



# Python入门

• 一门语言。

1, 陈述 : `a = b`

2, 循环 : `for i in range(10):`

`print(i)`

3, 判断 : `if a > 3:`

`print(a)`

4, 函数 `def func(x):`  
`return x`

data structure

circulation

judge

function

class

assignment

random

numpy

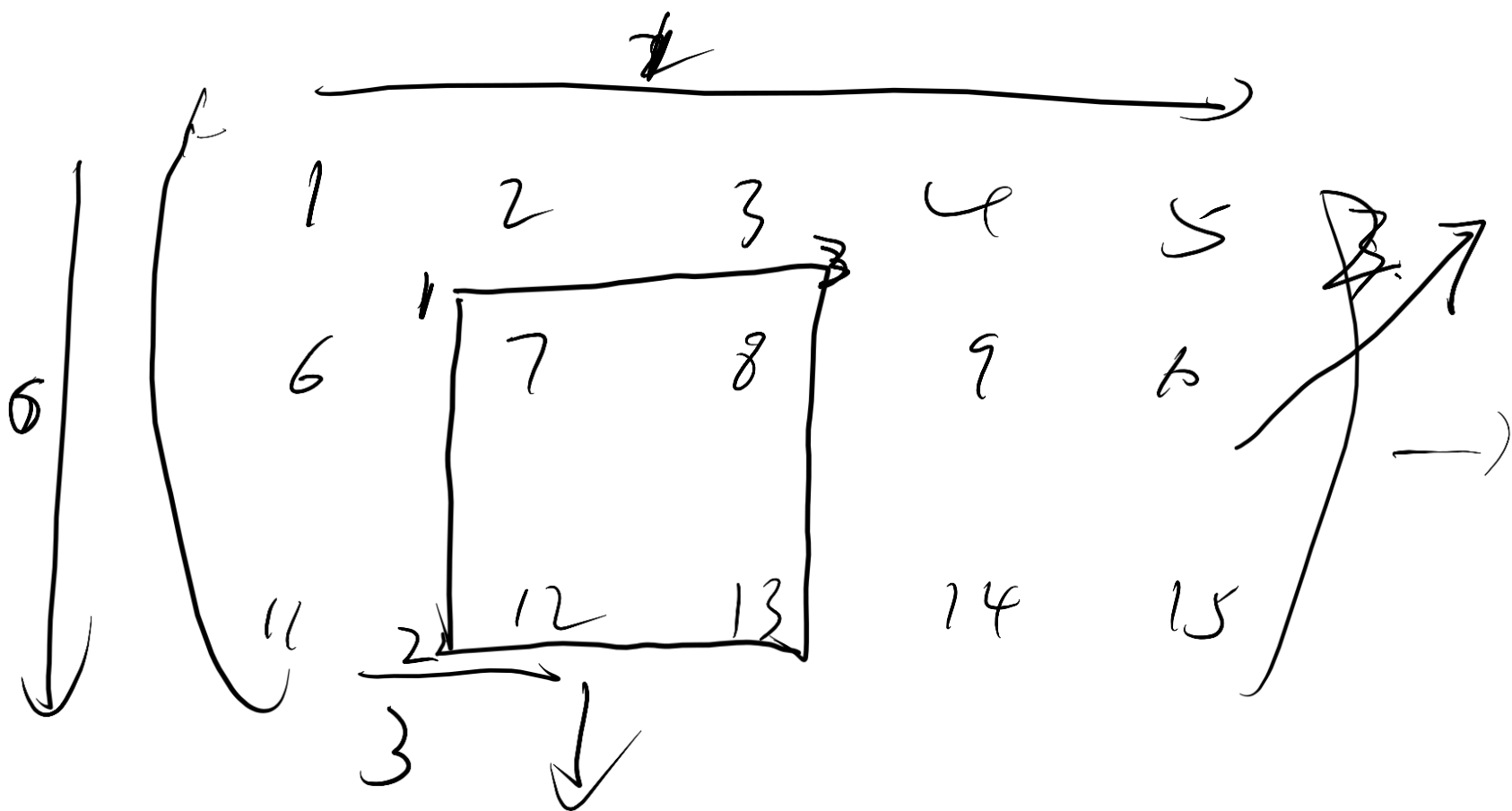
tensor

