



北京航空航天大学  
COLLEGE OF SOFTWARE 软件学院  
BEIHANG UNIVERSITY

# 人工智能

## 第10讲：生成式人工智能

### 自编码器、变分自编码器

张晶

2025年春季

- 参考资料：吴飞，《人工智能导论：模型与算法》，高等教育出版社
- 在线课程：<https://www.icourse163.org/course/ZJU-1003377027?from=searchPage>
- 本部分参考：李宏毅，《机器学习》课程，台湾大学



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 提纲

**一、生成式人工智能概念**

**二、从主成分分析到自编码器**

**三、从自编码器到生成模型**

**四、变分自编码器**



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 提纲

## 一、生成式人工智能概念

## 二、从主成分分析到自编码器

## 三、从自编码器到生成模型

## 四、变分自编码器



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 人工智能 (Artificial Intelligence, AI)

人工智能（目标）

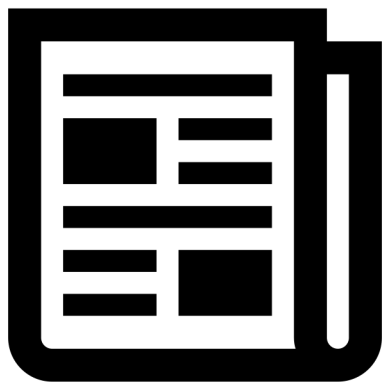
让机器展现“智能”



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 生成式人工智能（Generative AI）

- 让机器产生复杂有结构的对象



文章

由文字所构成



图像

由像素所组成



语音

由采样点构成



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 生成式人工智能

人工智能（目标）

生成式人工智能（目标之一）

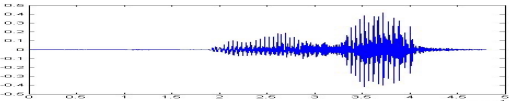




北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 机器学习: 从数据中学习知识


- 语音识别

$$f(\text{  }) = \text{“你好”}$$

- 图像分类

$$f(\text{  }) = \text{“猫”}$$

- 围棋游戏

$$f(\text{  }) = \text{“5-5”}$$

(下一步落子位置)



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 生成式人工智能

人工智能（目标）

生成式人工智能（目标之一）

机器学习（手段）

深度学习（更厉害的手段）





北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 生成式人工智能

人工智能（目标）

机器学习（手段）

深度学习（更厉害的手段）

生成式人工智能

目前的生成式人工智能多以深度学习为手段

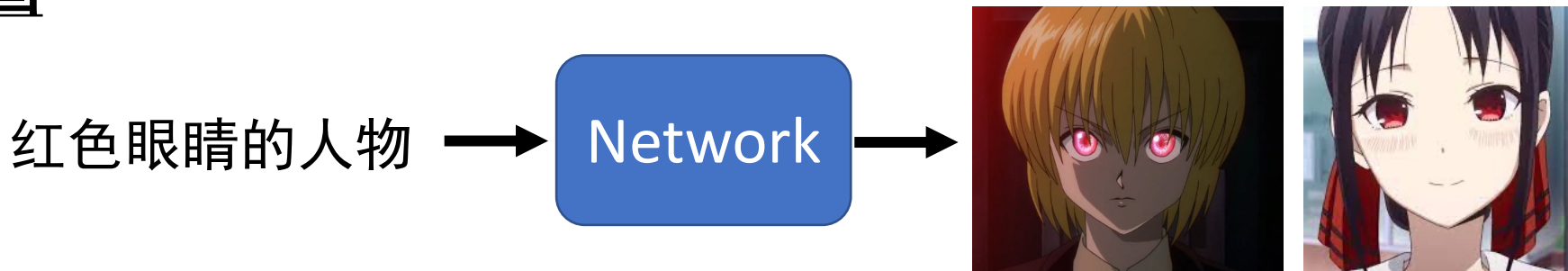


# 内容生成：AI-Generated Content (AIGC)

(同一输入可以产生不同输出。)

- AIGC是一种对“创意”有需求的任务

## 绘画

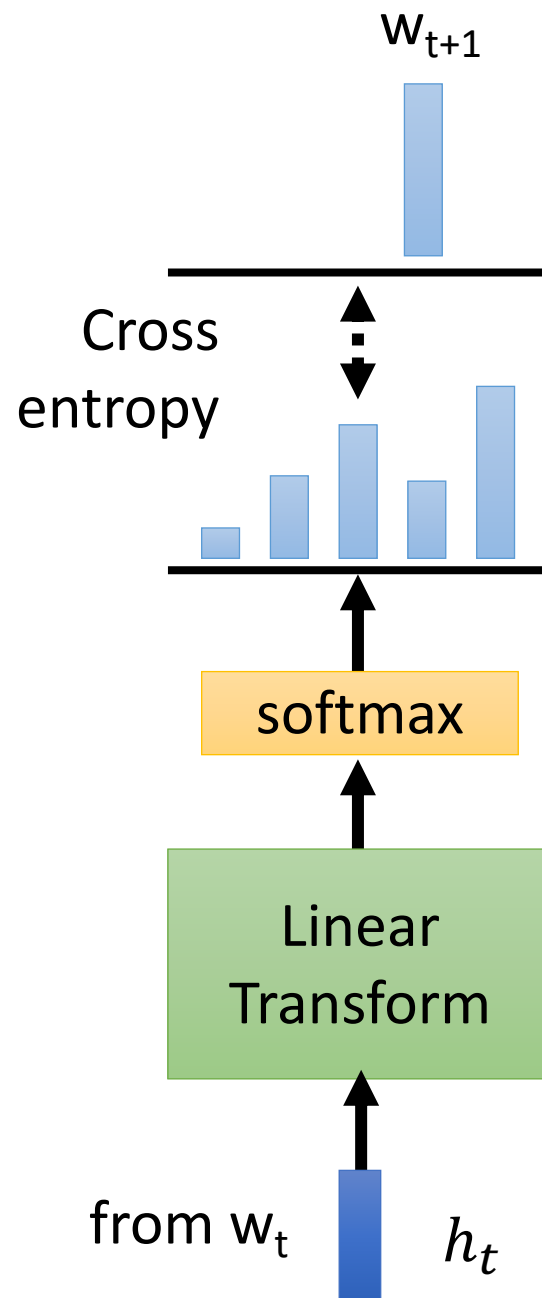
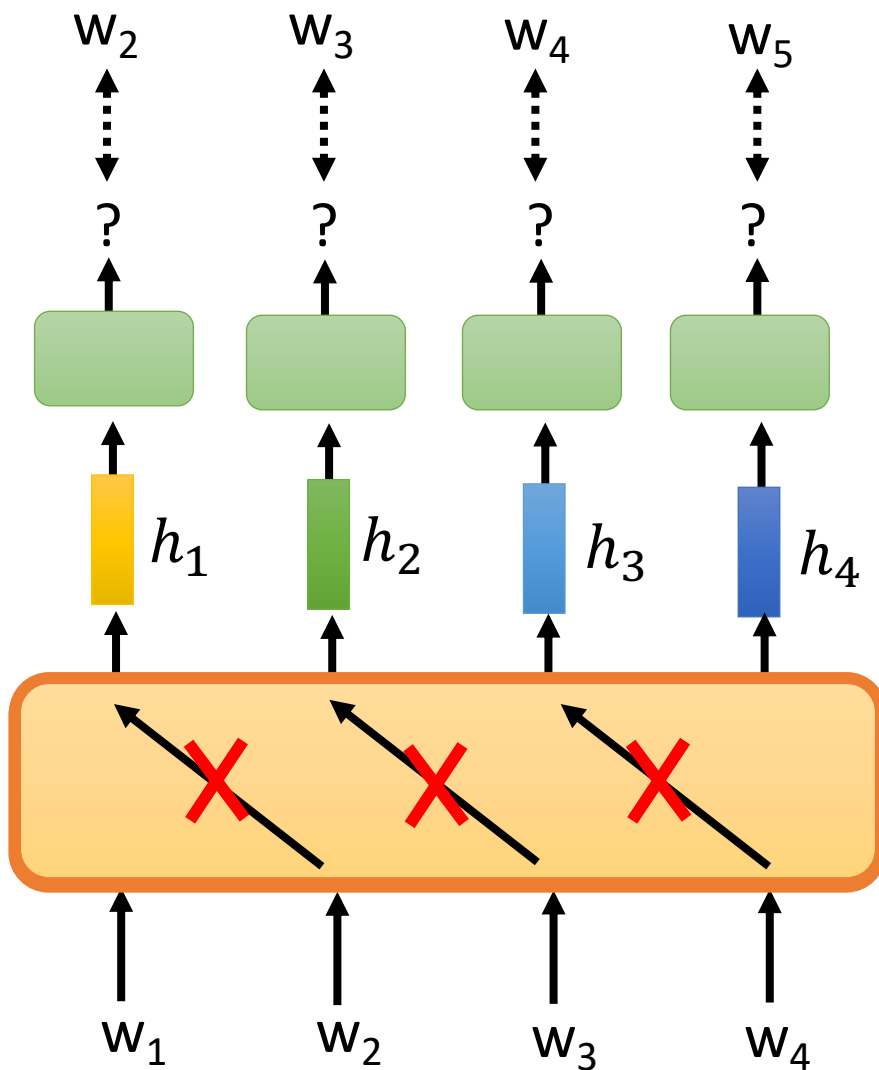


