BETA-VAE

1. 文章概要

如果想让机器有用类似人类的学习和思考方式，我们就希望机器在**无监督情况下学习到可因式分解的 (factorized)、独立的 (independent) 表征**。这种能力被称为**解耦能力** (disentanglement)。

基于经典的变分自编码 (VAE) 做了一点点改变，增加了一个β参数来控制允许用来构建输出图片的比特数。

如果推断出的潜在表征中的每个变量只对单一生成因素敏感，而对其他因素相对相对不变，我们就说这个表征是分解的或因素化的。分解表征通常带来的一个好处是良好的可解释性和易于推广到各种任务。例如，一个在人脸照片上训练的模型可能会在不同的维度上捕捉到人的温柔、肤色、头发颜色、头发长度、情绪、是否戴眼镜以及其他许多相对独立的因素。这样的拆分表示对面部图像的生成非常有利。

$$L_{\mathrm{BETA}}(\phi, \beta) = -\mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x} \mid \mathbf{z}) + \beta D_{\mathrm{KL}}\left(q_\phi(\mathbf{z} \mid \mathbf{x}) \| p_\theta(\mathbf{z})\right)$$

β-VAE

Maximize

$$\mathbb{E}_{p(\mathbf{x})}[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \beta D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z}))]$$

β > 1 → constraints on the capacity of the latent bottleneck → encouraging the encoder to learn a disentangled representation

AnnealedVAE

$$\mathbb{E}_{p(\mathbf{x})}[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \gamma|D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})) - C|]$$

Small C → Small KL →low encoding capacity

Increase C gradually → larger KL budget → new variation encoded (one at a time)