切片

包的引用

矩阵

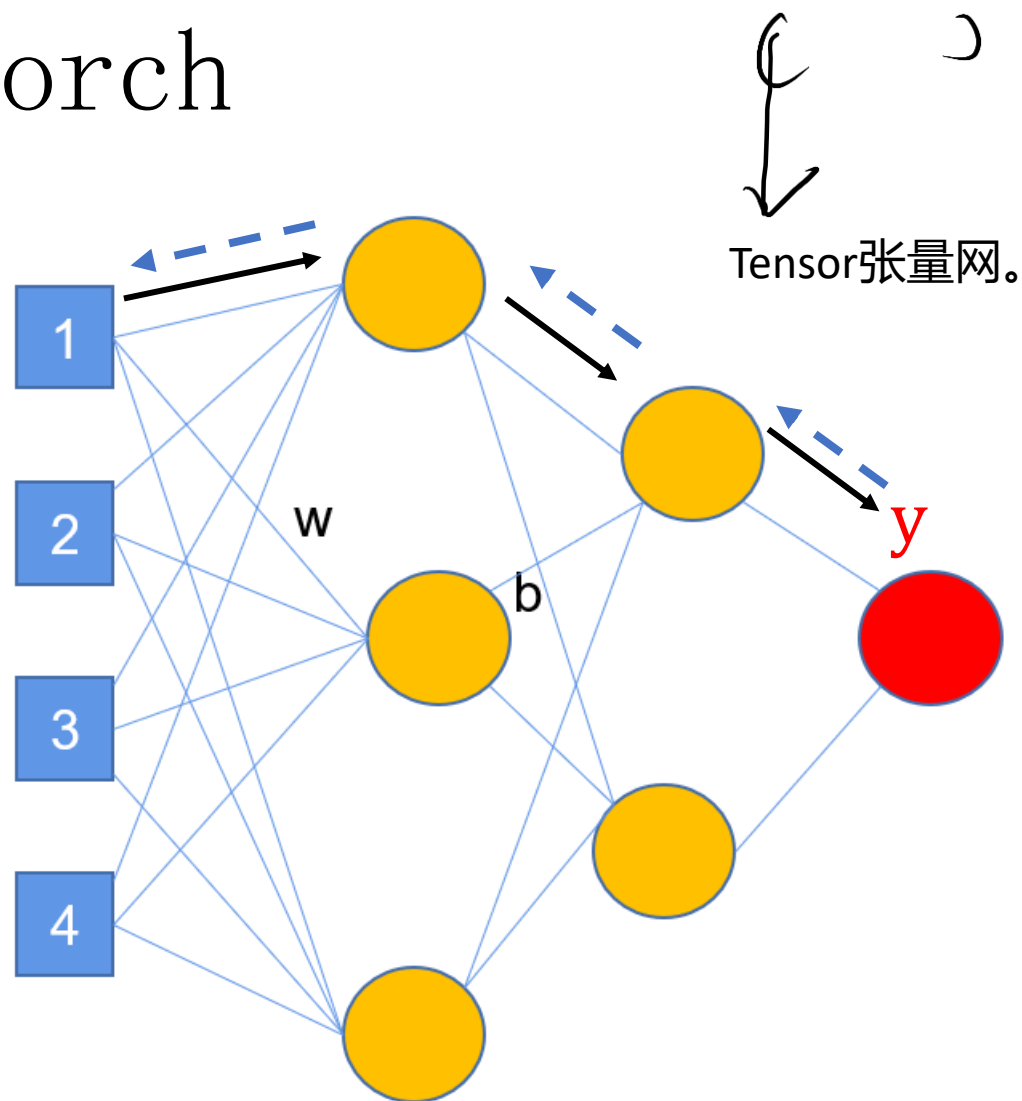
张量



# Pytorch



李哥 考研



Detach 表示取下来

带大家亲手从0 写一个线性回归。知道一个神经网络基本训练过程。

# 计算loss



李哥考研

x	$y = 2x + 1 + \epsilon$	$\hat{y} = 3x + 2$	loss
1	3.1	5	1.9
2	5.1	8	2.9
3	6.9	11	4.1
4	8.7	14	5.3
5	10.8	17	6.2
6	13.5	20	6.5
7	?		

