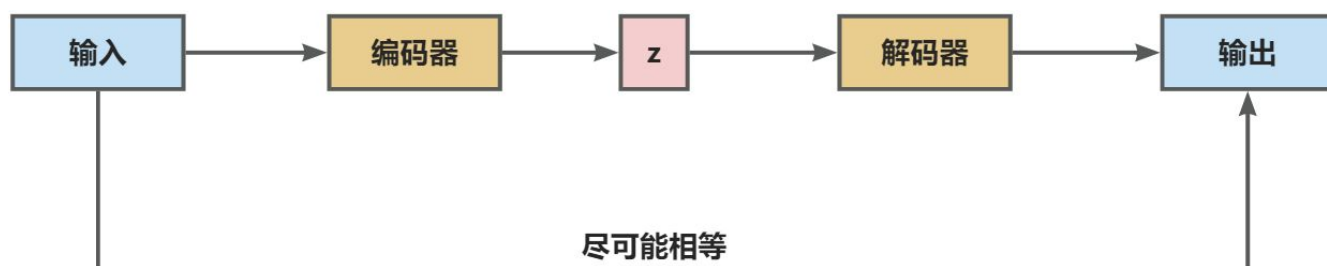


# VAE

## 1. 自动编码器 (AE)

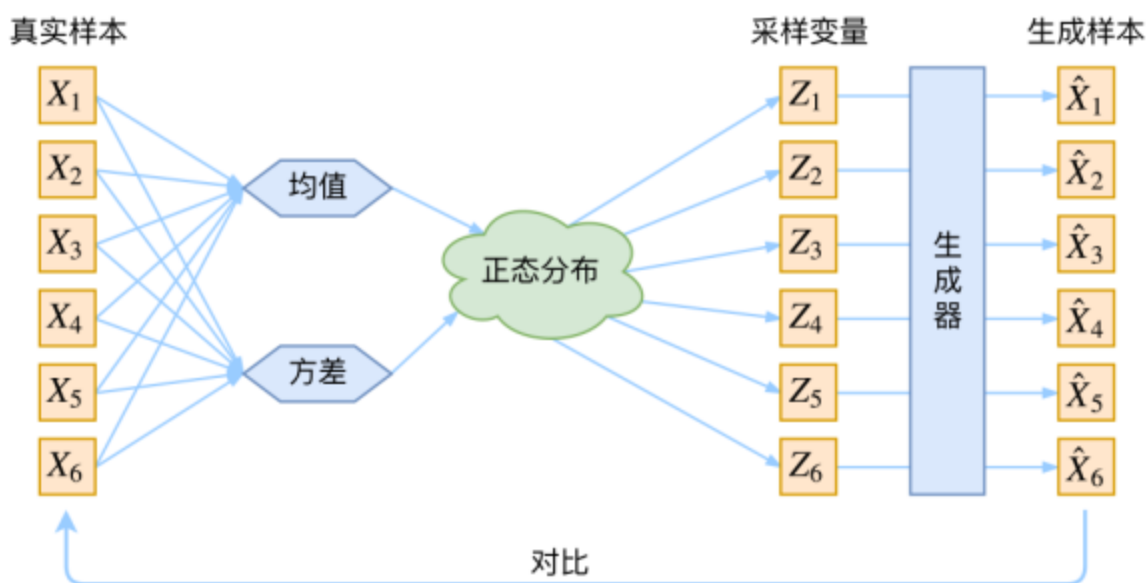
AE (Autoencoder)，自动编码器。其初衷是为了进行数据降维，希望原有的数据可以被降维成一个低维的数据 $z$ ，然后最后又可以被解码回到原始数据，并尽可能保持一致。



我们所常用的encoder-decoder结构即为一种简单的AE。由于AE的训练不需要对数据进行标注，所以其是一种无监督的学习方法。目前自编码器的主要用途就是降维/去噪，图像生成。

