回归问题

神经网络的输入值

总体的个数

$Y = f(x)$

神经网络进阶

$L = \frac{1}{2}(y - y')^2$

超过线的

模型输出

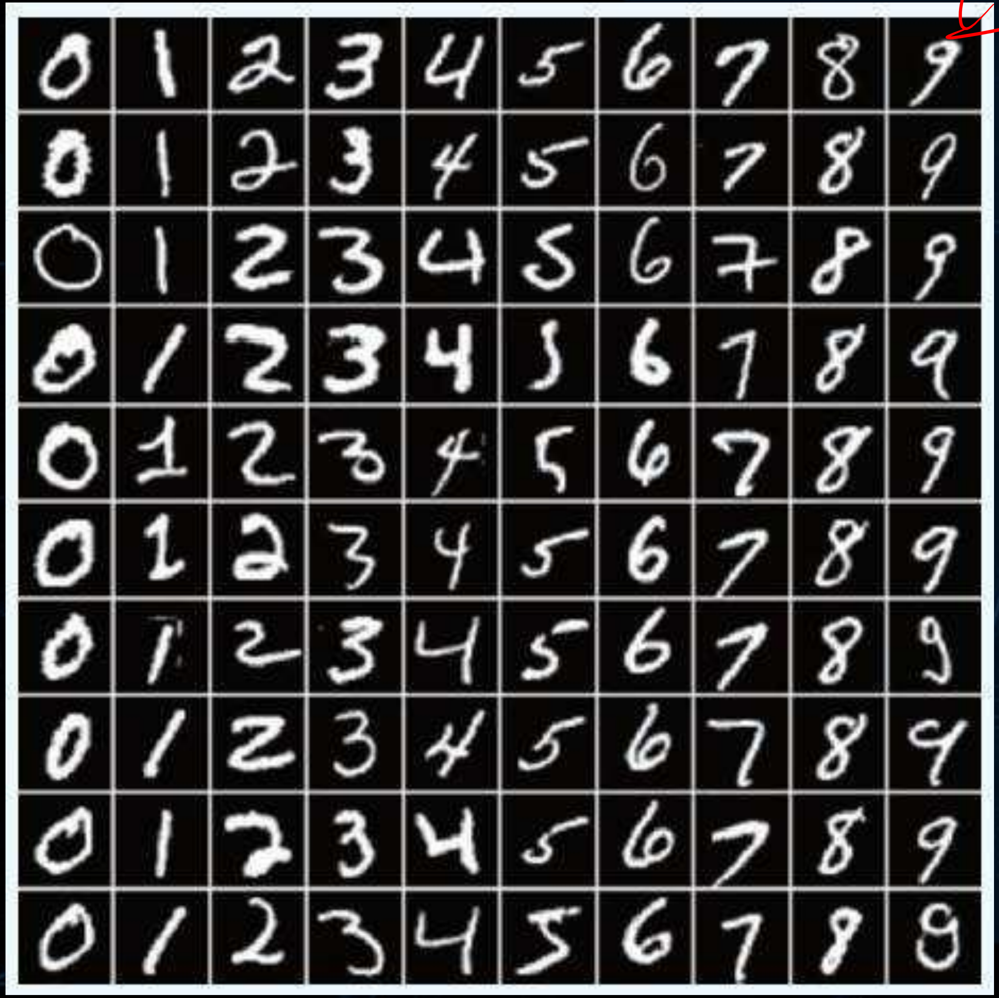
Steven Tang

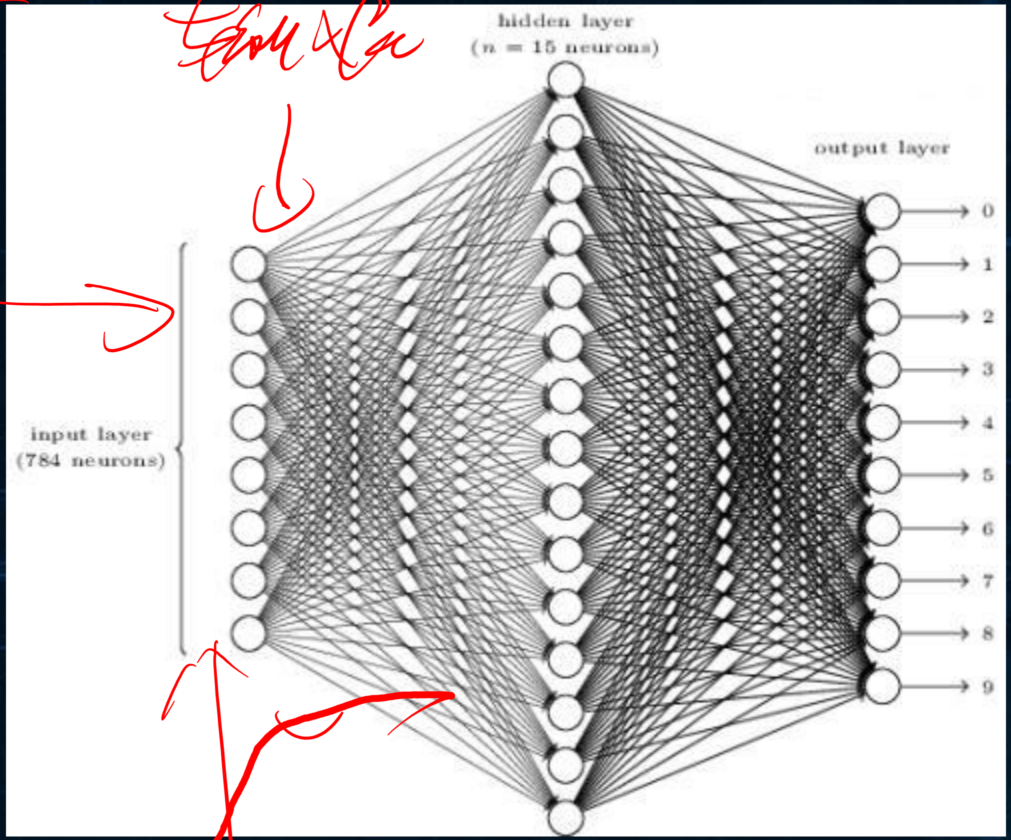
主要内容

- 神经网络分类问题
- Softmax函数
- 交叉熵损失
- 基于tensorflow的神经网络分类问题编程

渡一教育
Duyi Education

分类问题




hidden layer
(n = 15 neurons)
output layer