4.48

Loss: 
$$L = \frac{1}{N} \sum_n l$$

# 数据

1 样: - | - | - | -

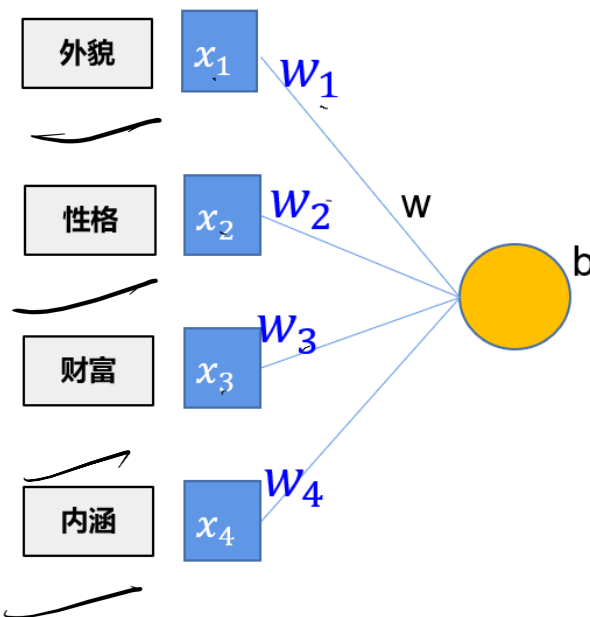


李哥考研

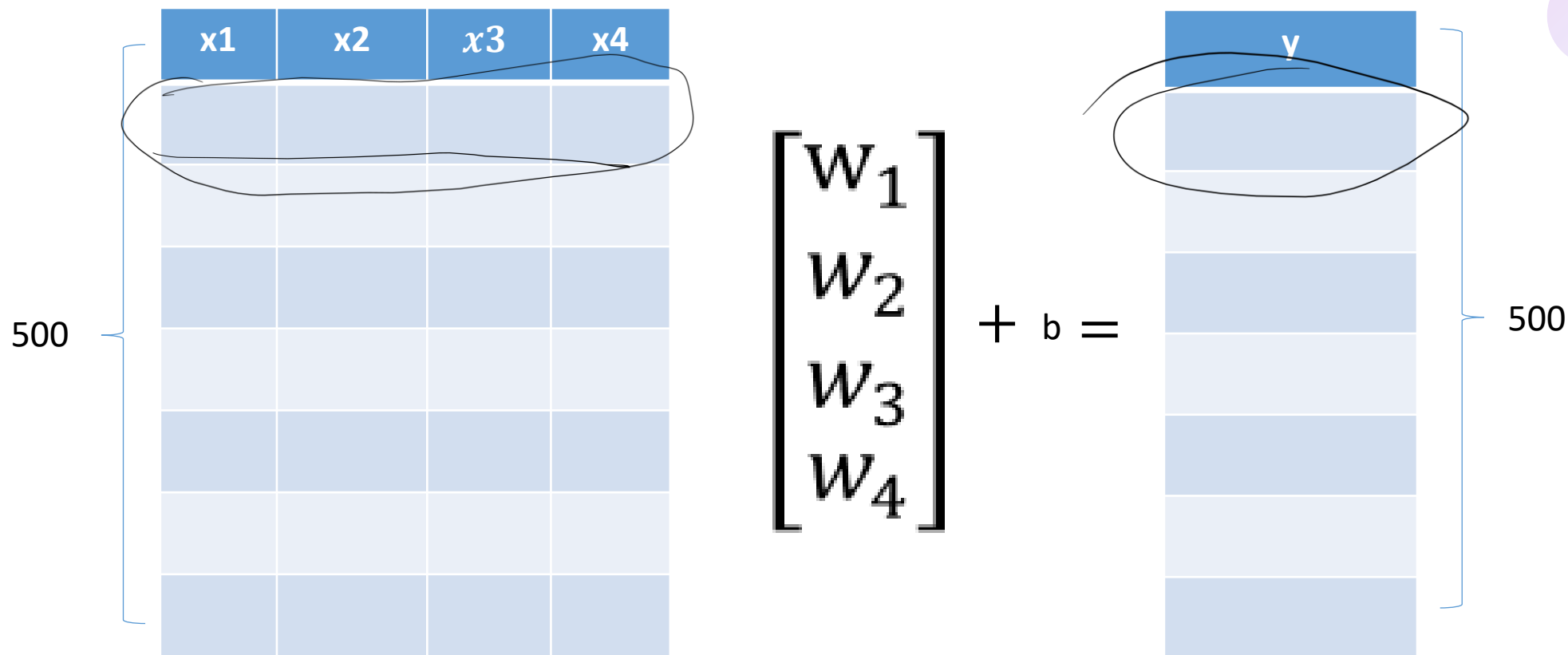
$$[y] = [b] + [w_{11} \ w_{12} \ w_{13} \ w_{14}] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$y = b + Wx$$

