

VAE变分自编码器

1. VAE 概述

变分自编码器（VAE，Variational Autoencoder）是自编码器（Autoencoder）的扩展，具有**生成数据**的能力。VAE 的核心思想是通过一个**概率生成模型**来建模数据的潜在空间，使得模型能够学习到数据的隐式分布，并能够根据这个分布生成新的样本。

- 目标：**给定一个训练集，**学习到一个隐空间的分布**，使得该分布与训练数据分布 $p_{\text{data}}(x)$ 接近，从而能够从隐空间生成新的样本。
- 生成能力：**VAE不仅学习如何将**输入映射到低维潜在空间**（通过编码器），还学习如何**从潜在空间生成数据**（通过解码器）。

2. 编码器原理

VAE 中的编码器（Encoder）将输入数据 x 映射到潜在空间，并输出潜在变量的**均值** $\mu(x)$ 和**标准差** $\sigma(x)$ 。通过从潜在分布（通常是高斯分布）中采样，获得潜在变量 z 。

2.1 变分推断

VAE 通过 **变分推断**（Variational Inference）来近似真实的后验分布 $p(z|x)$ ，因为直接计算后验分布是非常困难的，尤其是对于复杂的模型。我们引入一个 **变分分布** $q(z|x)$ 来近似真实后验分布：

$$q(z|x) \approx p(z|x)$$

- 后验分布：**
- 变分分布：** $q(z|x)$ 是由编码器网络输出的潜在变量的分布，通常是高斯分布：

$$q(z|x) = \mathcal{N}(z; \mu(x), \sigma(x)^2)$$

这里， $\mu(x)$ 和 $\sigma(x)$ 是通过编码器网络从输入数据 x 中得到的参数。

2.2 重参数化技巧

为了使得变分推断可以通过梯度下降进行优化，VAE 使用了 **重参数化技巧 (Reparameterization Trick)**。通常，从变分分布中采样是不可导的，因此不能直接进行反向传播。通过重参数化技巧，将随机采样过程转化为确定性操作，使得我们可以通过反向传播来训练编码器。

- **重参数化公式：**

$$z = \mu(x) + \sigma(x) \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

这里， ϵ 是从标准正态分布 $\mathcal{N}(0, I)$ 中采样得到的噪声， $\mu(x)$ 和 $\sigma(x)$ 是由编码器网络输出的均值和标准差。

3. 解码器原理

VAE 中的解码器 (Decoder) 接收到从编码器采样得到的潜在变量 z ，然后通过解码器网络生成数据 x 的条件概率分布 $p(x|z)$ 。解码器的目标是根据潜在变量 z 生成与原始数据 x 尽量相似的样本。

- **生成模型：**解码器生成 x 的条件概率分布 $p(x|z)$ ，并通过该分布重建数据。
- **数据生成过程：**从潜在空间采样得到 z ，然后通过解码器生成新的数据 x_{new} 。

$$x_{\text{new}} = \text{Decoder}(z)$$

4. 损失函数

VAE 的损失函数由两部分组成：**重构损失**（Reconstruction Loss）和 **KL 散度**（KL Divergence）。

4.1 重构损失

重构损失度量了生成数据 x_{new} 和原始数据 x 之间的差异。常见的损失函数有均方误差（MSE）和交叉熵损失（Binary Cross-Entropy）。

$$\mathcal{L}_{\text{recon}} = -\mathbb{E}_{q(z|x)}[\log p(x|z)]$$

- 目标：**最大化生成数据与原始数据的相似度，即最大化 $\log p(x|z)$ 。

4.2 KL 散度

KL 散度度量了编码器输出的潜在变量分布 $q(z|x)$ 与先验分布 $p(z)$ （通常为标准正态分布 $\mathcal{N}(0, I)$ ）之间的差异。

$$\mathcal{L}_{\text{KL}} = \text{KL}(q(z|x) \parallel p(z)) = -\frac{1}{2} \sum_{i=1}^{D_z} (1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2)$$

- 目标：**最小化 KL 散度，使得潜在空间的分布接近标准正态分布，从而避免过拟合。

4.3 总损失函数

VAE 的总损失函数是重构损失和 KL 散度的和：

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{KL}}$$

- 重构损失：**确保生成的数据与原始数据尽可能相似。
- KL 散度：**确保潜在空间的分布接近标准正态分布。

5. 总结

VAE 是一个生成模型，通过学习数据的潜在空间分布来进行数据生成。其编码器将输入数据映射到潜在空间，解码器根据潜在变量生成数据。VAE 通过最大化变分下界（ELBO），优化重构损失和 KL 散度，从而使模型能够生成新的数据。