input layer
(784 neurons)
0
1
2
3
4
5
6
7
8
9

softmax

Wx+b

logits

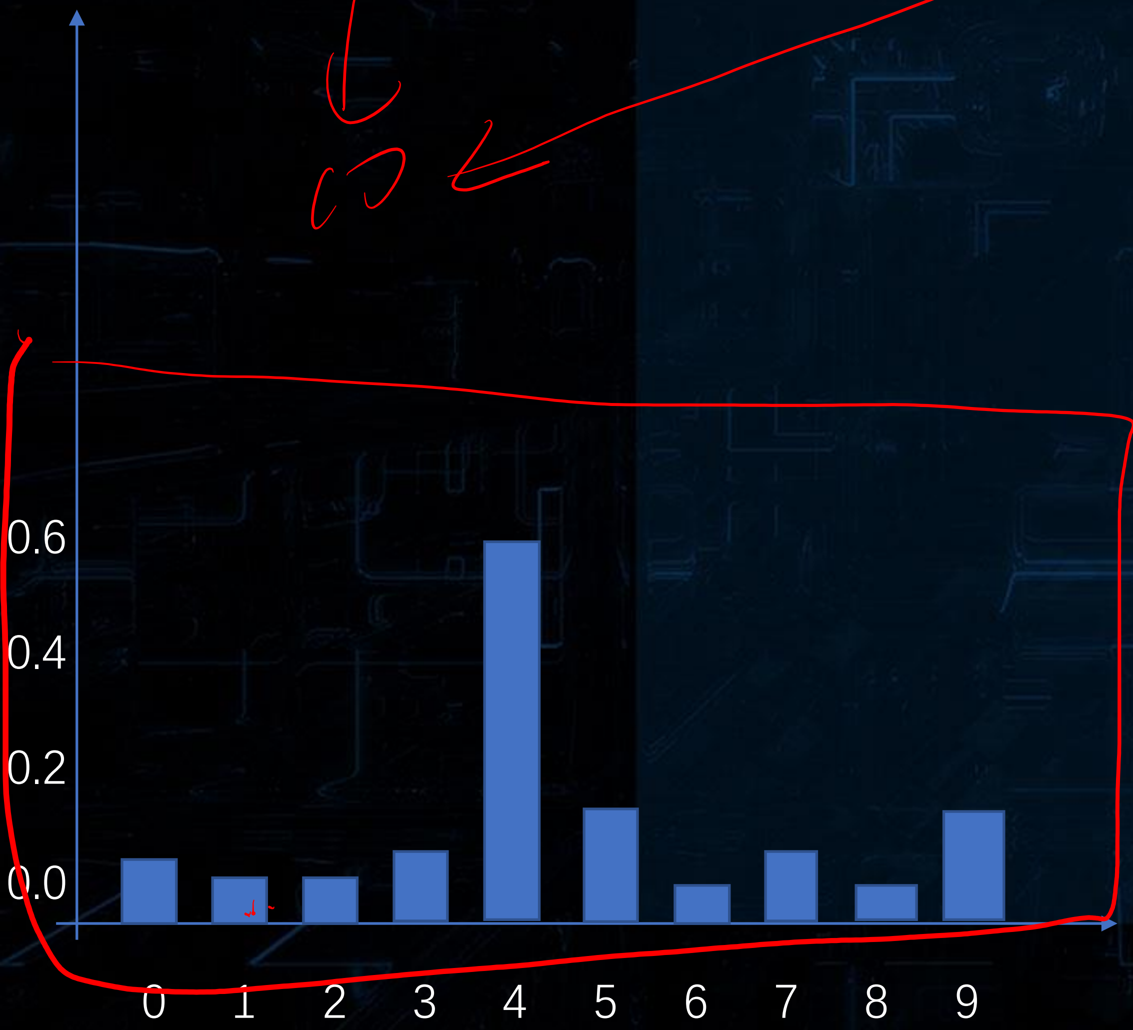
$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

$$\begin{bmatrix} 1.2 \\ 0.9 \\ 0.4 \end{bmatrix} \longrightarrow \boxed{\text{Softmax}} \longrightarrow \begin{bmatrix} 0.46 \\ 0.34 \\ 0.20 \end{bmatrix}$$