## 聊天机器人





# 生成模型举例：GPT





北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 生成策略

复杂的输入

文章

由文字构成



较小的单位

字

字

字

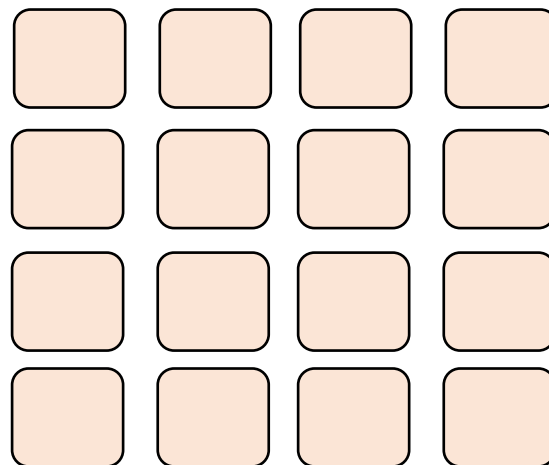
字

字

依照某种固定的顺序依序生成

图片

由像素构成



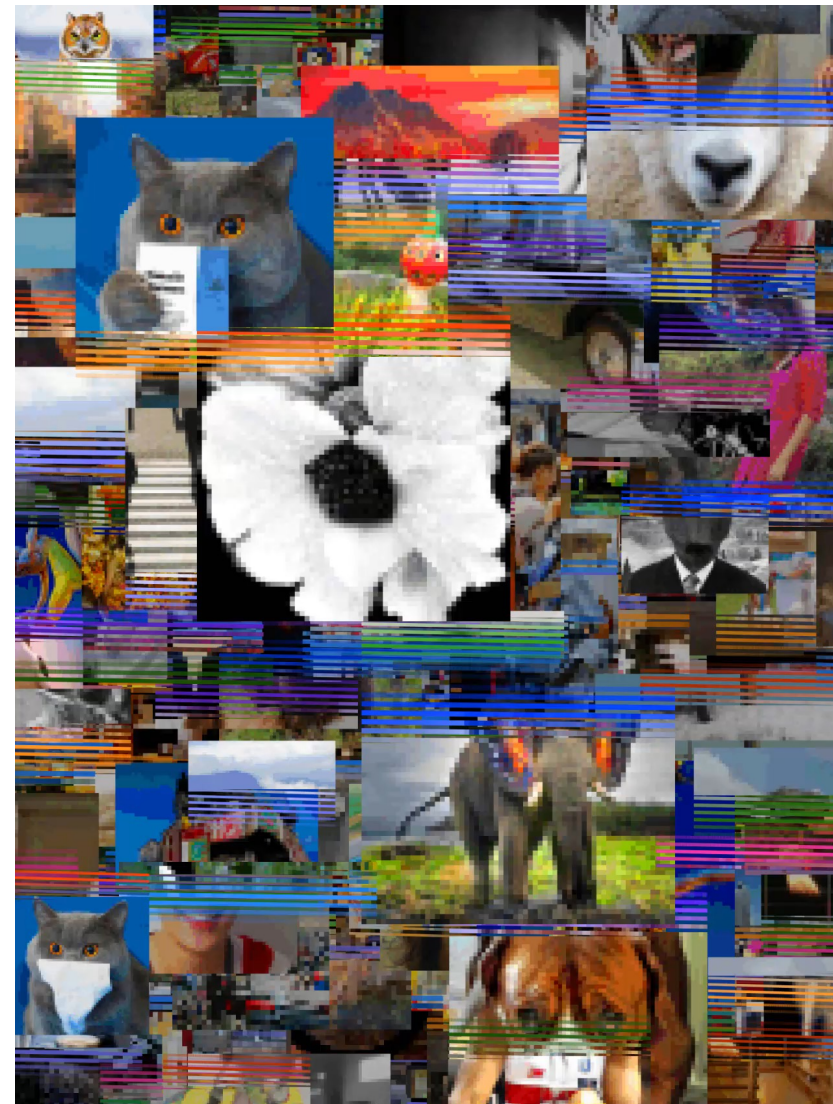
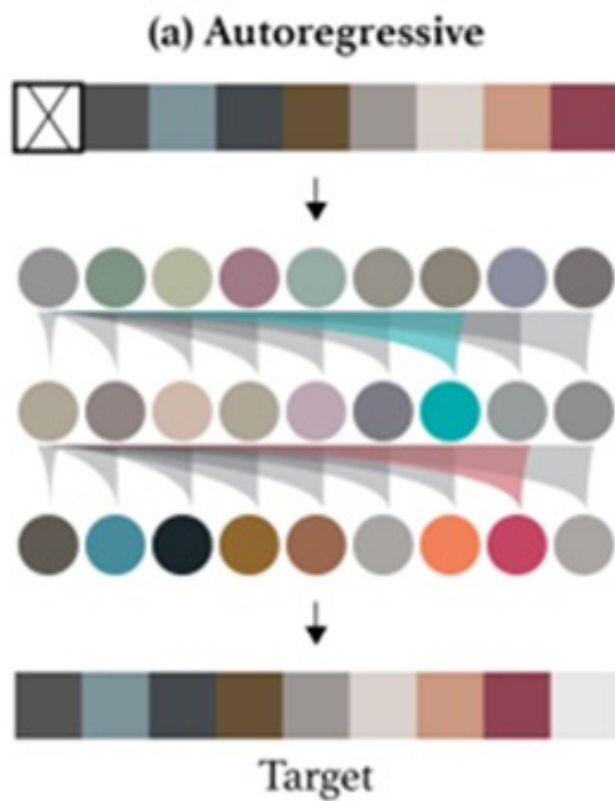
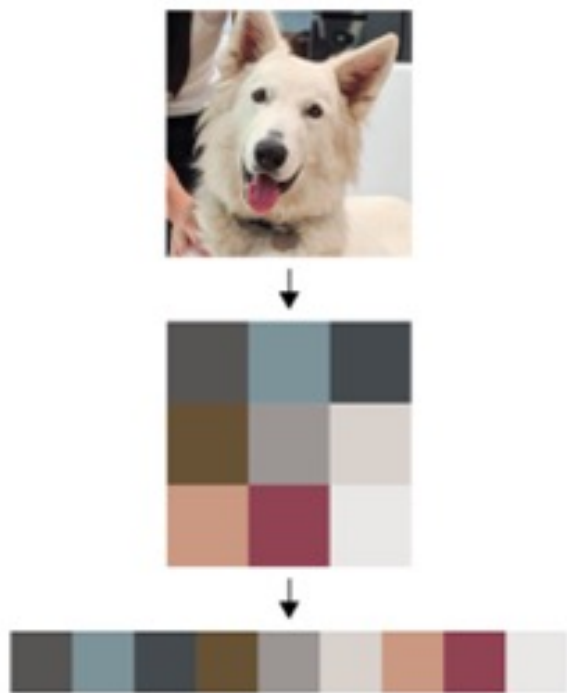
Autoregressive Generation



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 生成策略

## 图像版 GPT



<https://openai.com/blog/image-gpt/>



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 提纲

**一、生成式人工智能概念**

**二、从主成分分析到自编码器**

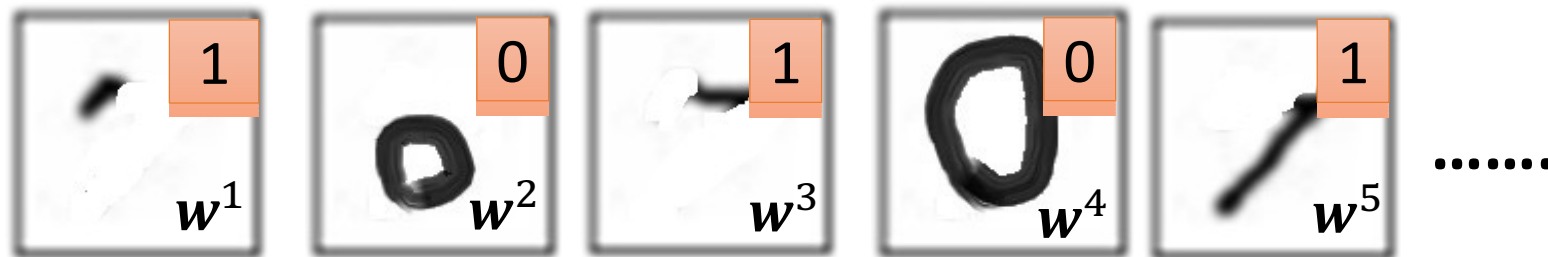
**三、从自编码器到生成模型**

**四、变分自编码器**

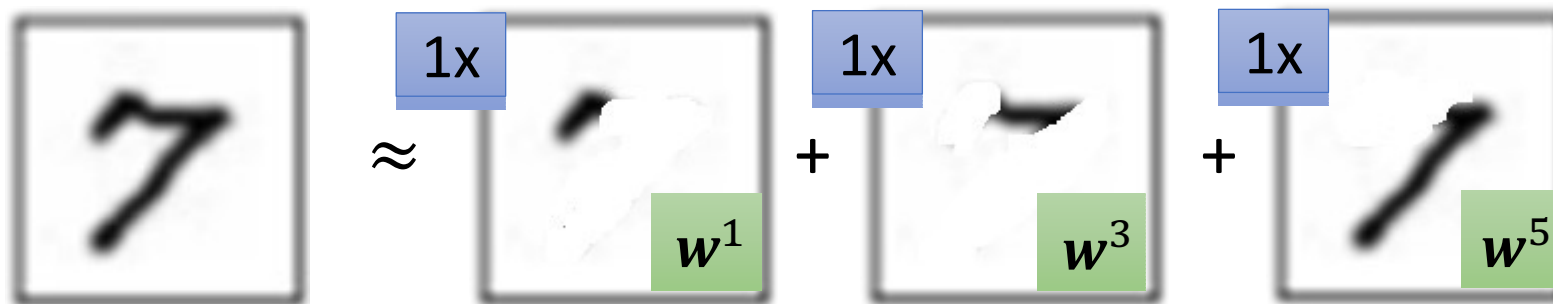


# 回顾：主成分分析

主成分:



$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ \vdots \end{bmatrix}$



$$x \approx c_1 w^1 + c_2 w^2 + \dots + c_K w^K + \bar{x}$$

手写数字  
图像的像  
素值表达


主成分

$\begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_K \end{bmatrix}$

手写数字  
图像新特  
征表达



# 主成分分析

$$\mathbf{x} - \bar{\mathbf{x}} \approx c_1 \mathbf{w}^1 + c_2 \mathbf{w}^2 + \cdots + c_K \mathbf{w}^K = \hat{\mathbf{x}}$$


重构误差:  $\|(\mathbf{x} - \bar{\mathbf{x}}) - \hat{\mathbf{x}}\|_2$

优化问题: 通过最小化 $L$ 得到 $\mathbf{w}$ ,

$$L = \min_{\{\mathbf{w}^1, \dots, \mathbf{w}^K\}} \sum \left\| (\mathbf{x} - \bar{\mathbf{x}}) - \left( \sum_{k=1}^K c_k \mathbf{w}^k \right) \right\|_2$$





# 主成分分析

$$L = \min_{\{w^1, \dots, w^K\}} \sum \left\| (x - \bar{x}) - \left( \sum_{k=1}^K c_k w^k \right) \right\|_2$$

