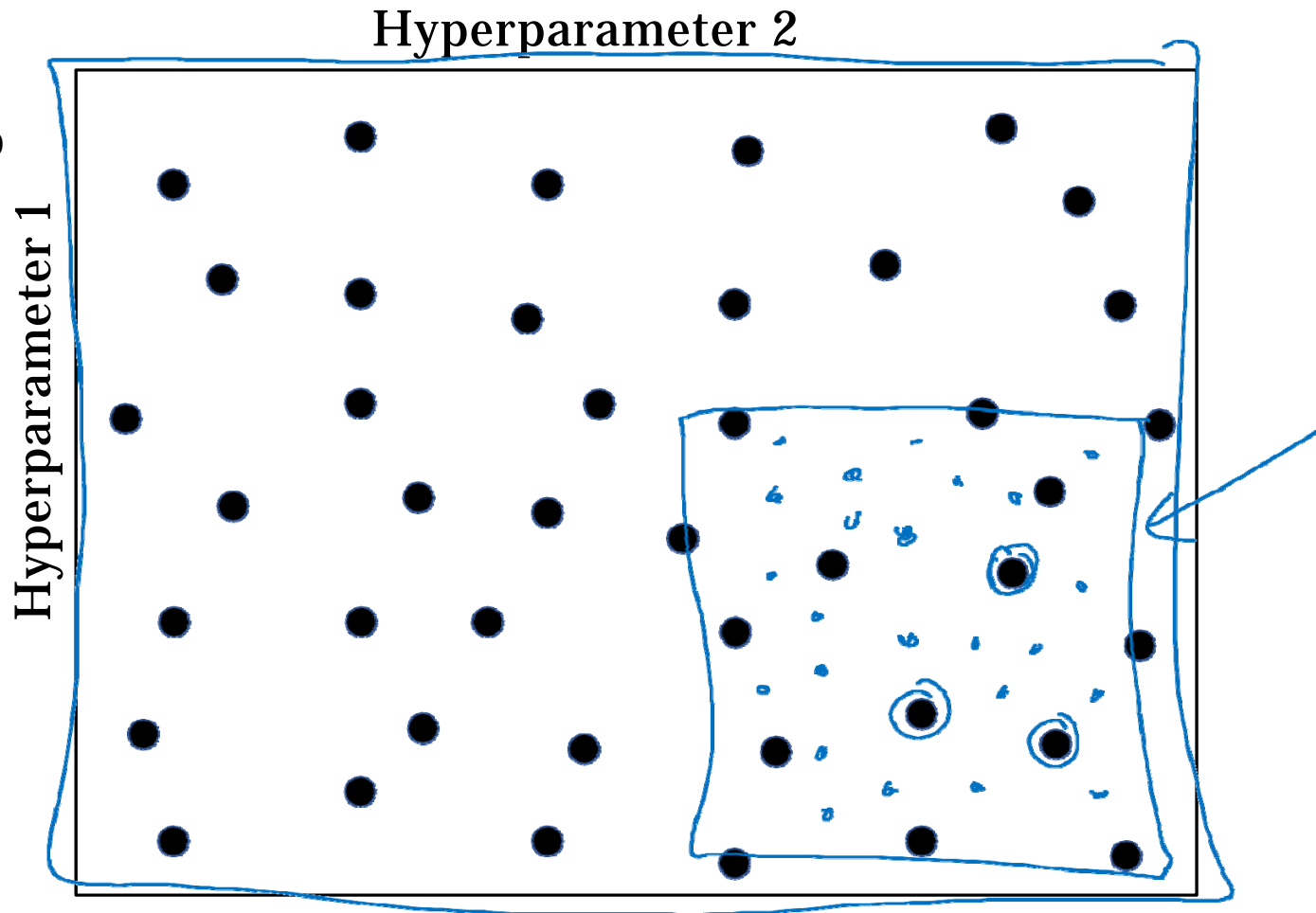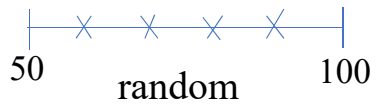②   #hidden units