softmax

mnist

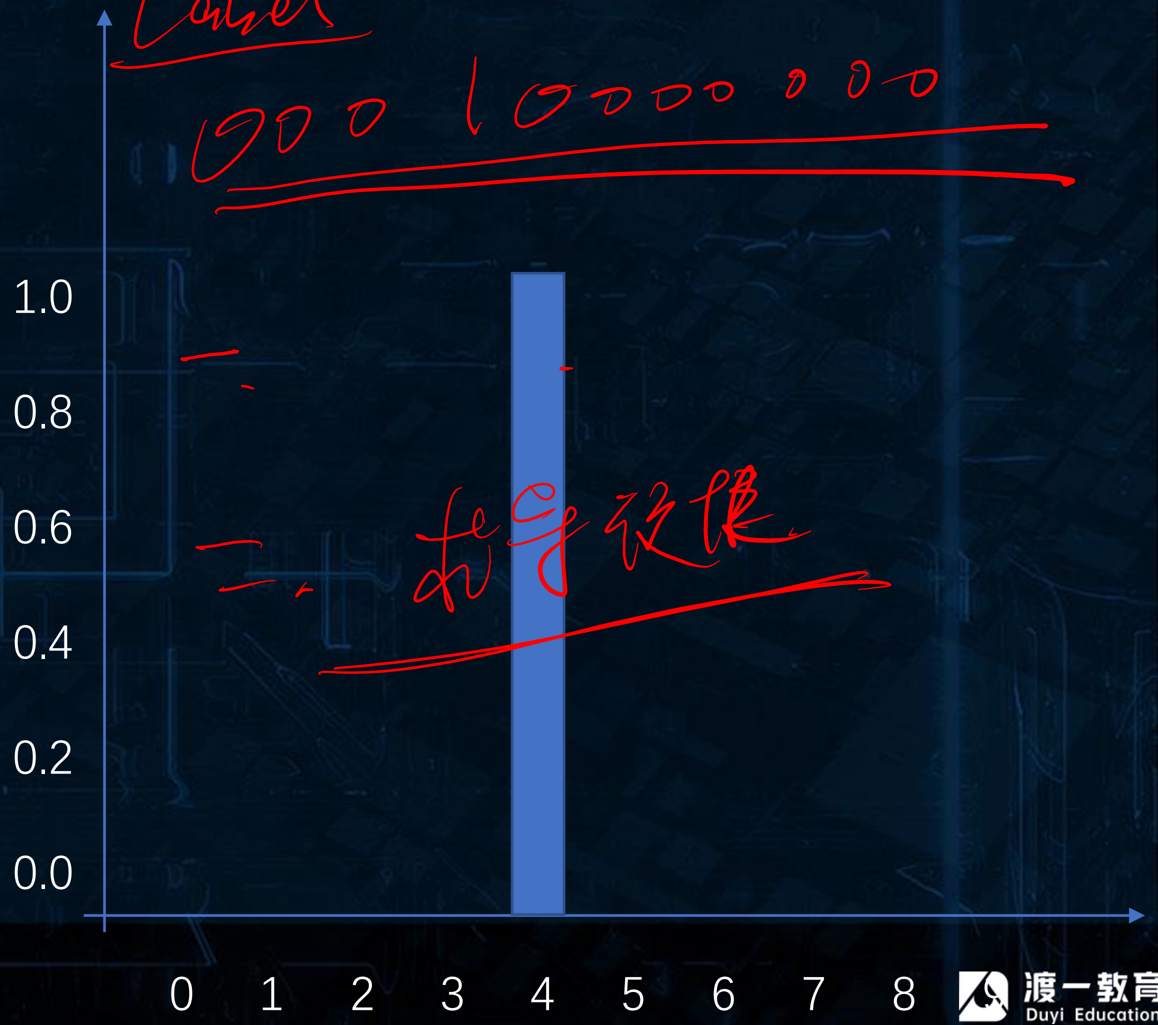
10



one_hot_encoding

Label

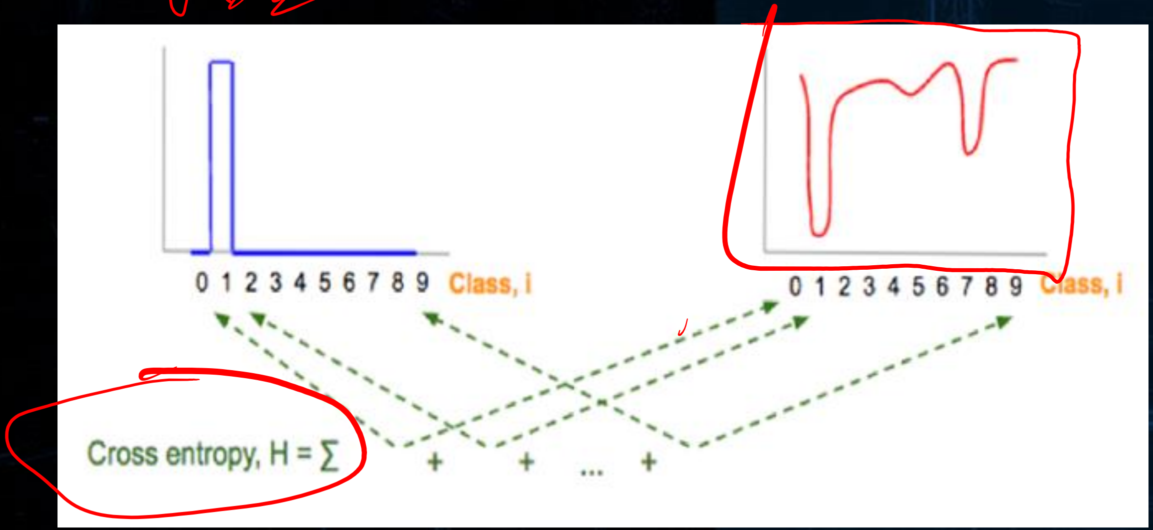
哑变量



Crossentropy（交叉熵）

logits	softmax	probability	crossentropy	one-hot
1.2	$$S(y_i)=\frac{e^{y_i}}{\sum_j e^j}$$	$p_0=0.46$	$$-\sum y_i \log(\hat{y}_i)$$	$c_0=1$
0.9		$p_1=0.34$		$c_1=0$
0.4		$p_2=0.2$		$c_2=0$