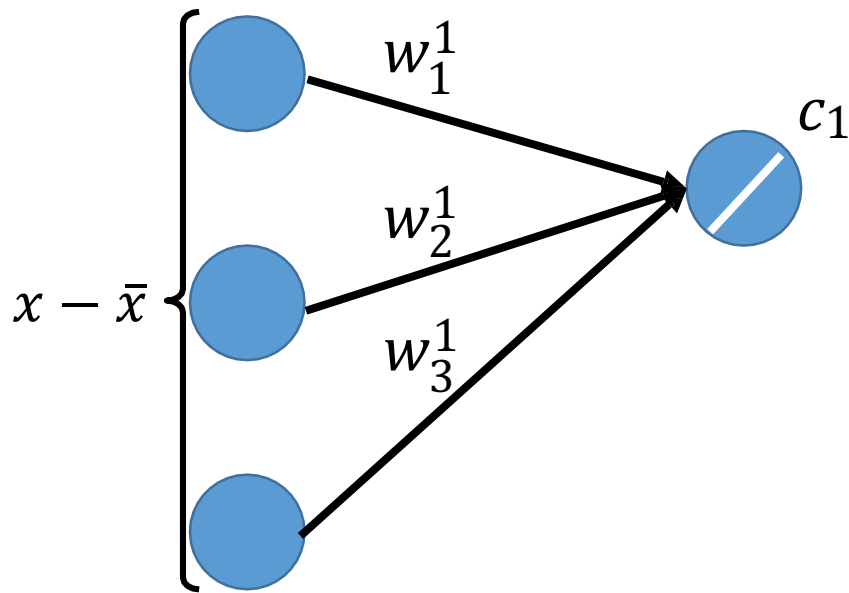
如果  $\{w^1, w^2, \dots, w^K\}$  即为主成分  $\{u^1, u^2, \dots, u^K\}$

$$\hat{x} = \sum_{k=1}^K c_k w^k \longleftrightarrow x - \bar{x}$$

最小化重构误差:

$$c_k = (x - \bar{x}) \cdot w^k$$

$K = 2$ :





# 主成分分析

$$L = \min_{\{w^1, \dots, w^K\}} \sum \left\| (x - \bar{x}) - \left( \sum_{k=1}^K c_k w^k \right) \right\|_2$$

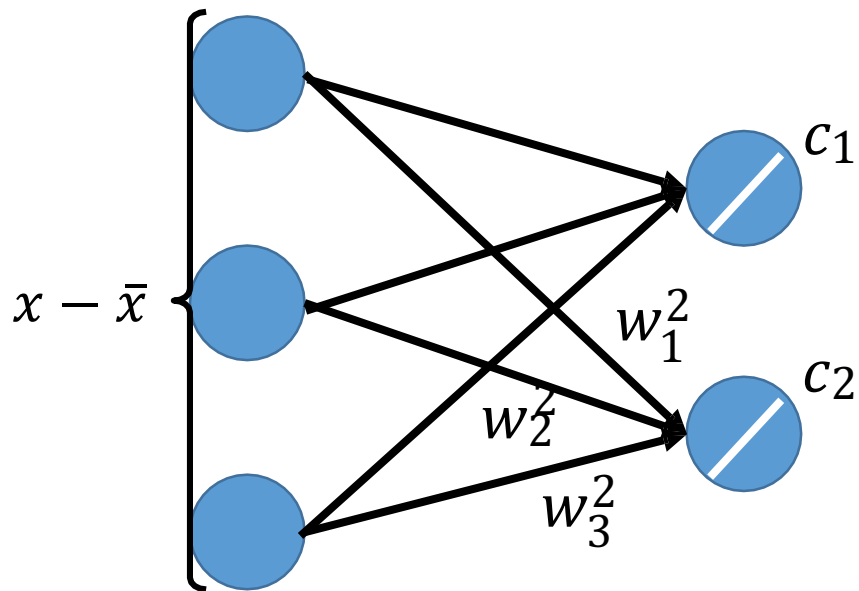
如果  $\{w^1, w^2, \dots, w^K\}$  即为主成分  $\{u^1, u^2, \dots, u^K\}$

$$\hat{x} = \sum_{k=1}^K c_k w^k \longleftrightarrow x - \bar{x}$$

最小化重构误差:

$$c_k = (x - \bar{x}) \cdot w^k$$

$K = 2$ :





# 主成分分析

$$L = \min_{\{w^1, \dots, w^K\}} \sum \left\| (x - \bar{x}) - \left( \sum_{k=1}^K c_k w^k \right) \right\|_2$$

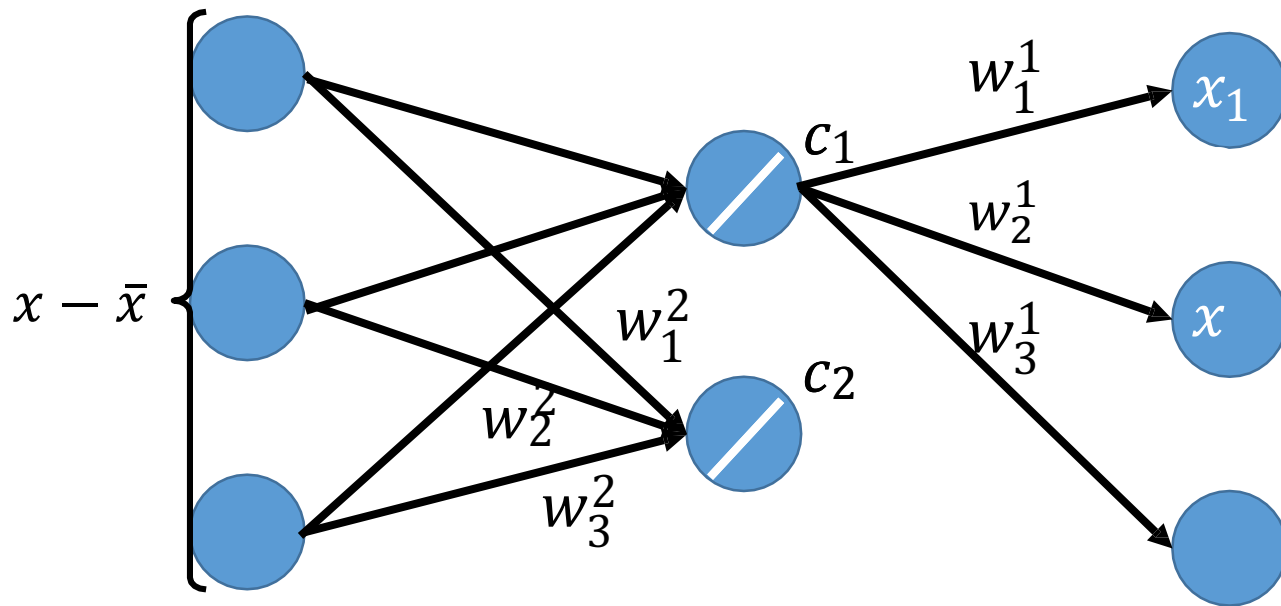
如果  $\{w^1, w^2, \dots, w^K\}$  即为主成分  $\{u^1, u^2, \dots, u^K\}$

$$\hat{x} = \sum_{k=1}^K c_k w^k \longleftrightarrow x - \bar{x}$$

最小化重构误差:

$$c_k = (x - \bar{x}) \cdot w^k$$

$K = 2$ :





# PCA 可以看做单隐层的神经网络 (线性激活)

自编码器

如果  $\{w^1, w^2, \dots, w^K\}$  即为主成分  $\{u^1, u^2, \dots, u^K\}$

$$\hat{x} = \sum_{k=1}^K c_k w^k \longleftrightarrow x - \bar{x}$$

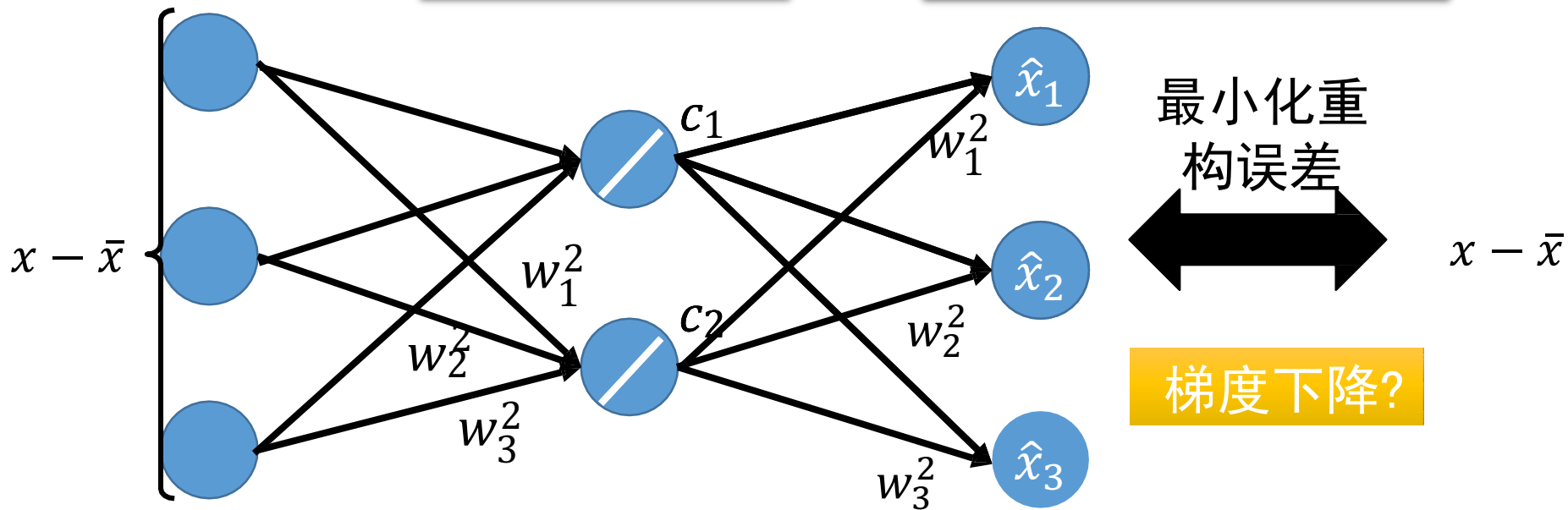
最小化重构误差:

$$c_k = (x - \bar{x}) \cdot w^k$$

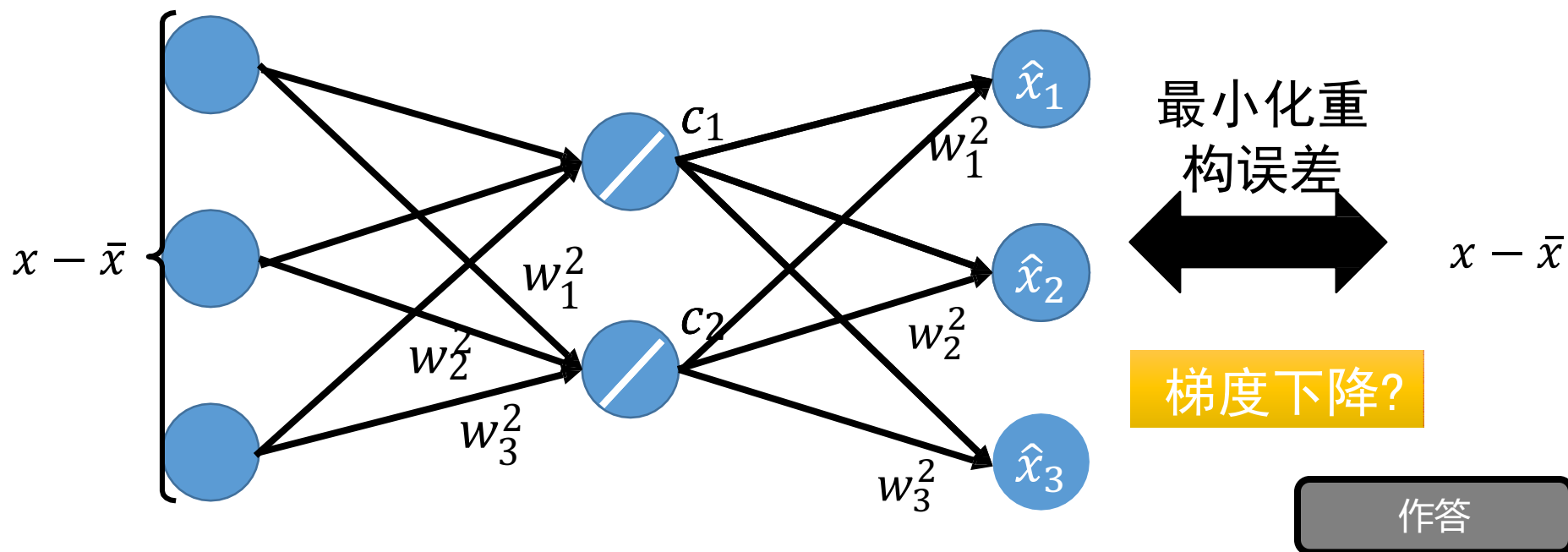
$K = 2$ :

加深层数

深度自编码器



请思考，单隐藏层线性自编码器经过梯度下降得到的结果跟主成分分析PCA结果一样吗？为什么？

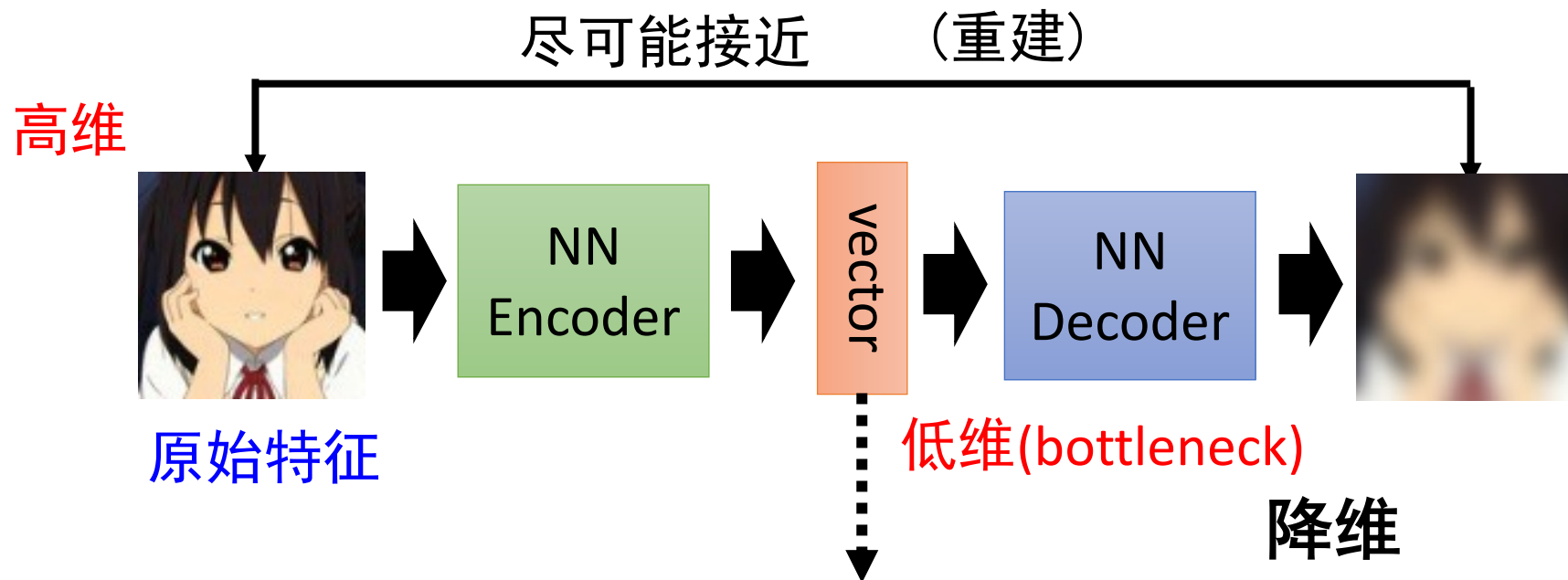
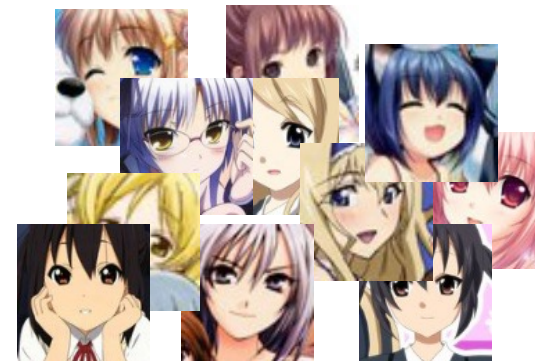




北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 自编码器 Auto-encoder

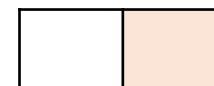
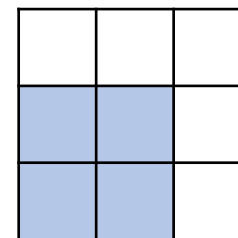
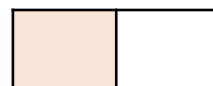
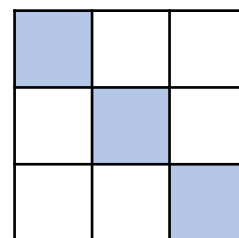
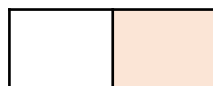
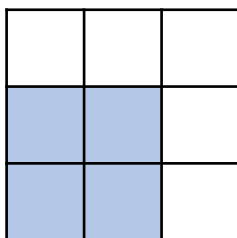
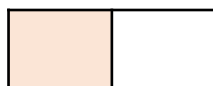
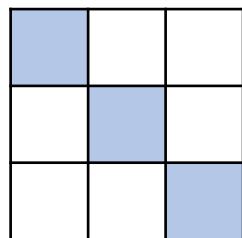
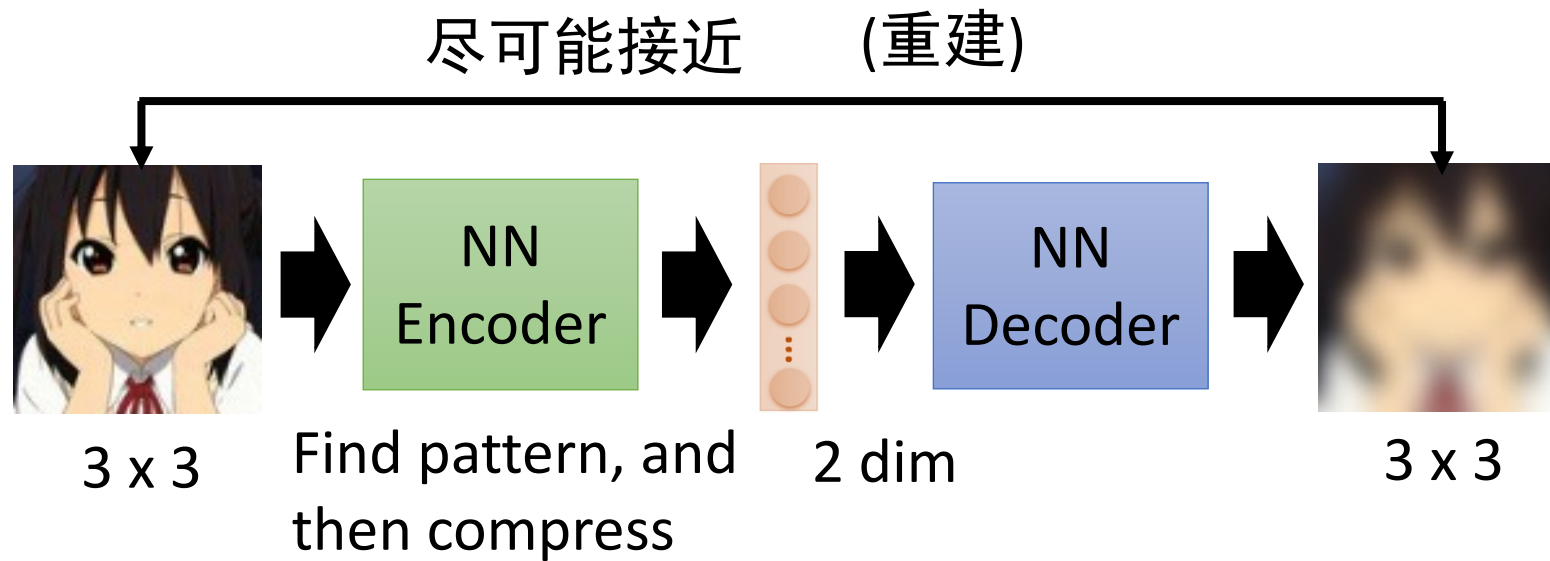
无标注图像