③   Learning rate decay

②   mini-batch size
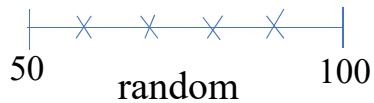
Try random values: Don't use a grid

Coarse to fine

- $n^{[l]} = 50, \ldots, 100$

50 ──×──×──×──×── 100
        random

Hyperparameter 2

Hyperparameter 1

Andrew

Using an appropriate scale to pick hyperparameters

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

50 ———×——×——×——×——— 100

random

- #Layers

  L: $2-4$

  $2,3,4$

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

0.0001 0.1 ———×——×——×——×——×——×——×——— **Uniform** **1**

0.0001   0.001 0.01   0.1   **1**
10^{-4} 10^{-3} 10^{-2} 10^{-1} 10^0

Andrew
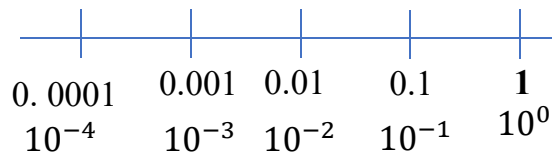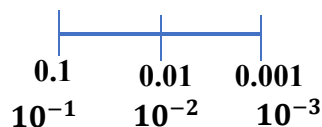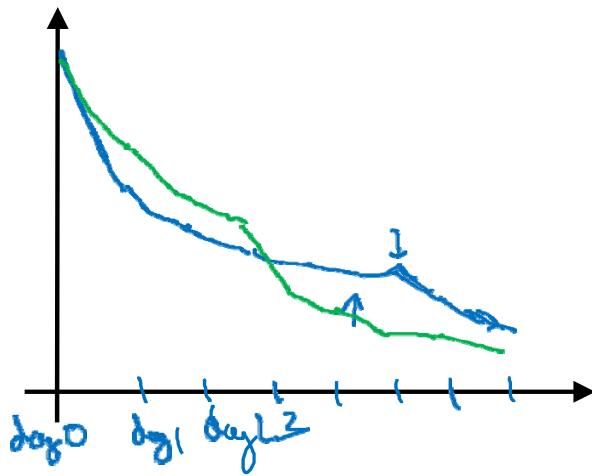
Using an appropriate scale to pick hyperparameters

- $r = -4 * np \cdot random \cdot rand()$ $r \in [-4, 0]$
- $\alpha = 10^r$
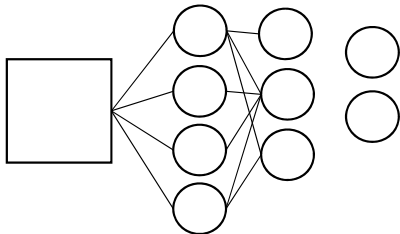- $\beta = 0.9 \cdots 0.999$
- $1 - \beta = 0.1 \cdots 0.001$

0.1	0.01	0.001
10^{-1}	10^{-2}	10^{-3}

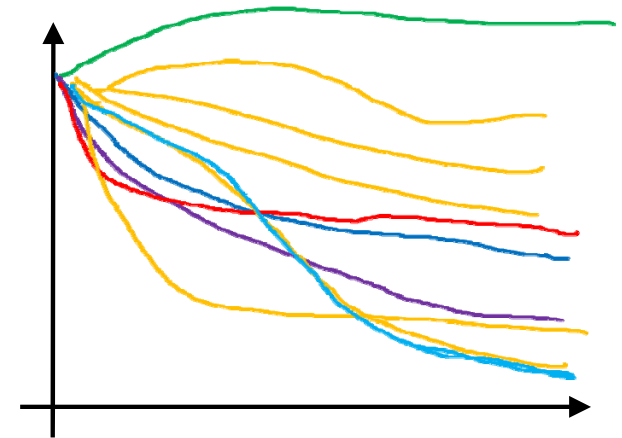- $r \in [-3, -1]$
  $1 - \beta = 10^r$
  $\beta = 1 - 10^r$

Babysitting one model

Panda

Training many models
in parallel

Caviar

Andrew