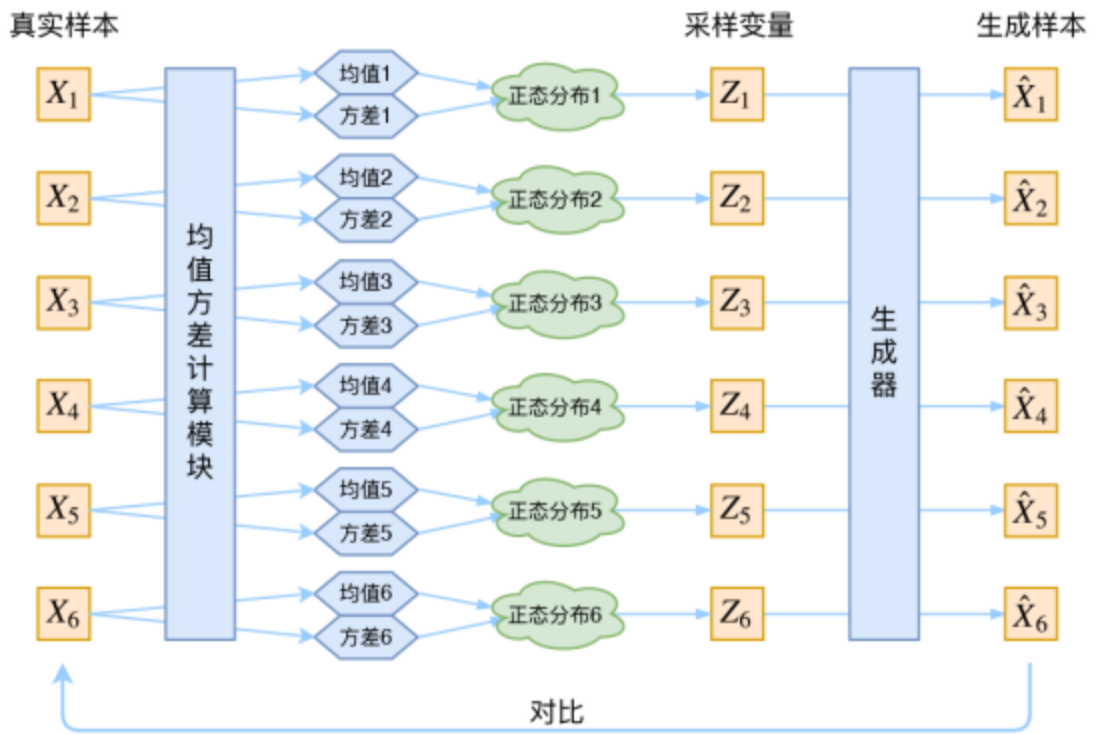
## 2. 变分自编码器 (VAE)

VAE假定了一个服从某些常见分布的隐变量 $Z$ ，然后训练一个模型  $X = g(Z)$ ，将原来的概率分布映射到训练集的概率分布，从而生成类似训练集的数据。结构大致如下图所示：



但本图中存在一个问题，因为生成的图片与训练集中图片之间不存在对应关系，所以无法判断生成数据的分布是否为训练集中的概率分布。

所以，在实际实践中，我们所假设的并非  $p(Z)$  为正态分布，而是  $p(Z|X)$  为正态分布，这样才把每个采样出的  $Z_t$  对应到了  $X_t$  上，因此还原出的  $\hat{X}_t$  可以与其对应的真实样本求loss，从而对生成器进行优化。

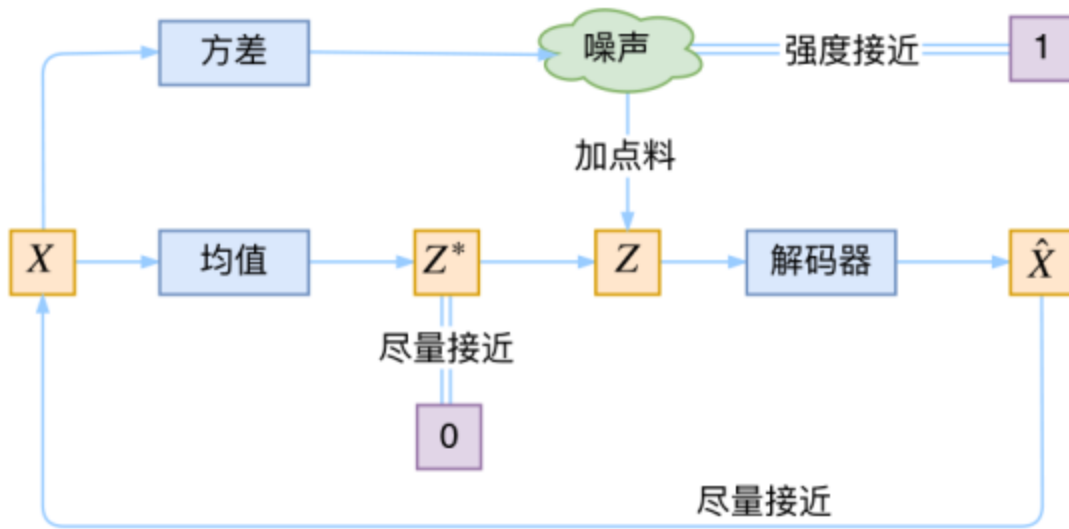


此时，结构图变为了上图所示，因为每张图片所对应的正态分布是不相同的，而控制正态分布的两个参数就是均值和方差，所以引入了一个均值方差计算模块，对每张图片预测其均值方差，再进行正态分布的采样，从而得到隐变量Z，最后得出生成样本  $\hat{X}$

新的问题进一步出现，如果还是使用  $X$  与  $\hat{X}$  进行差值最小化，那么作为随机变量的方差（控制噪声），会逐渐趋近于0，因为越小的噪声实际上就越有助于恢复原来的图片。

因此，实践中还会加入一个新的loss，使得被估计出来的均值和方差更加接近标准正态分布。

### 3. 理解



VAE的原理实际上就如图所示，由 $X$ 预测出方差和均值。均值整体需要更加接近0，然后作为隐变量 $Z$ 的基本成分，而方差强度需要接近1，且作为噪声加入到 $Z$ 中，并且对这个 $Z$ 进行解码，从而得到最后的预测输出。

预测输出想要更加靠近输入，就希望噪声更低，但噪声本身的强度也有loss，所以会在输出效果较差时，生成较小的噪声，而在效果较好之后，噪声强度也逐步增强，直到稳定在1附近。