Embedding, Representation, Code  
为下游任务学习新的特征表示

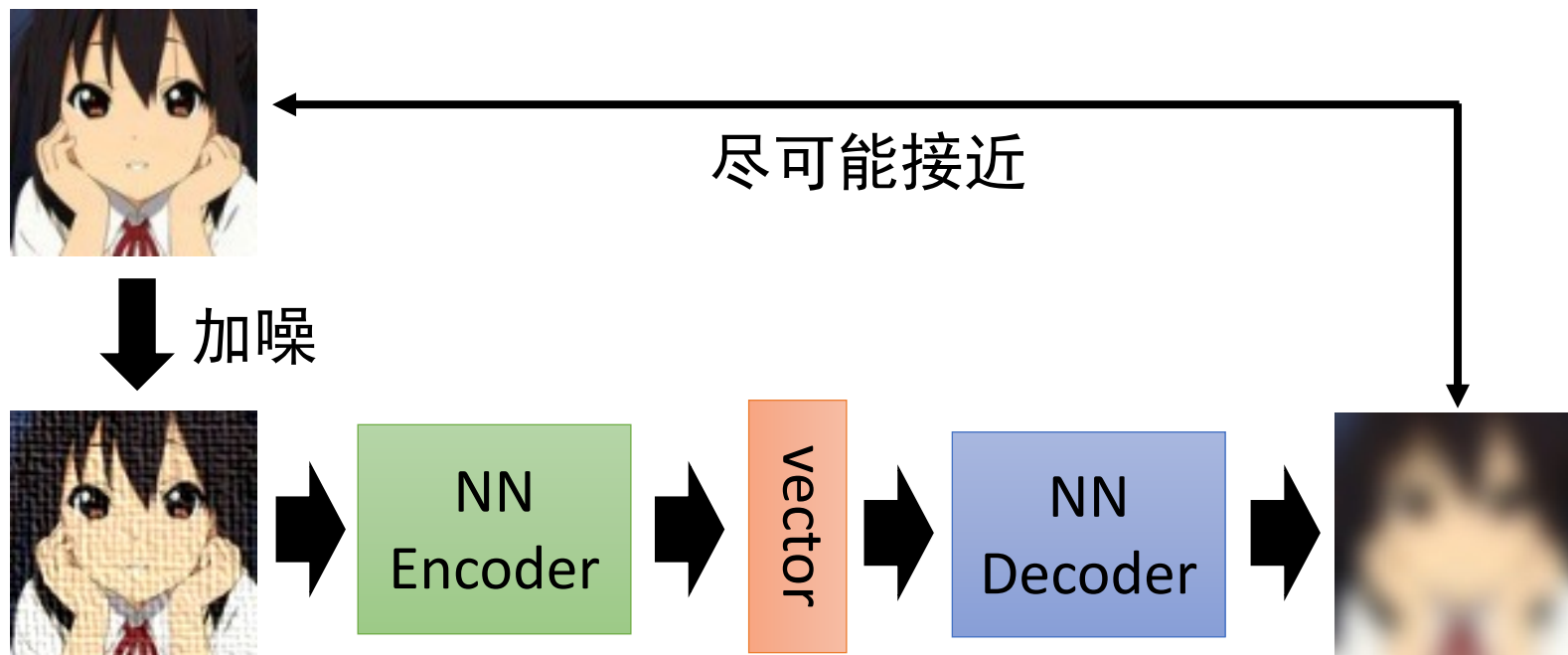


# Why Auto-encoder?





# 去噪自编码器 De-noising Auto-encoder



The idea sounds familiar? 😊

Vincent, Pascal, et al. "Extracting and composing robust features with denoising autoencoders." *ICML*, 2008.

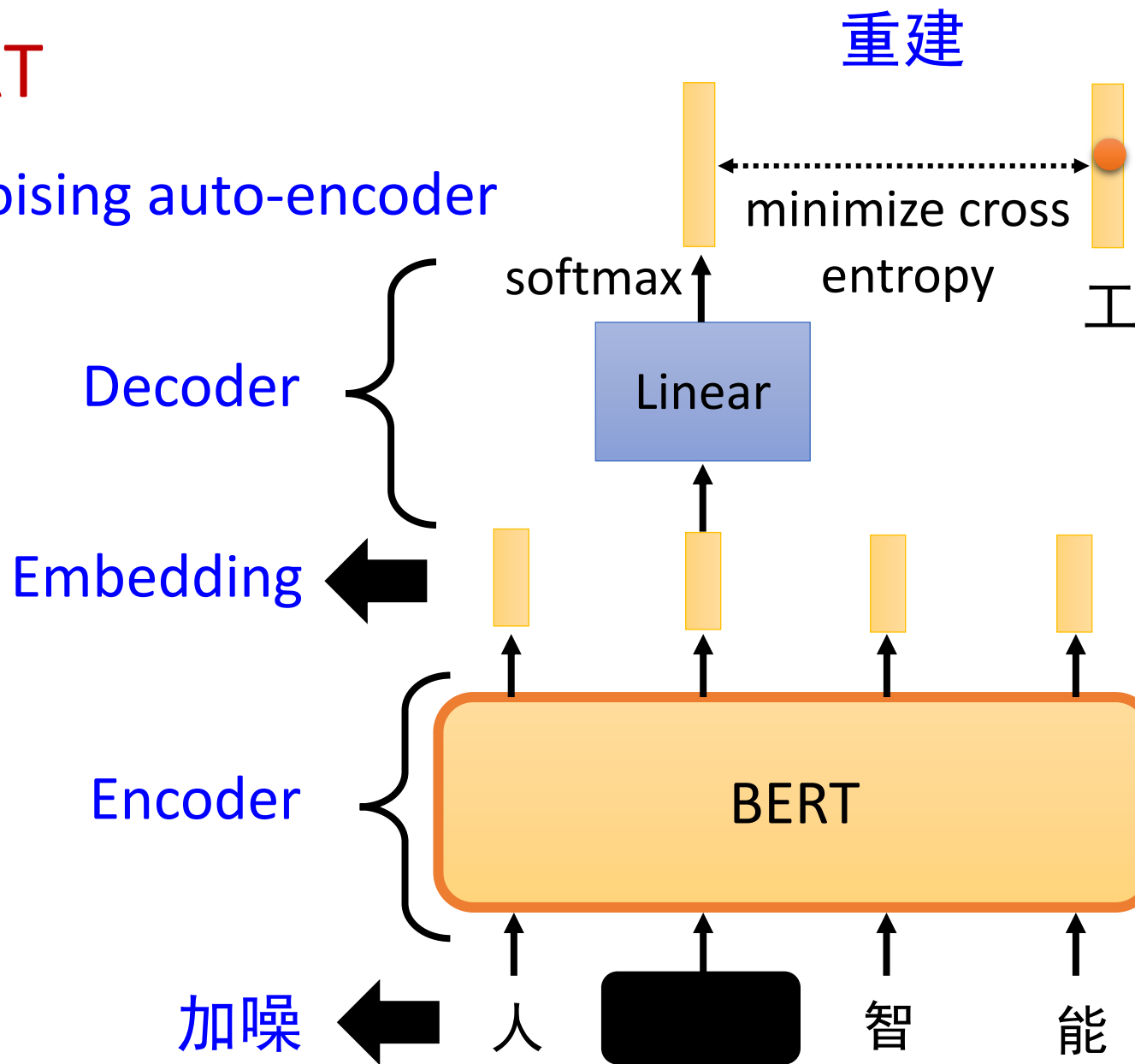


请回顾，我们在深度学习中介绍的哪个模型类似于去噪自编码器？

作答

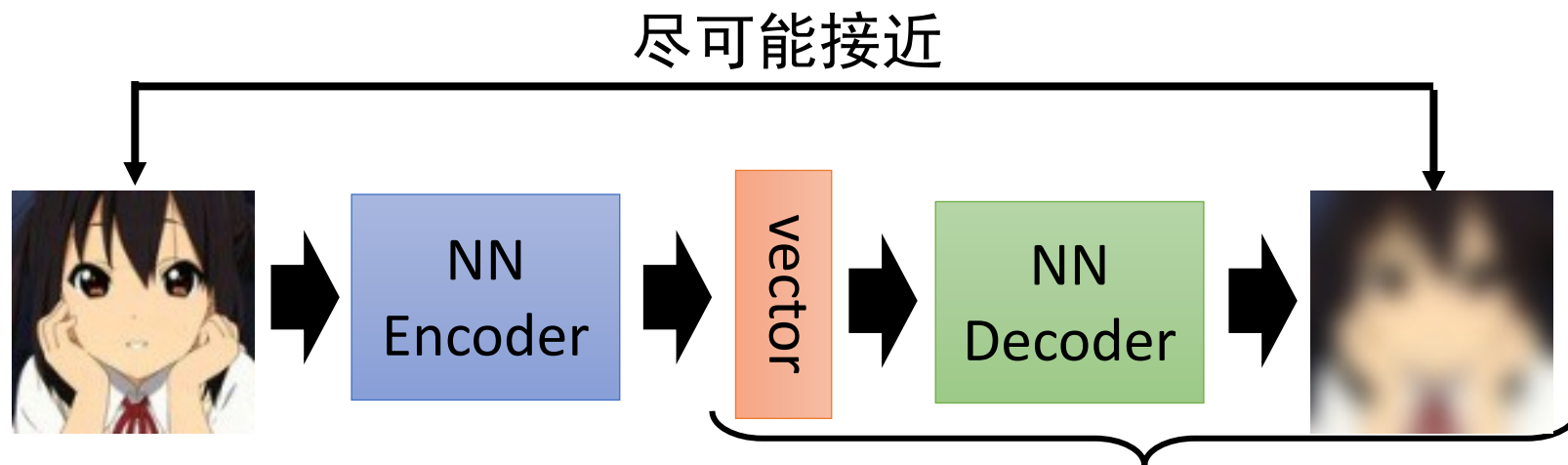
# 回顾: BERT

A de-noising auto-encoder



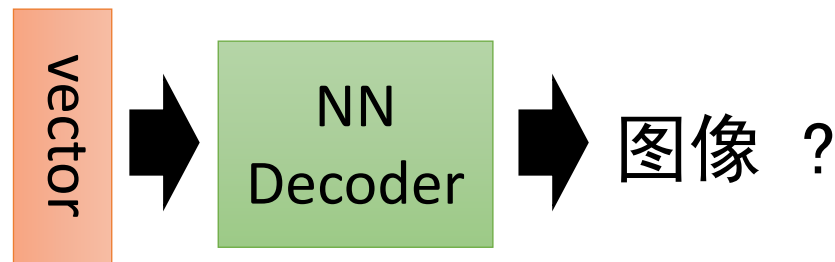


# 图像生成：自编码器 Auto-encoder



看做生成器?