Crossentropy（交叉熵）



$$\hat{\mathbf{y}} = \begin{bmatrix} 0.1 \\ 0.5 \\ 0.4 \end{bmatrix} \qquad D(\hat{\mathbf{y}}, \mathbf{y}) = -\sum_j y_j \ln \hat{y}_j \qquad \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$\to \ln 0.5$

Crossentropy（交叉熵）

输入	输出logits	概率probability	标签	结果
	老虎：1.5 猫：0.7 兔子：0.2	0.57 0.25 0.17	1 0 0	正确
	老虎：1.1 猫：2.5 兔子：1.5	0.15 0.61 0.22	0 1 0	正确
	老虎：2.0 猫：1.5 兔子：1.1	0.49 0.30 0.21	0 0 1	错误

Crossentropy（交叉熵）

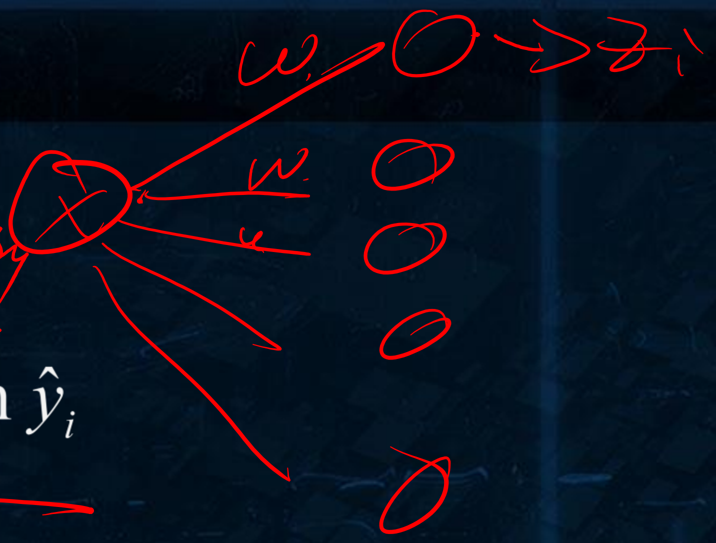
$$z_i = w_i * x + b$$

softmax

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

cross-entropy

$$L = -\sum_{i=1}^{n} y_i \ln \hat{y}_i$$



$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial z_i}\frac{\partial z_i}{\partial w_i}$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial z_i}\frac{\partial z_i}{\partial b}$$

$$\frac{\partial z_i}{\partial w_i} = x \qquad \frac{\partial z_i}{\partial b} = 1$$

$$\frac{\partial L}{\partial z_i} = \sum_{k=1}^{n} \frac{\partial L}{\partial \hat{y}_k}\frac{\partial \hat{y}_k}{\partial z_i}$$

$$\frac{\partial L}{\partial \hat{y}_k} = \frac{\partial\left(-\sum_{j=1}^{n} y_j \ln \hat{y}_j\right)}{\partial \hat{y}_k} = -\frac{y_k}{\hat{y}_k}$$

Crossentropy（交叉熵）

$$z_i = w_i * x + b$$

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

$$L = -\sum_{i=1}^{n} y_i \ln \hat{y}_i$$

$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial z_i} \frac{\partial z_i}{\partial w_i}$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial z_i} \frac{\partial z_i}{\partial b}$$

$$\frac{\partial w_i}{\partial b} = \mathrm{x}, \frac{\partial z_i}{\partial b} = 1$$

$$\frac{\partial L}{\partial z_i} = \sum_{k=1}^{n} \frac{\partial L}{\partial \hat{y}_k} \frac{\partial \hat{y}_k}{\partial z_i}$$

$$\frac{\partial L}{\partial \hat{y}_k} = \frac{\partial\left(-\sum_{j=1}^{n} y_j \ln \hat{y}_j\right)}{\partial \hat{y}_k} = -\frac{y_k}{\hat{y}_k}$$

Crossentropy（交叉熵）

softmax

$(f\cdot u)' = f'u + f\cdot u'$

最后来求 $\frac{\partial \hat{y}_k}{\partial z_i}$，这个分为两个情况：

第一种情况当k≠i时：

$$\frac{\partial \hat{y}_k}{\partial z_i} = \frac{\partial\left(\frac{e^{z_k}}{\sum_{j=1}^{n} e^{z_j}}\right)}{\partial z_i} = -e^{z_k}\frac{1}{\left(\sum_{j=1}^{n} e^{z_j}\right)^2}e^{z_i} = -\frac{e^{z_k}}{\left(\sum_{j=1}^{n} e^{z_j}\right)}\frac{e^{z_i}}{\left(\sum_{j=1}^{n} e^{z_j}\right)} = -\hat{y}_k\hat{y}_i$$

第二种情况是当k=i时：

$$\frac{\partial \hat{y}_k}{\partial z_i} = \frac{\partial \hat{y}_i}{\partial z_i} = \frac{\partial\left(\frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}\right)}{\partial z_i} = \frac{e^{z_i}\sum_{j=1}^{n} e^{z_j} - (e^{z_i})^2}{\left(\sum_{j=1}^{n} e^{z_j}\right)^2} = \frac{e^{z_i}}{\left(\sum_{j=1}^{n} e^{z_j}\right)}\frac{\sum_{j=1}^{n} e^{z_j} - e^{z_i}}{\left(\sum_{j=1}^{n} e^{z_j}\right)} = \hat{y}_i(1-\hat{y}_i)$$

