

学习基础知识

论文复现有一点点小难，！！截止日期要到了只能先交部分笔记了，面试之前再继续研究尝试提交（如果允许的话）

encoder（编码器）与decoder（解码器）

编码器将输入序列转化成成一个向量，解码器将这个激活状态生成目标译文。整个过程将现实问题转化为数学问题并求解，再应用于解决现实问题。



KL散度(相对熵)

相对熵用于衡量两个概率分布之间的差异，它是从一个分布到另一个分布的信息损失。

$$D_{KL}(p||q) = H(p, q) - H(p) = - \sum_{x \in X} p(x) \log(q(x)) + \sum_{x \in X} p(x) \log(p(x)) = - \sum_{x \in X} p(x) \log\left(\frac{q(x)}{p(x)}\right)$$

性质：

KL散度具有如下性质：

1. 非负性；
2. 非对称性；
3. 当KL散度的取值位于 $(0, \infty)$ ，越接近于0，说明分布P与分布Q越匹配。

此外，概率分布P和分布Q之间的KL散度还可以表示为两个概率密度函数p和q之间对数差的期望。假设随机变量x为概率分布函数P的一个概率值，E为期望，那么KL公式还可如下定义：

$$D_{KL}(P||Q) = E_{x \sim P(x)}[\log \frac{p(x)}{q(x)}]$$

重参数化

结构重参数化构造一系列结构用于训练，并通过等价转换参数为另一组参数用于推理。这样，原始的结构可以等价转换为另一系列结构。在训练阶段，可能会使用具有更高精度的较大结构；而在推理阶段，则使用转换后的较小结构，同时保留相同的精度或其他性质。

隐变量

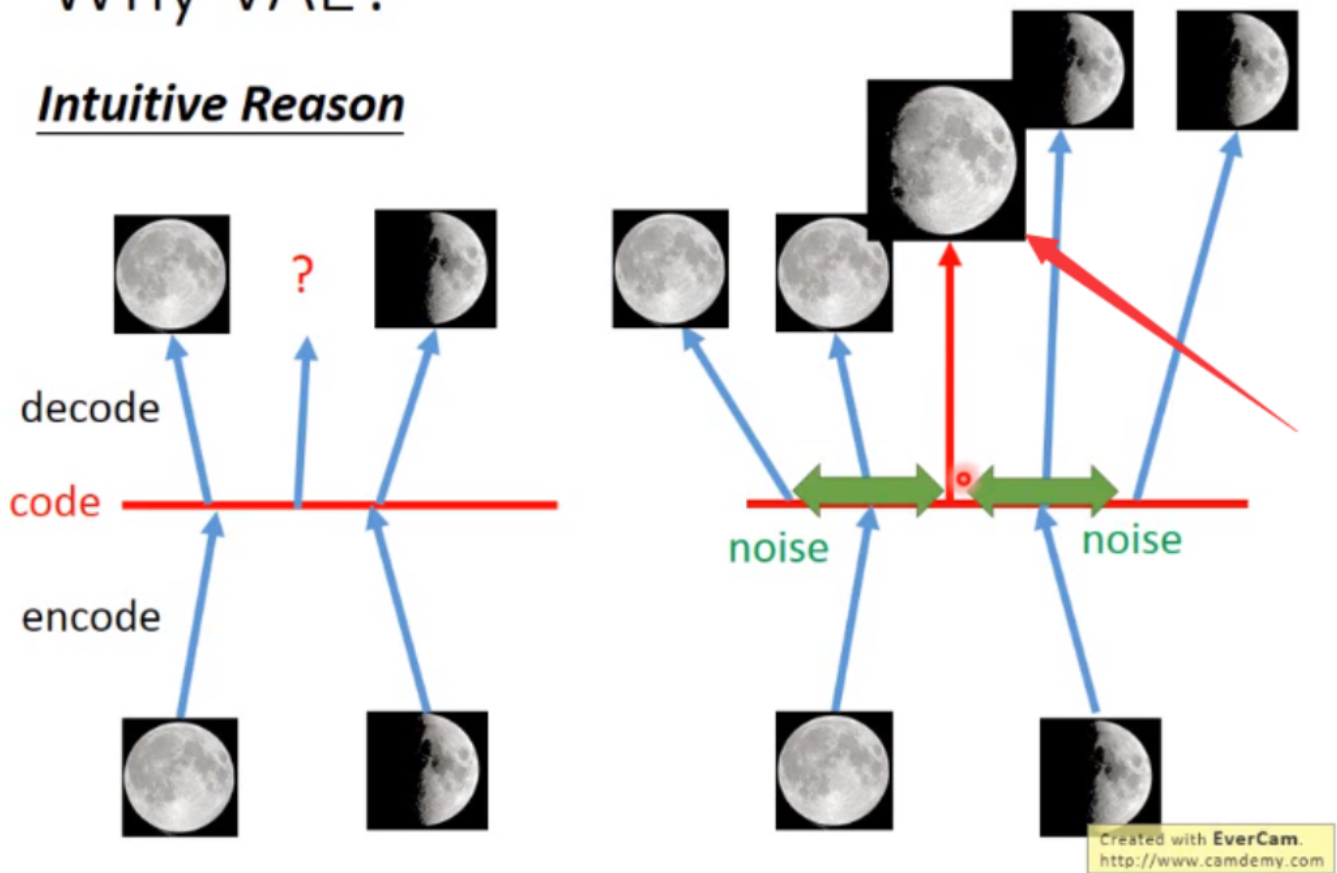
隐变量不可直接观测的综合性变量。在生成手写数字的过程中(0~9)，如果数字的左半部分包含数字5的左半部分，则右半部分不能包含0的左半部分，否则字符将很明显看起来不像任何真正的数字。简单来说，如果模型在给任何像素赋值之前，先决定生成哪个字符（或者说具有某种特征的字符），这对赋值像素将会很有帮助。这种决定在形式上被称为隐变量(Latent Variable)。