随机从某个分布采样向量



请思考，直接用自编码器的解码器作为生成器，会有什么问题？

作答



北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 提纲

**一、生成式人工智能概念**

**二、从主成分分析到自编码器**

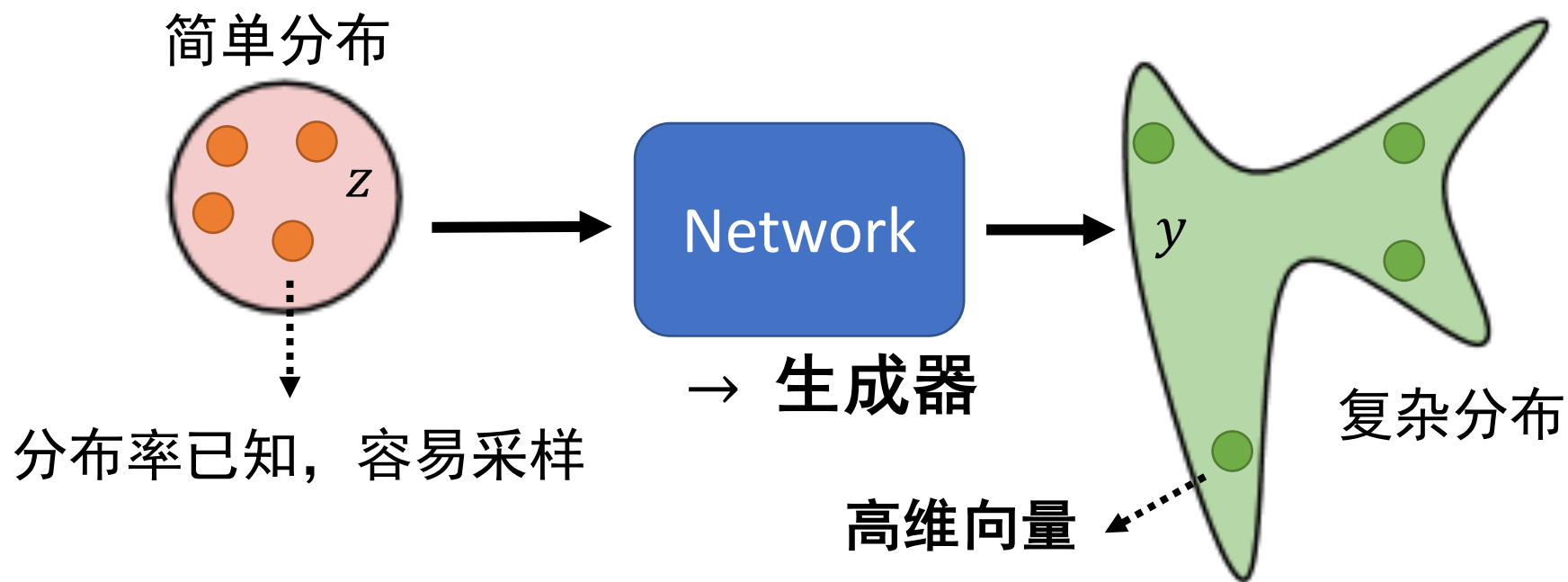
**三、从自编码器到生成模型**

**四、变分自编码器**



# 图像生成：Network as Generator

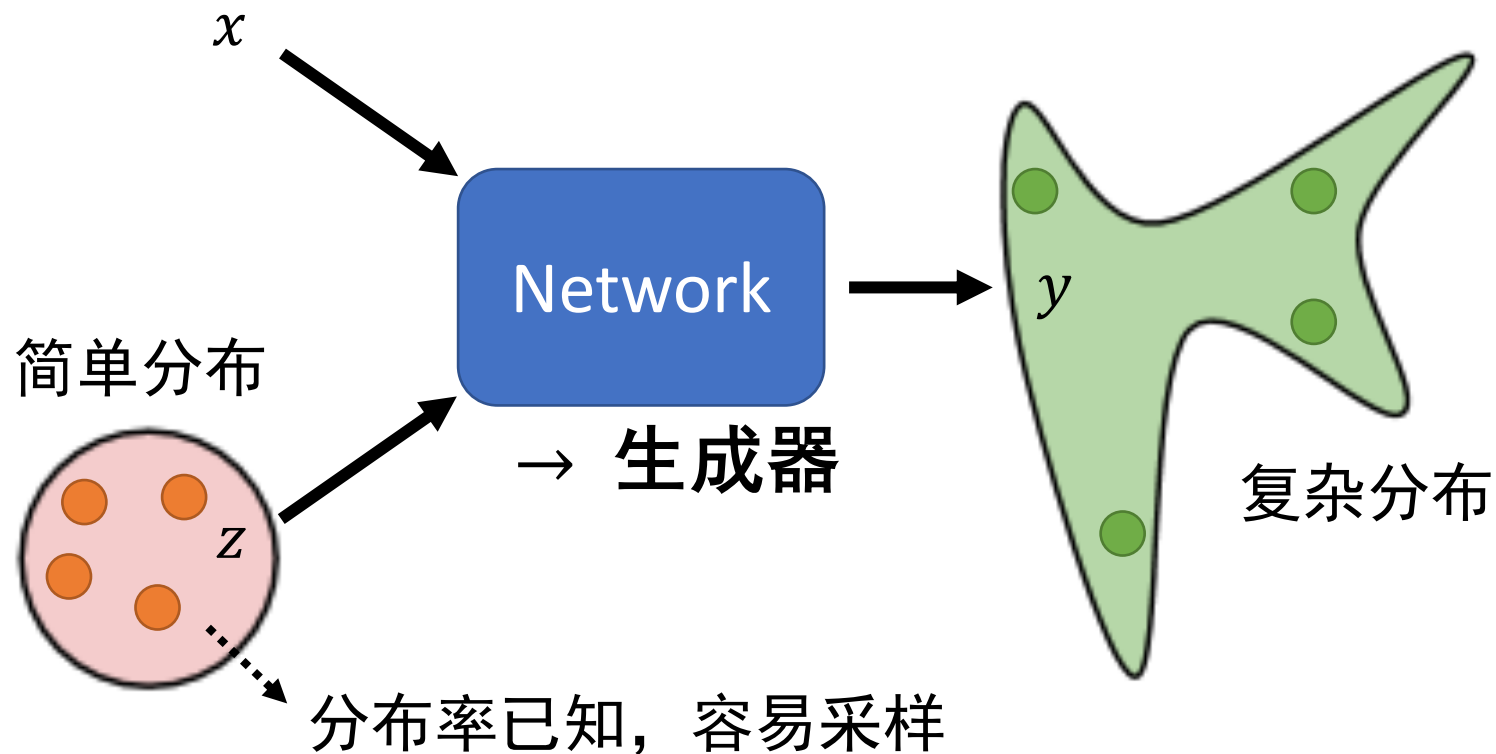
- 无条件生成





# 图像生成：Network as Generator

- 带条件生成

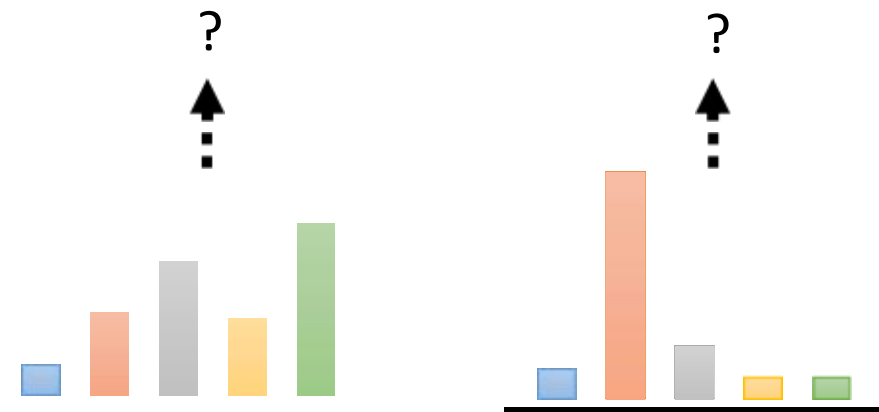




北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 图像生成: Why distribution?