Crossentropy（交叉熵）

$$\frac{\partial L}{\partial z_i}=\sum_{k=1}^{n}\left[\frac{\partial L}{\partial \hat{y}_k}\frac{\partial \hat{y}_k}{\partial z_i}\right]=\sum_{k=1}^{n}\left[-\frac{y_k}{\hat{y}_k}\frac{\partial \hat{y}_k}{\partial z_i}\right]$$

$$=-\frac{y_i}{\hat{y}_i}\frac{\partial \hat{y}_i}{\partial z_i}+\sum_{k=1,k\neq i}^{n}\left[-\frac{y_k}{\hat{y}_k}\frac{\partial \hat{y}_k}{\partial z_i}\right]$$

$$=-\frac{y_i}{\hat{y}_i}\hat{y}_i(1-\hat{y}_i)+\sum_{k=1,k\neq i}^{n}\left[-\frac{y_k}{\hat{y}_k}(-\hat{y}_k\hat{y}_i)\right]$$

$$=y_i(\hat{y}_i-1)+\sum_{k=1,k\neq i}^{n}y_k\hat{y}_i$$

$$=-y_i+y_i\hat{y}_i+\sum_{k=1,k\neq i}^{n}y_k\hat{y}_i$$

$$=-y_i+\sum_{k=1}^{n}y_k\hat{y}_i$$

由于在多分类问题中，$\sum_{k=1}^{n} y_k=1$ ，所以最终的结果为：

$$\hat{y}_i-y_i$$

将上式代入得到：

$$\hat{y}_i-y_i\ \frac{\partial L}{\partial w_i}=(\hat{y}_i-y_i)\bullet x$$

$$\frac{\partial L}{\partial w_i}=\hat{y}_i-y_i$$

渡一教育 Duyi Education

迭代训练法

- 一个epoch代表所有的数据都训练过了一遍，总共调整多少次参数有batch_size决定。
- 通常会使用训练数据集训练多个epochs，在中间每隔若干个epochs可以将验证数据集输入查看模型性能。
- 所有的epochs训练完以后将测试数据集输入，查看最终的模型性能。
-

