

什么是变分自编码器

我们已经知道自编码器是一种两头宽中间窄，状如哑铃的神经网络，网络学习的目的是在输出端还原输入数据。由于中间的瓶颈层比较窄，只有那些最重要的信息才会在编码 z 中保留下来。因此，自编码器常用于提取数据中的关键信息。

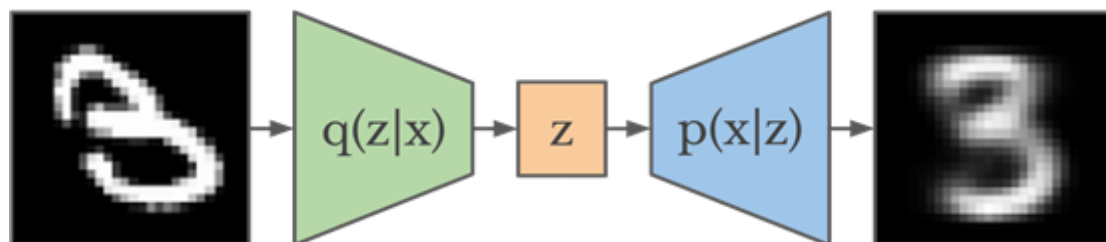


图 1：自编码器

事实上，瓶颈层的编码 z 由于维度较低，可以认为是处于原始数据的一个低维的子空间，如图 2 所示。在这个子空间里，只有那些最重要的信息保留下来，比如数字的类别，方向，笔迹粗细等，而那些不是特别重要的信息，如每个像素点的灰度值，会被忽略。这个低维子空间一般称为“流形”。假设我们在这个流形上改变 z 的值，就可以改变图片的属性，如图 2 所示。

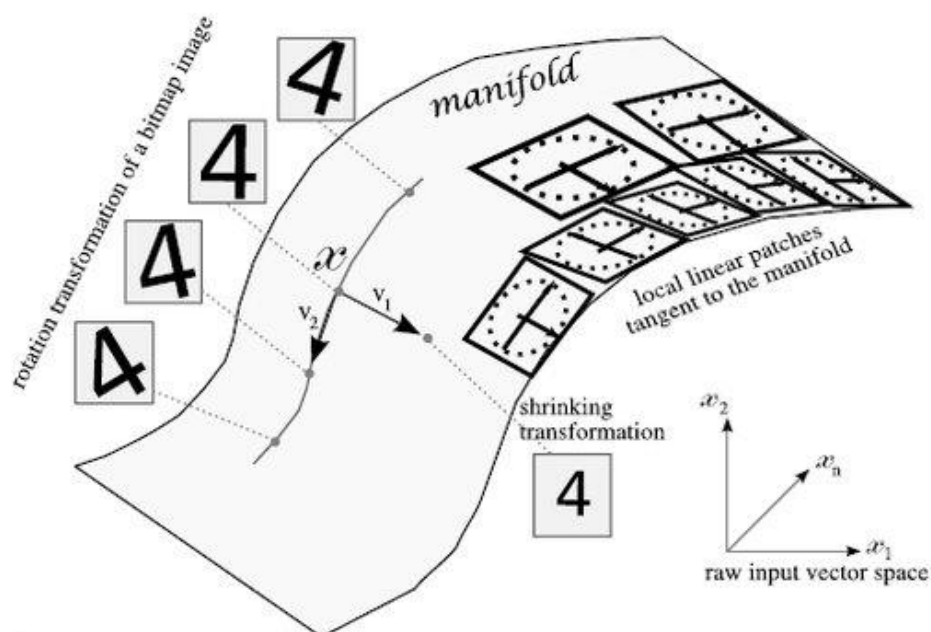


图 2：自编码器的流形空间[4]

现在的问题是，这个流形的空间是无限大的，那些真实数据所对应的位置既不规律，也不聚堆，使得子空间的操作很困难。一个方法是在训练时对 z 进行约束，使它符合某个简单的分布 $p(z)$ ，例如标准高斯分布 $N(0,1)$ 。有了这个约束， z 所处的位置就基本确定了。这样的模型称为变分自编码器(VAE)。

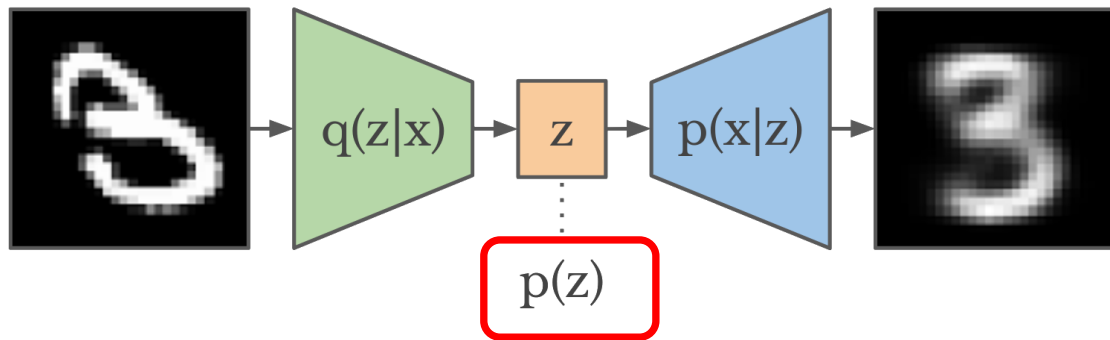


图 3：变分自编码器(VAE)

变分自编码器可以用作生成模型。这是因为模型训练完成后，所有真实数据的编码 z 符合既定的分布 $p(z)$ ，因此，如果我们可以从 $p(z)$ 中随机选出一个点，经过解码器后即可生成真实图片。DeepMind 的研究者发现，利用 VAE 模型可以得到非常真实的图片。



图 4：由 VAE 生成的人脸照片[2]

同时，我们也可以在 z 空间中游走，得到连续变化的图片，如图 5 所示。



图 5：由 VAE 生成的连续变化数字图片[3]

总结起来，变分自编码器继承了传统自编码器的哑铃结构，但对瓶颈编码 z 进行了约束，使之符合特定的分布。有了这个约束，就可以对编码进行随机生成和修改，从而得到各种有趣的生成效果。

[1]Kingma D P, Welling M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.

[2]Razavi A, Van den Oord A, Vinyals O. Generating diverse high-fidelity images with vq-vae-2[J]. Advances in neural information processing systems, 2019, 32. <http://www.gwylab.com/pdf/vqvae2.pdf>

[3]<https://ml-showcase.paperspace.com/projects/variational-autoencoder-vae>

[4] <https://www.quora.com/Can-auto-encoders-be-used-to-perform-manifold-learning>