自编码器（AE）与变分自编码器（VAE）

1. 自编码器是通过将输入信息作为学习目标，对输入信息进行表征学习。VAE 模型是一种包含隐变量的生成模型，它利用神经网络训练得到两个函数（也称为推断网络和生成网络），进而生成输入数据中不包含的数据。
2. VAE能产生输入数据中不包含的数据，而AE只能产生尽可能接近或者就是以前的数据，甚至可能出现混乱的图片。
3. 训练时，AE训练得到Encoder和Decoder模型，而VAE除了得到这两个模型，还获得了隐藏层的分布模型。也正是因为VAE中包含隐变量的生成模型，才使得它能产生新的数据。
4. 变分自编码的核心思想：（离散变连续）

Why VAE?

Intuitive Reason



如左图所示，自编码器模型能还原这两张图片。接下来，我们在两张图片的编码点中间取一点，然后将这一点交给解码器，实际的结果是，生成图片是模糊且无法辨认的乱码图。因为编码和解码的过程使用了深度神经网络，这是一个非线性的变换过程，所以在code空间上点与点之间的迁移是非常没有规律的。

如右图所示，变分自编码器在给两张图片编码的时候加上一点噪音，使得每张图片的编码点出现在绿色箭头所示范围内，于是在训练模型的时候，绿色箭头范围内的点都有可能被采样到，这样解码器在训练时会把绿色范围内的点都尽可能还原成和原图相似的图片。我们可以试图把噪音无限拉长，使得对于每一个样本，它的编码会覆盖整个编码空间。

跨模态

跨模态通常指整合视觉（图像或视频）、听觉（声音或语音）、文本或其他类型的数据，以便构建能够处理和理解多种类型输入的系统。

（个人拙见）在论文中，跨模态指的是整合 环境图像 和 门与无人机的相对状态。这样构建出一种紧凑的数据，让后续的控制策略更精确；同时这样也能有效避免过拟合，让模拟环境训练出的模型也能用于真实环境；除此之外，也让测试员能够清楚地知道决策过程，在出现异常时可以立即停止。