一只在奔跑的狗



256 x 256 images  $\rightarrow$  65536 pixels

每一个像素独立绘制







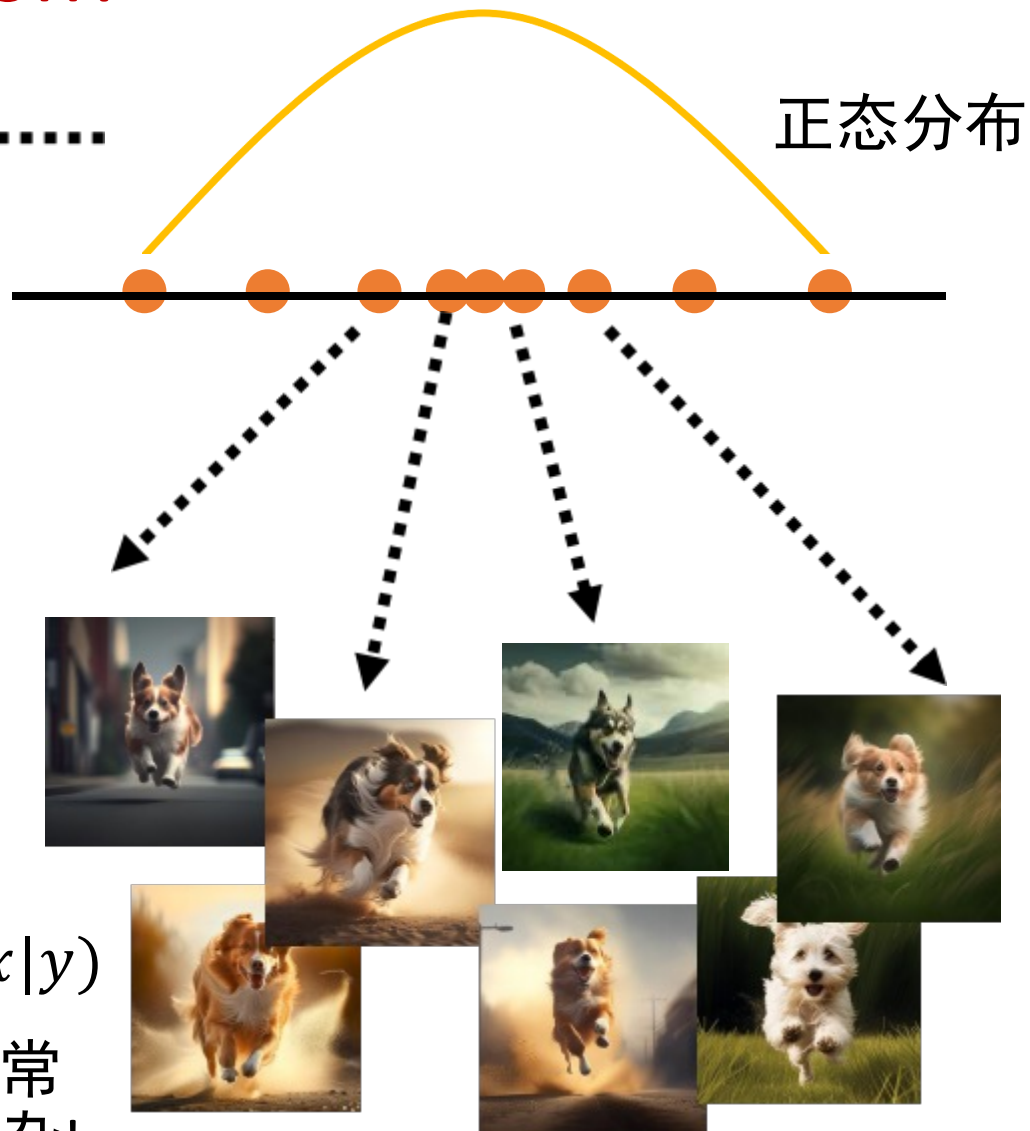
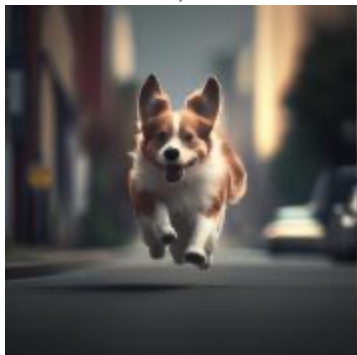
北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 图像生成: Why distribution?

