DDPM

理论部分

主要参考资料：[生成扩散模型漫谈（二）：DDPM = 自回归式VAE](), [Diffusion Model 入门（1）——概率生成式模型概述]()

从VAE概率角度看

变分：在真实后验概率较为复杂时，使用一个近似概率分布去近似真是概率分布的过程就叫做变分

推理：给定的隐变量时，用似然度对被观测变量进行采样。

扩散生成模型，本质上仍然是一个最大似然概率模型，也就是最大化解码器（生成器）在这个参数下，输出这个样本x的概率$p_\theta(x)$其中θ指的是生成器参数，x指的是数据样本。

Loss推导

首先让我们回归VAE的Loss部分，对于VAE而言，我们试图学习一种映射关系，可以将输入图像 x 映射到隐空间 z，然后再从隐空间 z 重构出原始图像，因此，我们想到了似然，这里的似然需要跟模型的隐变量进行挂钩，所以我们将似然写作边缘概率形式$p_\theta(x) = \int p_\theta(x,z)\,dz = \int p_\theta(x|z)p(z)\,dz$，由于神经网络采用的是端到端的学习，所以也就是输入x输出x，所以无法知道每个图像x和隐变量z之间的关系，无法监督的实现这个事情，所以直接计算这个积分是不可行的，因此我们想到了贝叶斯公式，可以有$p_\theta(x) = \frac{p_\theta(x,z)}{p_\theta(z|x)}$， 这样我们可以引入一个中间的过程$p_\theta(z|x)$这样每个输入的$x$都能知道其对应的隐变量$z$。然而由于我们$\theta$是解码器，其无法进行计算得到隐变量$z$，所以我们这里需要引入另外一个网络也就是编码器的近似分布$q_\phi(z|x)$用于近似解码器的真实分布$p_\theta(z|x)$。因此我们可以进行公示的推导了

$$p_\theta(x) = \int p_\theta(x,z)\,dz = \int \frac{p_\theta(x,z)}{q_\phi(z|x)} q_\phi(z|x) dz = \mathbb{E}_{z\sim q}\frac{p_\theta(x,z)}{q_\phi(z|x)}$$

对于解码器而言，他是能够知道$p_\theta(x|z)$的过程的，因此我们可以接着直觉上的对公式进行进一步的化简。

$$\mathbb{E}_{z\sim q}\frac{p_\theta(x,z)}{q_\phi(z|x)} = \mathbb{E}_{z\sim q}\frac{p_\theta(x|z)p_\theta(x)}{q_\phi(z|x)}$$

我们一般习惯性求的是对数似然，因此有

$$log p_\theta(x) = log \mathbb{E}_{z\sim q}\frac{p_\theta(x|z)p_\theta(x)}{q_\phi(z|x)}$$

利用詹森不等式，我们把期望和对数进行交换有

$$log p_\theta(x) = log \mathbb{E}_{x\sim q}\frac{p_\theta(x|z)p_\theta(z)}{q_\phi(z|x)} \geq \mathbb{E}_{x\sim q} log \frac{p_\theta(x|z)p_\theta(z)}{q_\phi(z|x)}$$

因此就求得了变分下界，我们要最大化变分下界也就是最小化负变分下界。

$$\mathcal{L} = -\mathbb{E}_{z\sim q} log \frac{p_\theta(x|z)p_\theta(z)}{q_\phi(z|x)} = D_{KL}(q_\phi(z|x)||p_\theta(z)) + \mathbb{E}_{z\sim q} - log p_\theta(x|z)$$

这样，我们就最终从似然的角度推导出了VAE的Loss。

其中第一项是先验分布和编码器近似分布的KL散度，目的是让我们模型的潜在空间分布近似于先验假设的标准正态分布，第二项是重构损失，也就是让我给定的x，通过这个x得到的z，通过这个z恢复的数据x之间的差异最小化。实际上VAE是一个对抗的过程，一方面，我们希望模型能够尽可能好地重构输入数据（通过最小化重构损失），这样会让隐空间变得复杂，不接近于先验的标准正态分布；另一方面，我们也希望模型的潜在空间分布能够接近我们

的先验假设（通过最小化KL散度损失）。

对于第一项KL散度和第二项为什么是MSE均有推导，这里不在赘述。