渡一教育 Duyi Education

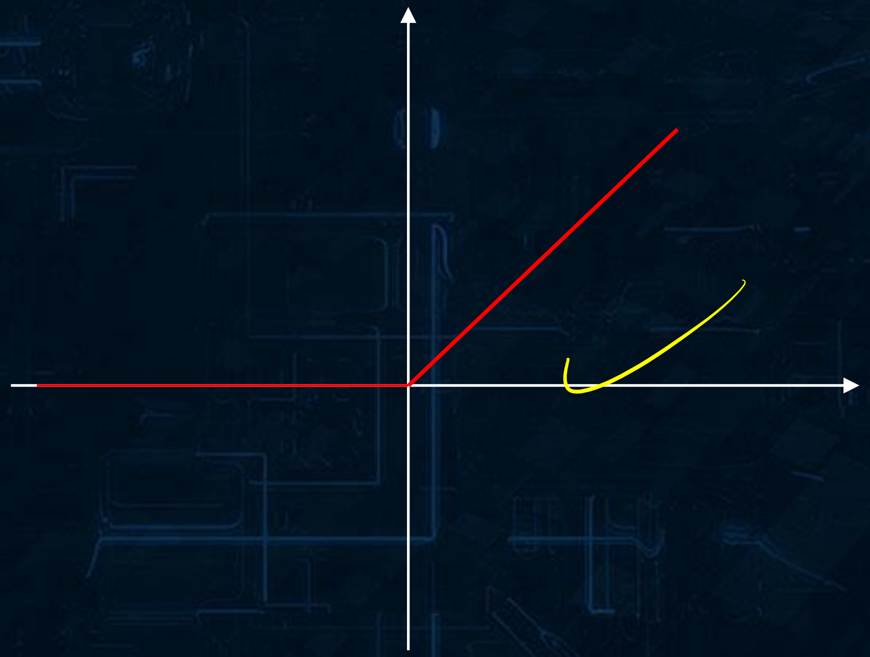
主要内容

- Relu激活函数
- 模型存储
- L1&L2正则项
- 自编码器神经网络

Relu激活函数

$$f(x) = \max(x, 0)$$

$$f'(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x <= 0 \end{cases}$$



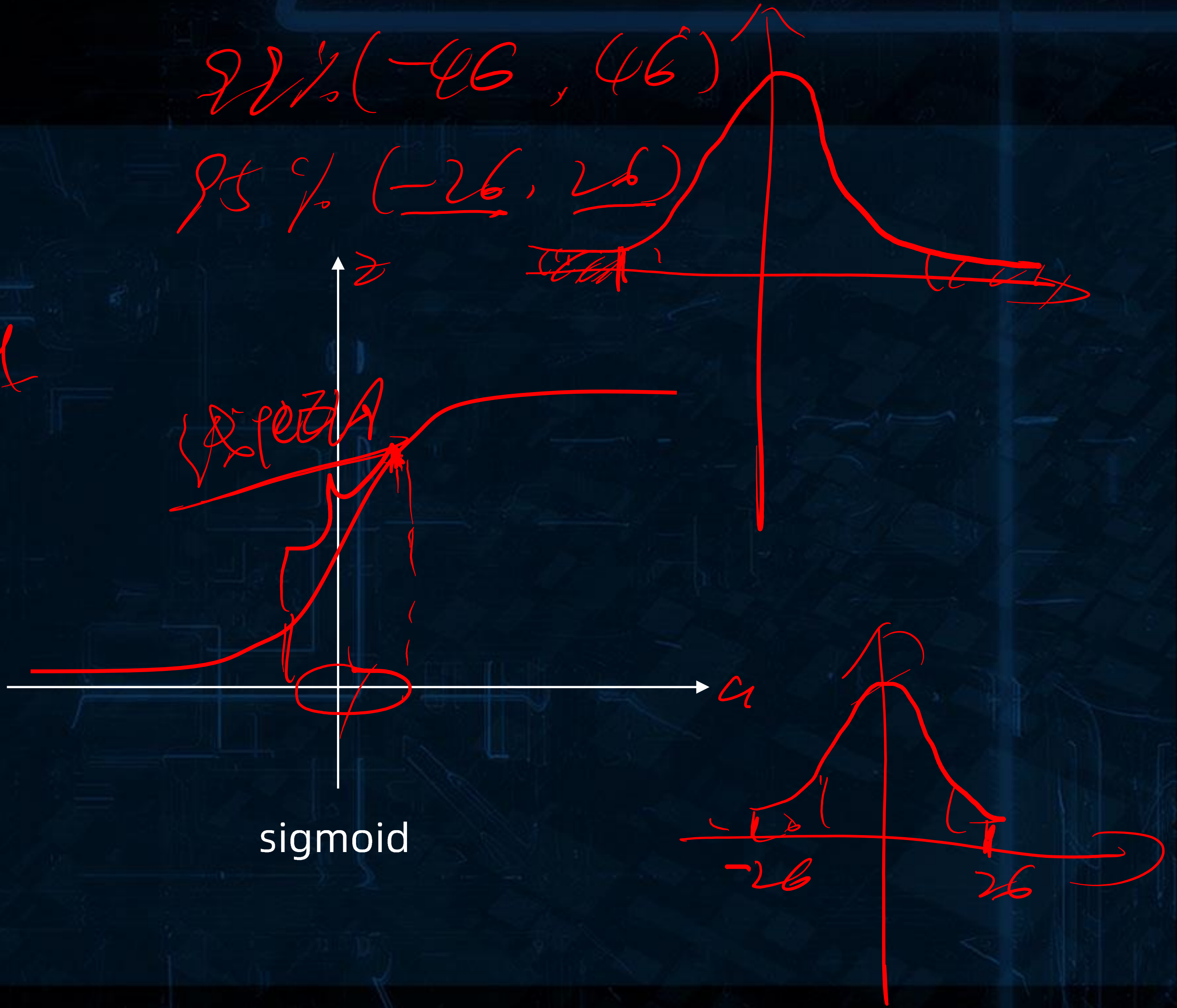
Relu

sigmoid激活函数的问题

truncated

$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x) * (1 - f(x))$$



sigmoid

Save

V1:(3,3)

V2(3,4)

V3(4,5)

Ckpt

V1: 3×3

V2 3×4

V3: 4×5

Resume

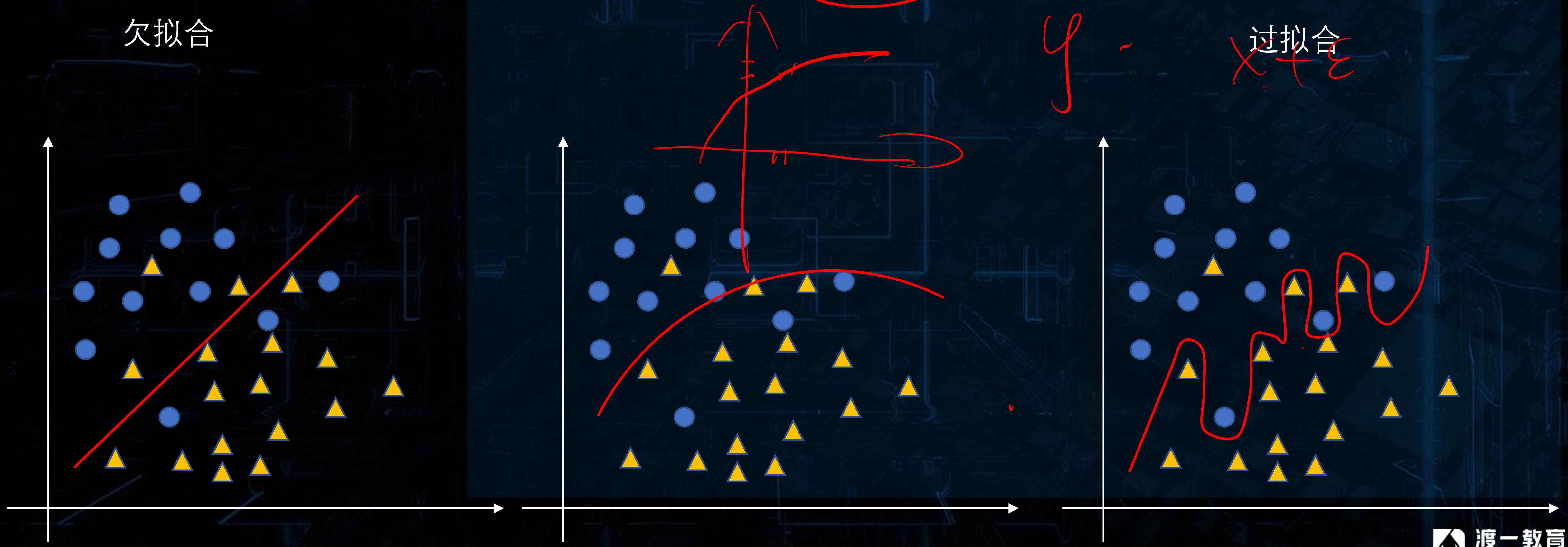
V1:(3,3)