$w = [8.1, 2, 2, 4]$        $b = 1.1$



$$\hat{y} = b + \sum_j w_j x_j$$



$$\begin{matrix} \text{Orange box} & \text{Blue box} & + & \text{Green box} & = & \text{Grey box} \\ x & W & & b & & y \\ 500 \times 4 & 4 \times 1 & & 1 & & 500 \times 1 \end{matrix}$$

Batch 16

# 如何找一个函数呢？



李哥考研

x	y= 2x+1+ε
1	3.1
2	5.1
3	6.9
4	8.7
5	10.8
6	13.5
7	?

loss

Linear model

模型：

线性模型

预测值  
 $\hat{y} = w x + b$

**weight**

**bias**

(权重和偏差, 未知参数)

Loss function of unknown para :  $L(w, b) = |\hat{y} - y| = |xw + b - y|$

- Loss 就是这些未知参数的函数
- Loss: 判断我们选择的这组参数怎么样。

数据 (feature)    标签 (label)

# 通过线性回归，回顾之前内容。



李哥考研

$$\hat{y} = b + \sum_j w_j x_j$$

$$\frac{\partial L}{\partial w} \Big|_{w=w^0}$$

$$L(w, b) = |\hat{y} - y|$$

$$\text{Loss: } L = \frac{1}{N} \sum_n l$$

$$\begin{bmatrix} \theta_1^1 \\ \theta_2^1 \\ \vdots \end{bmatrix} \leftarrow \begin{bmatrix} \theta_1^0 \\ \theta_2^0 \\ \vdots \end{bmatrix} - \begin{bmatrix} \eta \frac{\partial L}{\partial \theta_1} \Big|_{\theta=\theta^0} \\ \eta \frac{\partial L}{\partial \theta_2} \Big|_{\theta=\theta^0} \\ \vdots \end{bmatrix}$$