—在奔跑的狗

图像  
生成

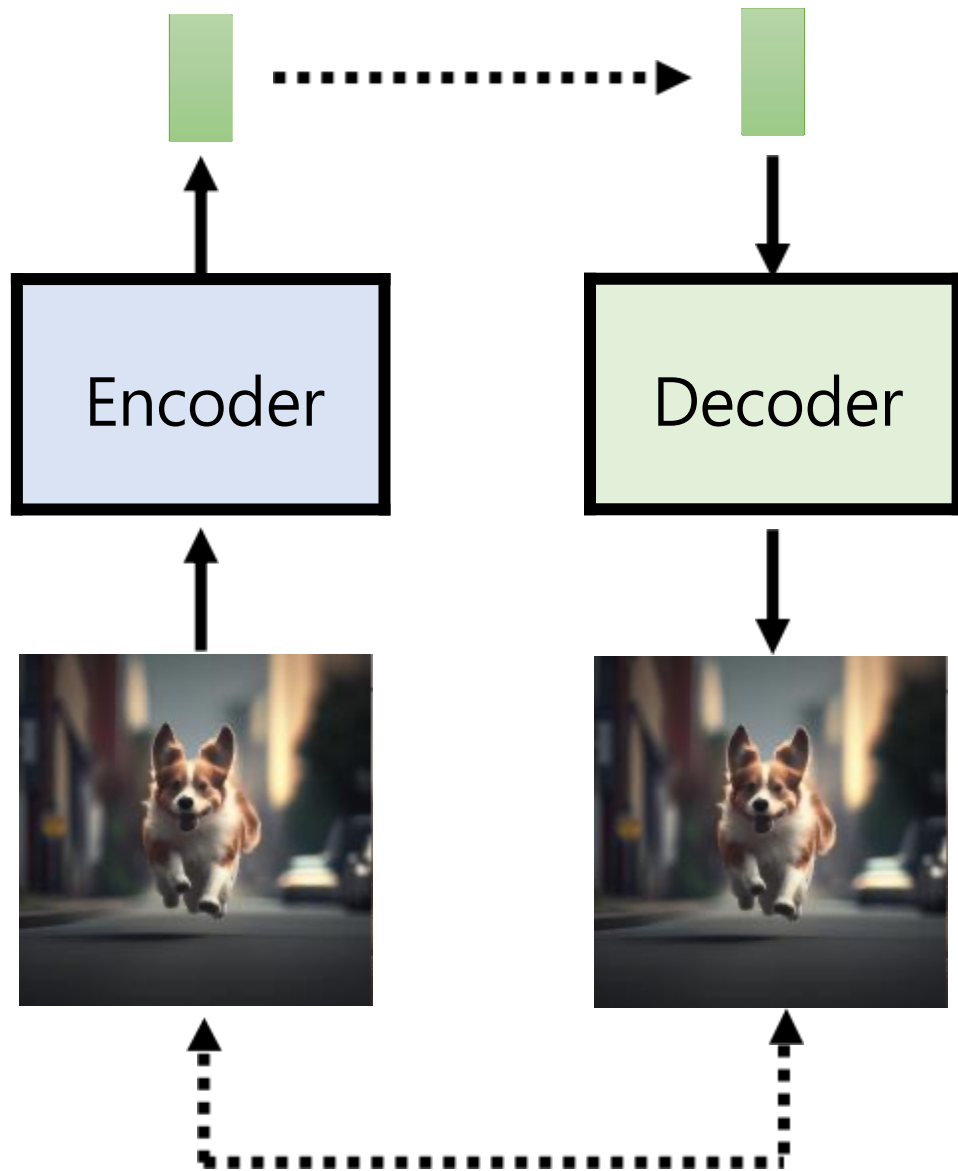
$y$ : 文字  
 $x$ : 图像





北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 基本模型：自编码器

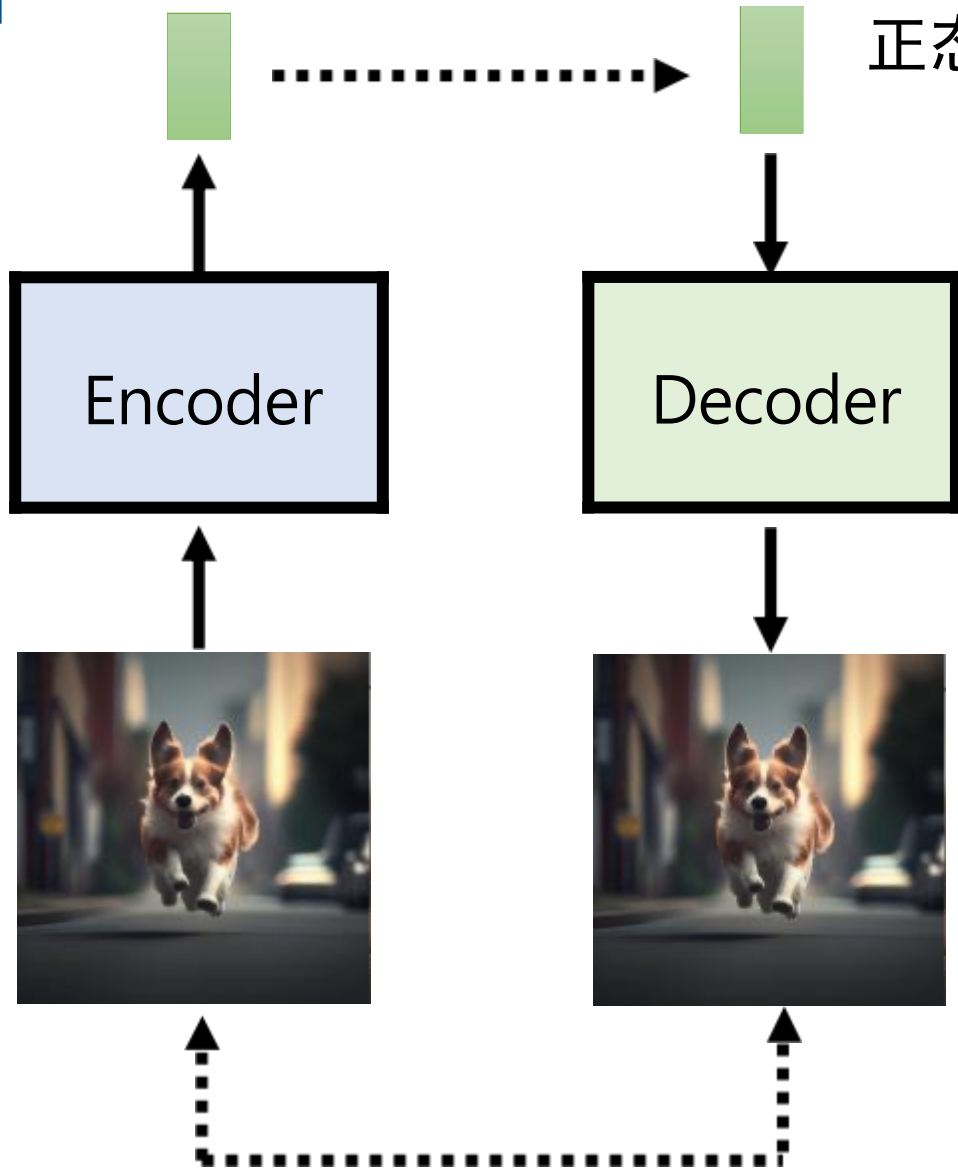


离散的特征向量



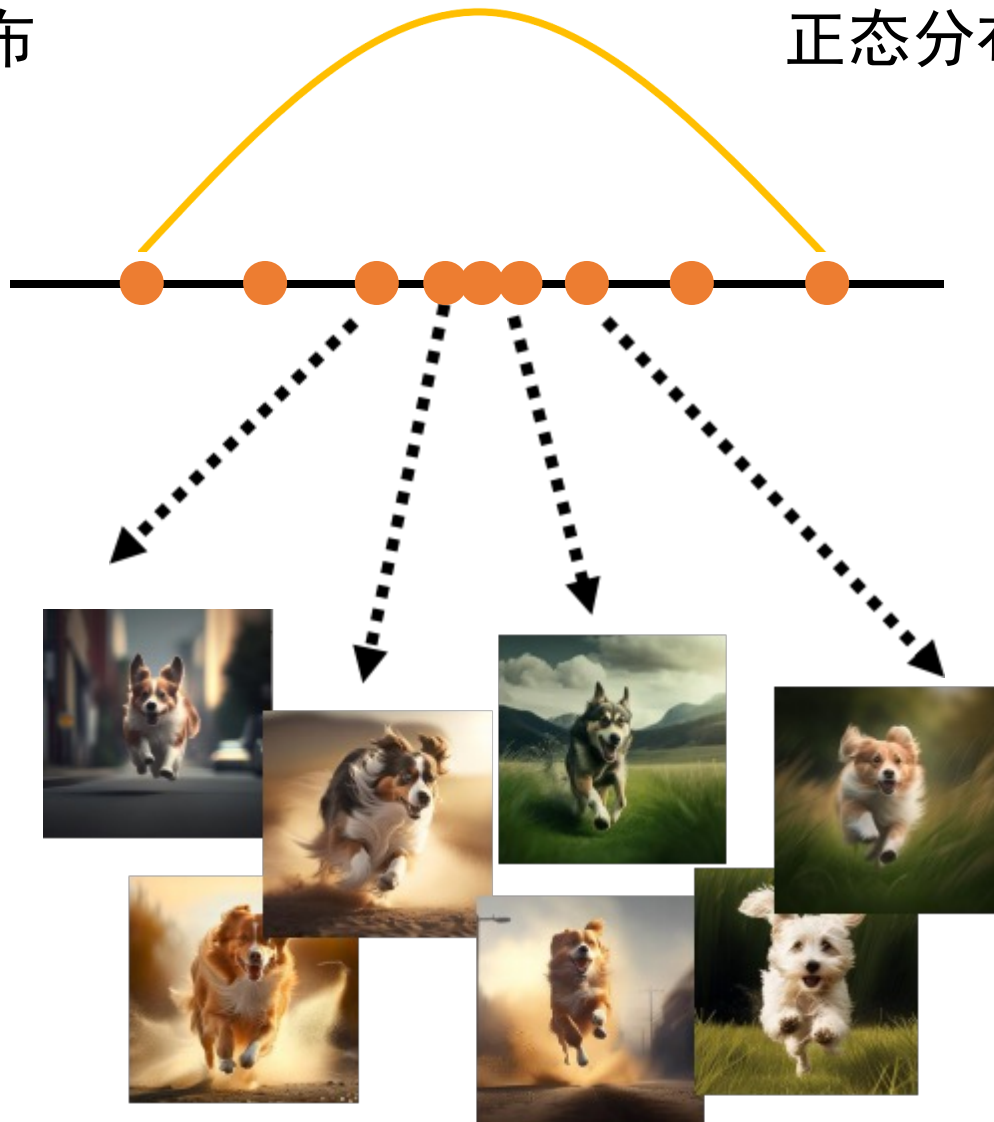
北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 图像生成模型：变分自编码器 VAE



正态分布

正态分布

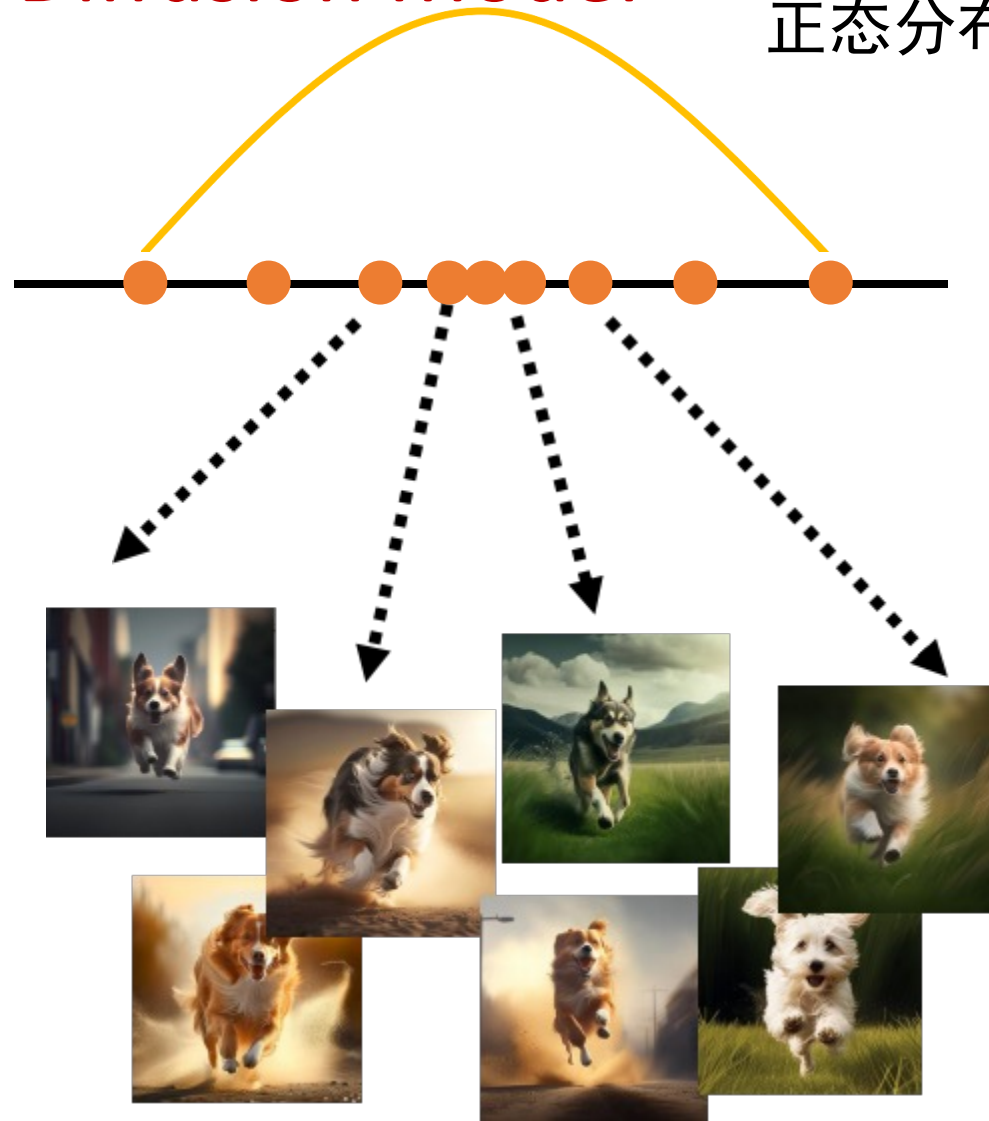
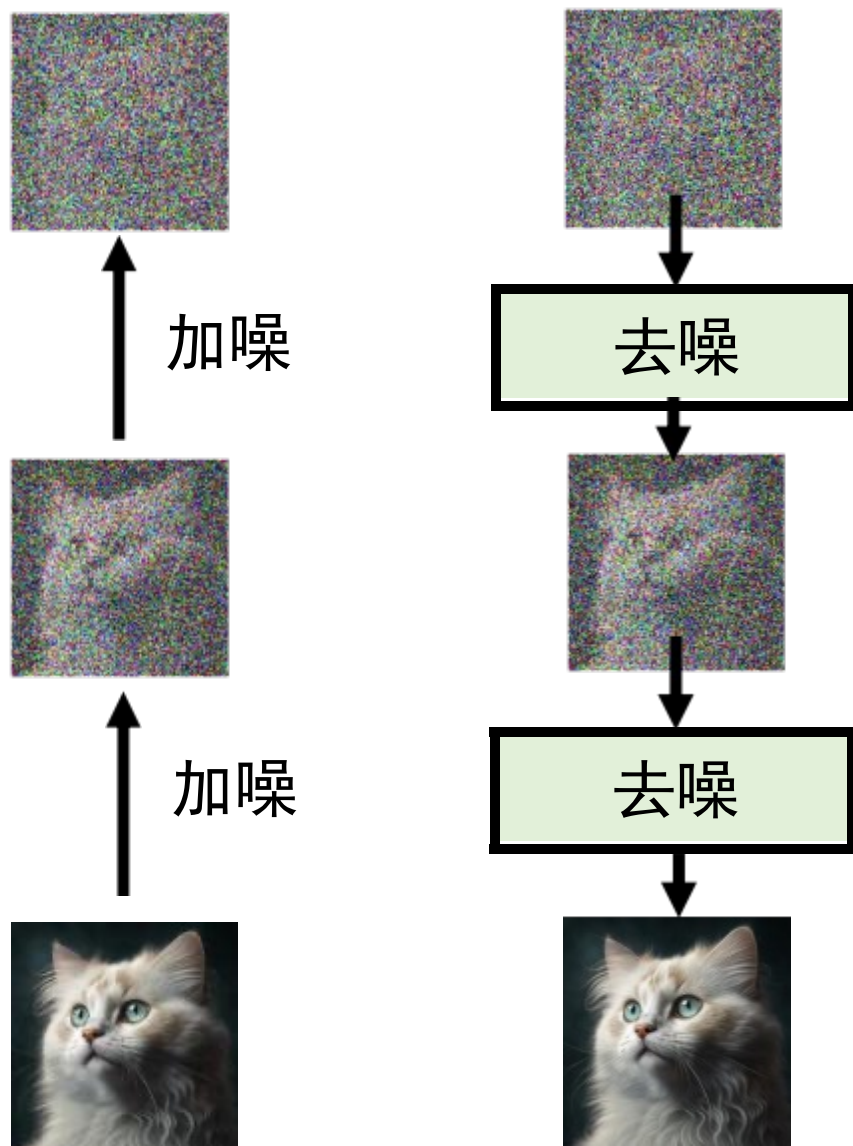




北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 图像生成模型：扩散模型 Diffusion Model

正态分布





北京航空航天大学  
COLLEGE OF SOFTWARE  
BEIHANG UNIVERSITY 软件学院

# 图像生成模型：生成对抗网络 GAN