正则化原因

过拟合　泛化能力过弱

参数　$Y = f_\theta(x)$

欠拟合　过拟合



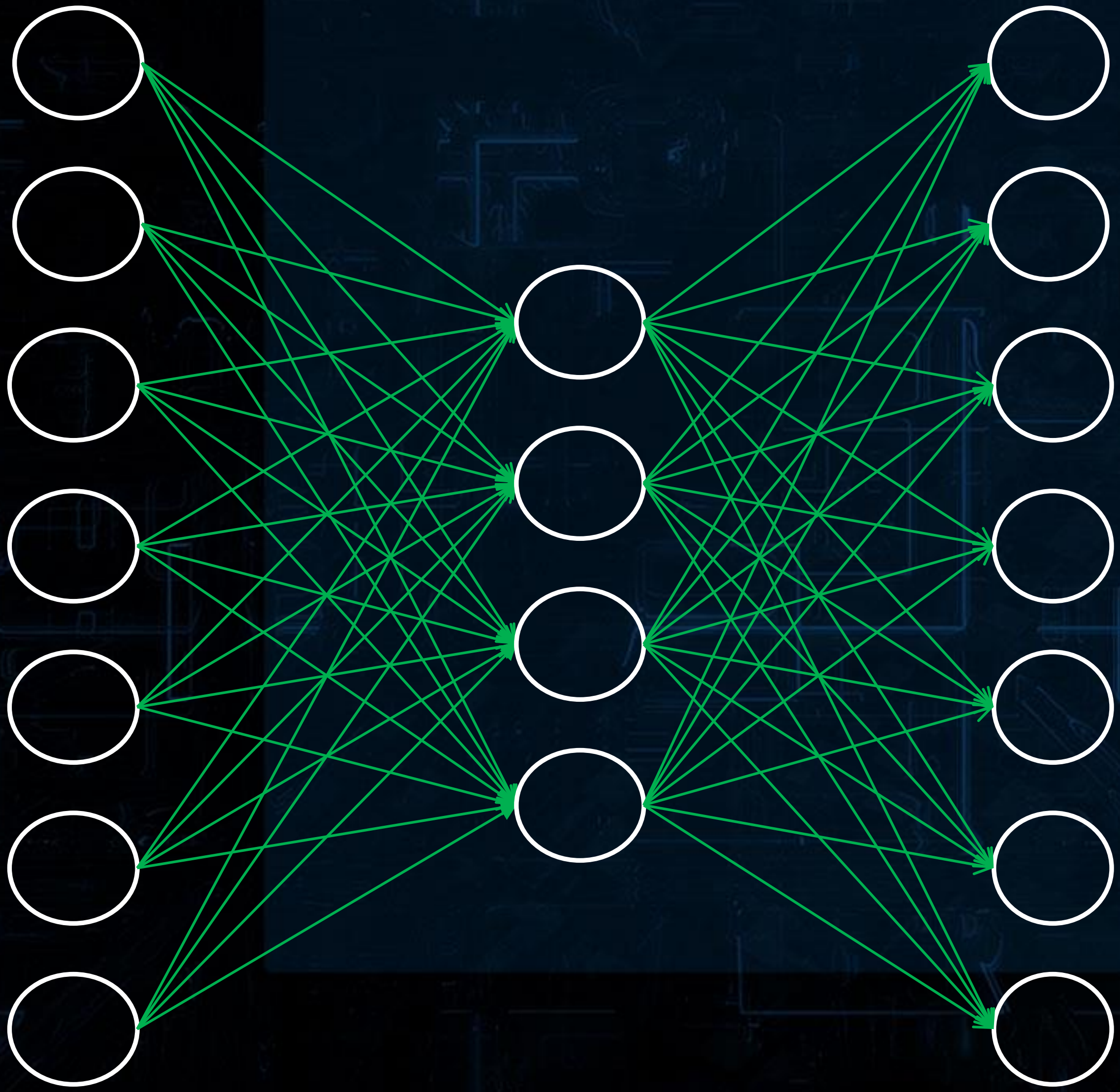
正则项（L1&L2范数)