代码文件: mylinear



### 3. Optimization



李哥考研

$$w^*, b^* = \arg \min_{w, b} L$$

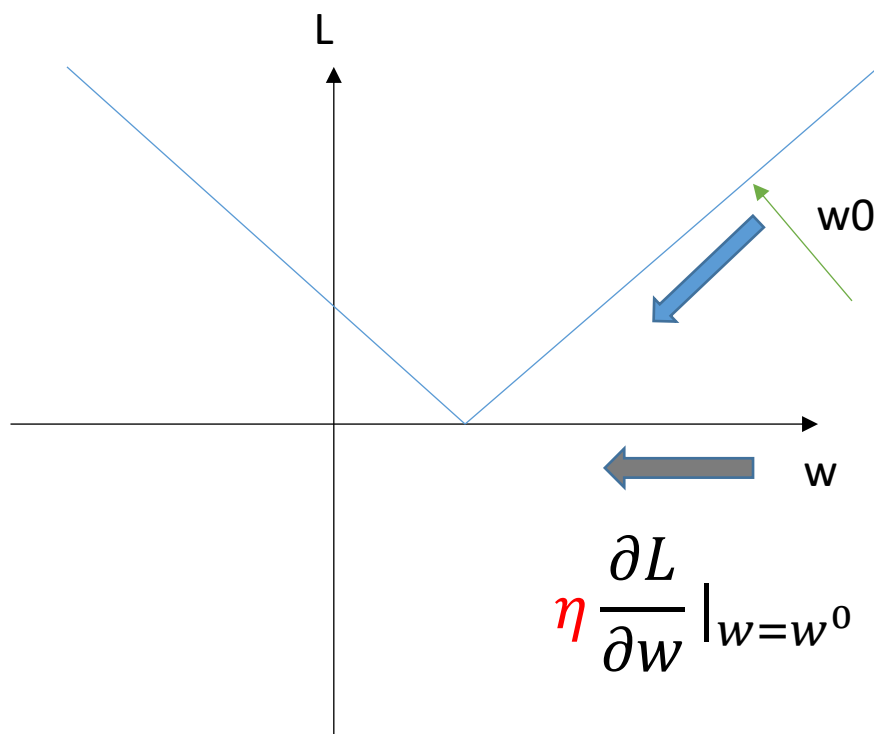
梯度下降

$$L(w, b) = |\hat{y} - y| = |xw + b - y|$$

➤ 随机选择一个  $w^0$

➤ 计算  $\frac{\partial L}{\partial w} \big|_{w=w^0}$

➤ 更新  $w$  的值



$$\eta \frac{\partial L}{\partial w} \big|_{w=w^0}$$

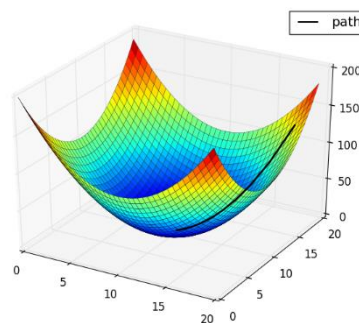
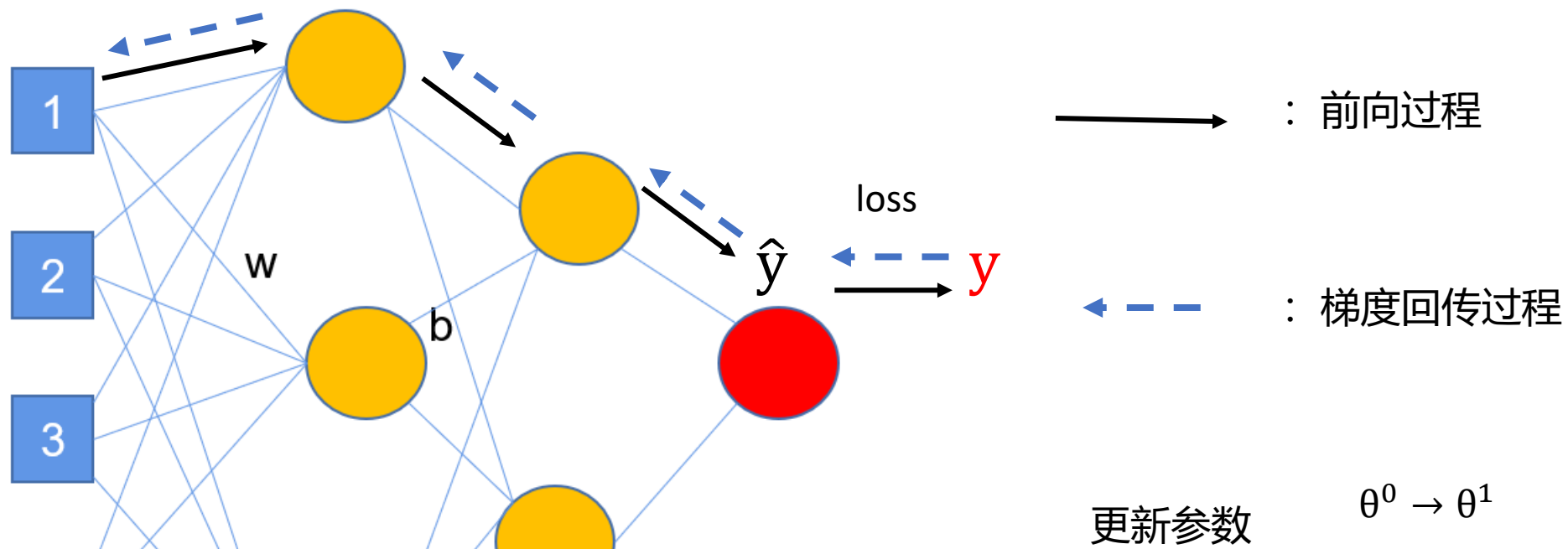
$\eta$ : learning rate