$$\|\mathbf{x}\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$

$$\|\mathbf{x}\|_1 = |x_0| + |x_1| + |x_2| + \ldots + |x_{n-1}|$$

$$\|\mathbf{x}\|_2 = \sqrt{x_0^2 + x_1^2 + x_2^2 + \ldots + x_{n-1}^2}$$

自编码器



$$F_\theta(X) = X$$

无监督学习

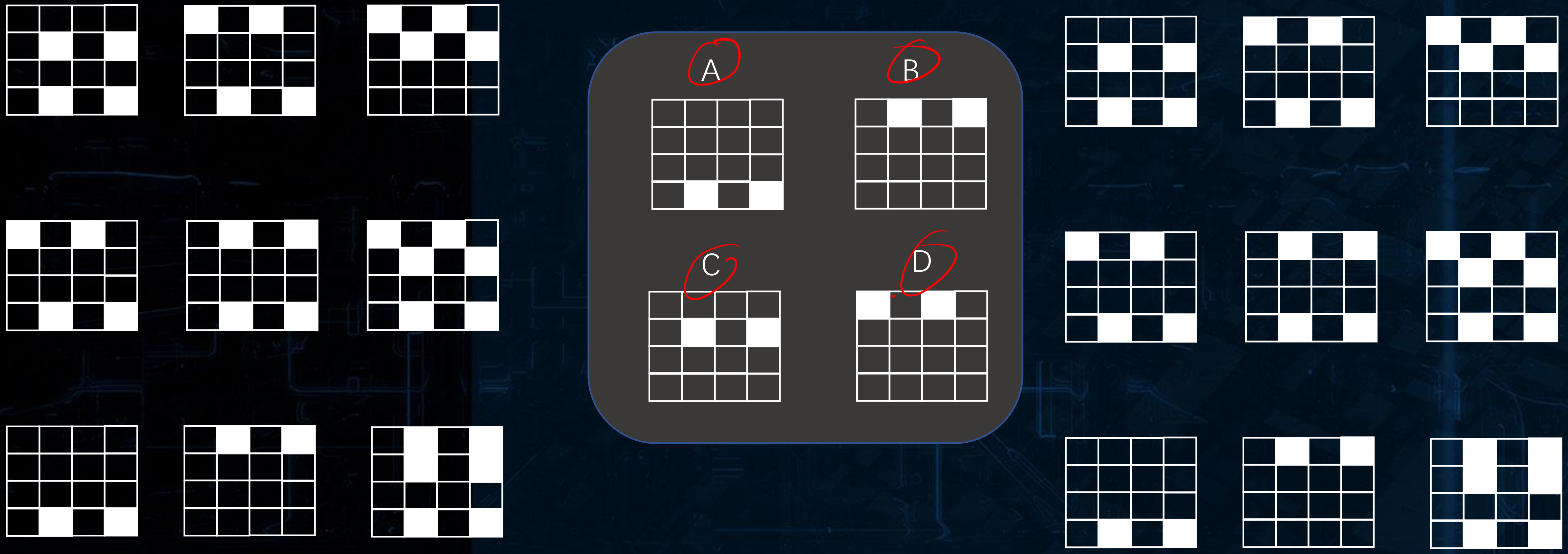
unsupervised learning

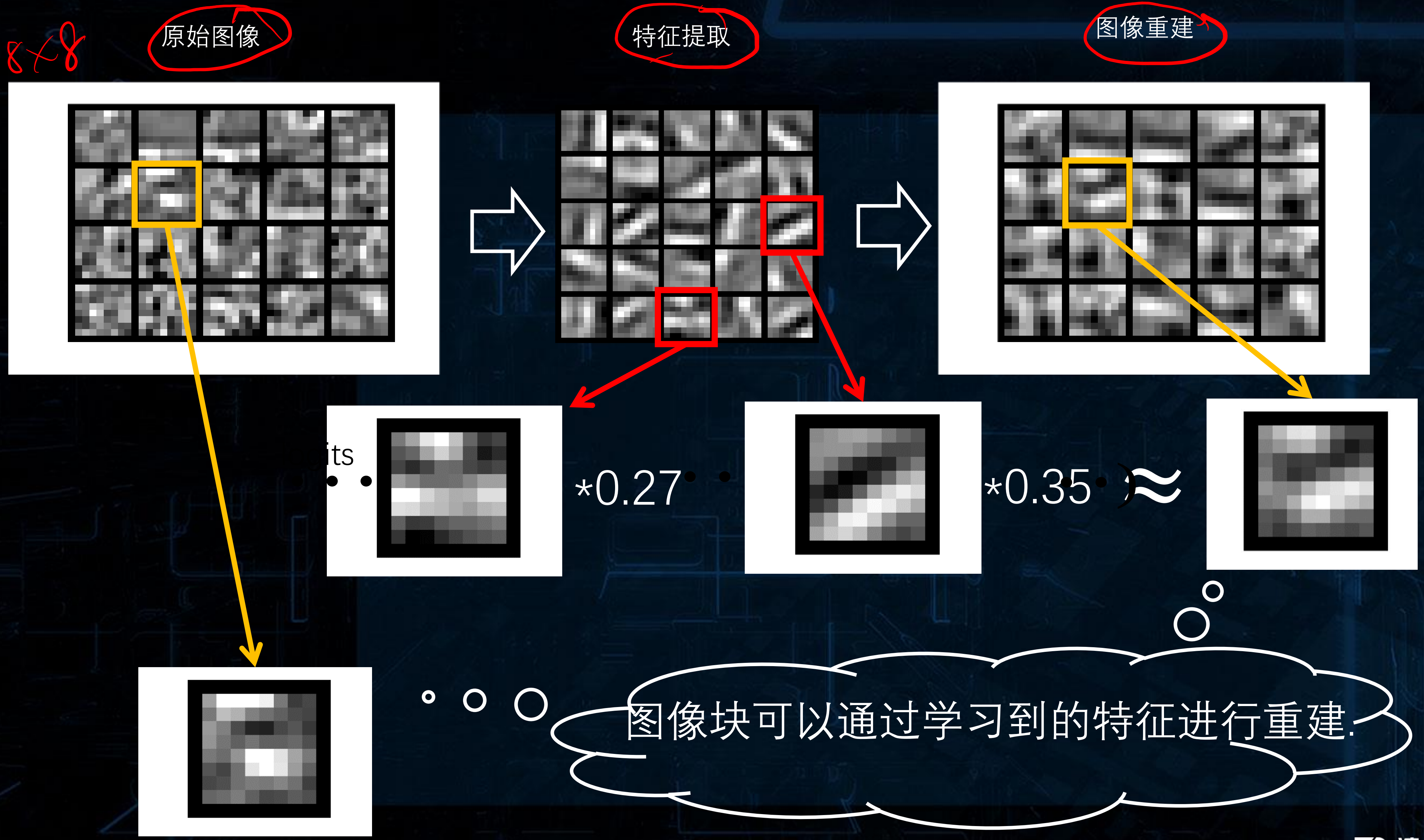
supervised learning

自编码器

0,0,0,0,0,1,0,1,0,0,0,0,0,1,0,1

1,0,1,0 ,　1,0,0,1




8×8
原始图像
特征提取
图像重建
its
*0.27
*0.35
≈
图像块可以通过学习到的特征进行重建.