超参数



# 深度学习的训练过程。

## Gradient Descent



# 作业



李哥考研

注册CSDN，书写博客，解析刚才的代码。 即： 每一步代码是在干什么。



# Optimization of New Model

$$\theta^* = \arg \min_{\theta} L$$

➤ (Randomly) Pick initial values  $\theta^0$

➤ Compute gradient  $\mathbf{g} = \nabla L^1(\theta^0)$

$$\text{update } \theta^1 \leftarrow \theta^0 - \eta \mathbf{g}$$

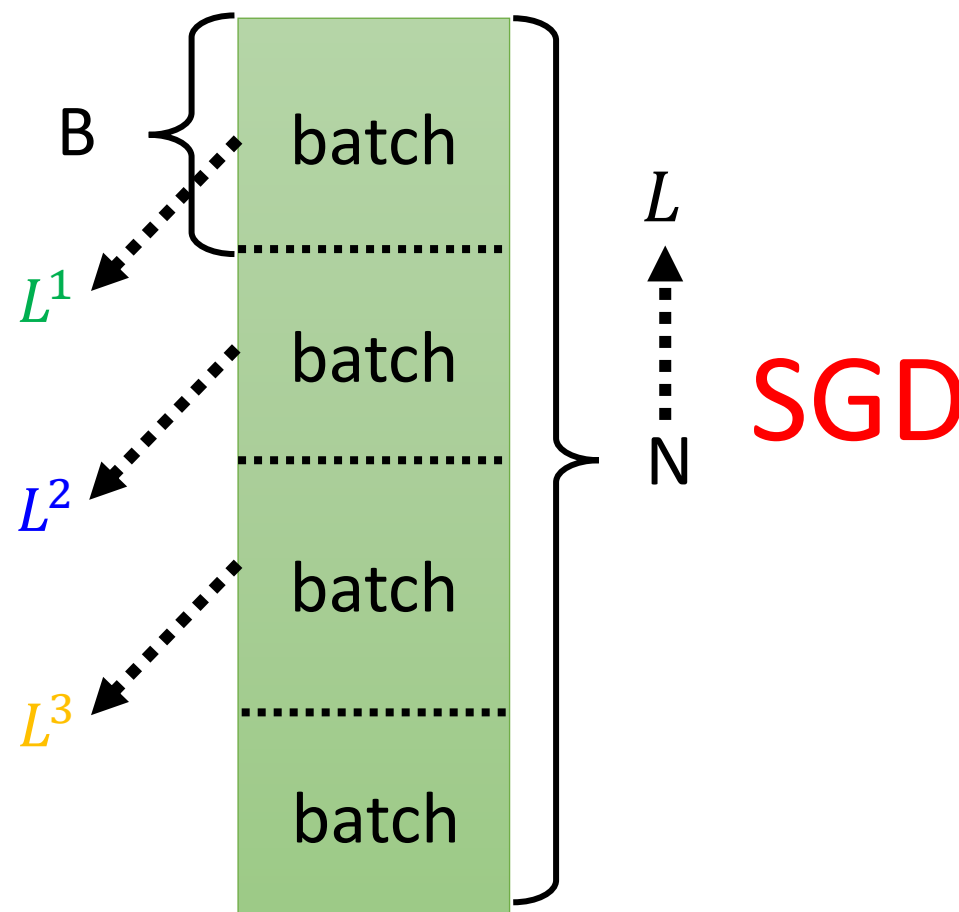
➤ Compute gradient  $\mathbf{g} = \nabla L^2(\theta^1)$

$$\text{update } \theta^2 \leftarrow \theta^1 - \eta \mathbf{g}$$

➤ Compute gradient  $\mathbf{g} = \nabla L^3(\theta^2)$

$$\text{update } \theta^3 \leftarrow \theta^2 - \eta \mathbf{g}$$

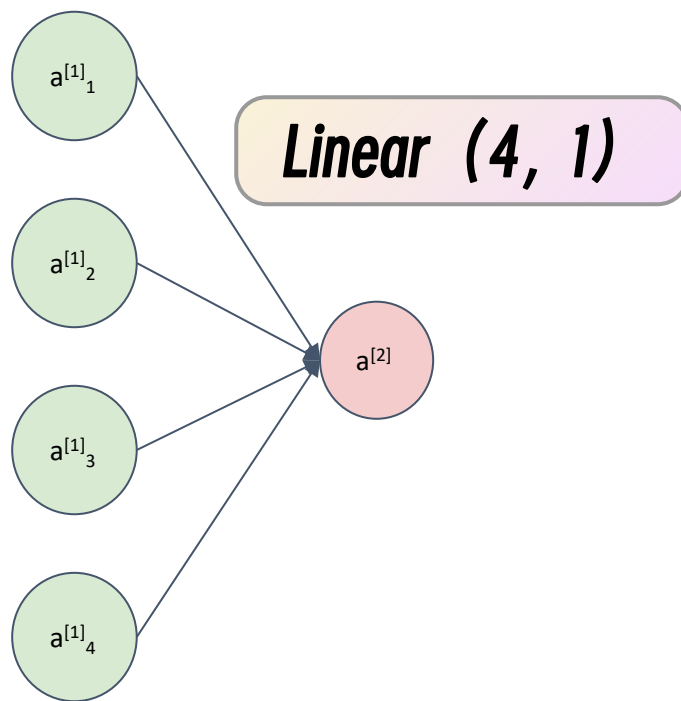
1 **epoch** = see all the batches once



# 回归



李哥考研





李哥考研

答疑和结束

THANKS

