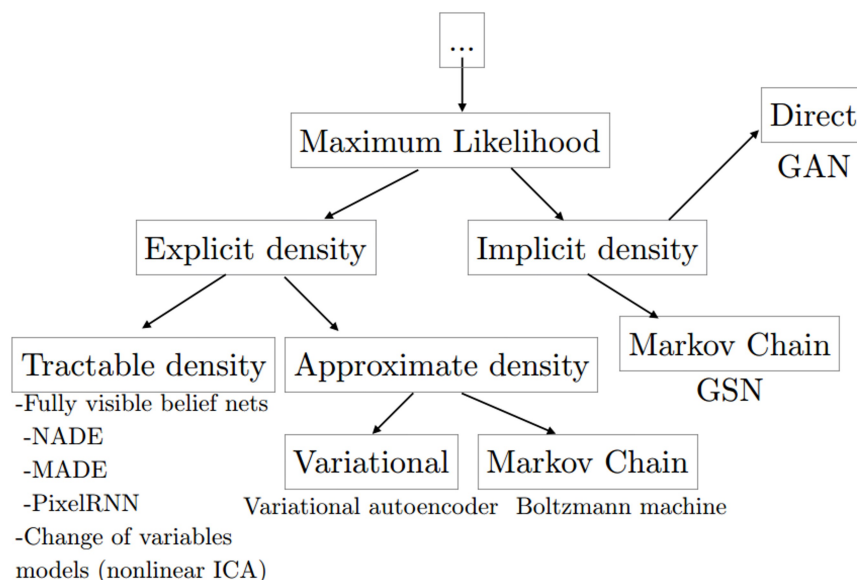


生成模型读书笔记六

2021年1月11日 21:51

1. 再谈生成模型家族

上面从最大似然开始介绍了这么多，现在我们再回过头去看生成模型的图谱，会发现前面讲的正是图中左边的部分分支。



从上图可以看到，下面正式讲解下这个图中的各个分支。

所有的模型都基于最大似然的原则，区别在于模型如何表示最大似然。因此有两个分支，一个是显式地对似然函数建模或者近似，然后最大化似然，求出分布的具体密度函数。二是隐式的，不求解具体的似然函数，而是学习从分布中采样或者直接从零生成符合分布的样本。

a. 显式概率密度

显式概率密度指直接对数据的概率密度函数建模，即直接定义 $p_{model}(x; \theta)$ ，然后依照最大似然求其中的参数。这类方法的难点在于，如何设计一个足够模拟数据复杂分布的模型。如上图所示，有两个解决方案，一是精心设计一个可解的概率密度形式，二是用一个近似的模型去逼近真实的密度分布。

i. 可解的密度模型

这类方法的常用算法都列在了图中，主要分为两大类，一是FVBN，一类是非线性ICA。

1) FVBN

数据中包含的所有变量都是可以观测到的，根据设计的概率图模型写出变量之间的关系，用概率链式法则把变量的联合分布概率写成各个变量条件概率相乘的形式，再去求解每个条件分布的参数

$$p_{\text{model}}(\mathbf{x}) = \prod_{i=1}^n p_{\text{model}}(x_i | x_1, \dots, x_{i-1})$$

典型模型：WaveNet

缺点：生成速度慢，需要按照变量依赖关系，先生成 x_1 ，再生成 x_2 ，...，从

而导致了不能并行运行，速度慢。

2) 非线性ICA

用于数据中存在隐变量的情况，定义一个连续的非线性的可逆的从隐变量空间到观测变量空间的变换 g ，再通过下式求得 x 的分布 $p(x)$ 。

$$p_x(\mathbf{x}) = p_z(g^{-1}(\mathbf{x})) \left| \det \left(\frac{\partial g^{-1}(\mathbf{x})}{\partial \mathbf{x}} \right) \right|.$$

缺点：对于变换函数 g 的限制太多，适用的分布太少。

本类方法特点：因为直接求解数据的真实分布(最大似然)，所以很有效准确，但是必须设计一个可解的密度函数，因此可选的函数有限，适用的分布也有限。

ii. 近似密度模型

既然直接解似然函数比较困难，于是就有了显式密度分支下的第二个分支：用一个比较容易求解的模型去近似真实分布。这种方法也可以分为两类，我们在前面"概率分布的推断"中就已经介绍过了，一类是变分法，另一类是基于采样近似的MCMC方法。

1) 变分近似

如前所述，变分近似的流程一般是找到似然函数/KL散度的一个下界，比如前面所说的ELBO，这个下界是容易求解的，从而将求解最大似然的问题转化成了最大化可求解下界的问题。

典型模型：VAE

缺点：回想在VAE中，近似的后验分布先验分布都是我们预先指定的，容易算的，但是如果他们的复杂度或者说capacity比真实分布小的话，那么即使求得最优解也只能是近似，无法做到无偏估计，最后求出的分布也不是真实的分布。因此VAE也经常得到较低质量的照片。

2) 马尔科夫链近似

这是一种用马尔科夫链进行样本采样的统计近似。在这个显式概率模型的分支下面，它被玻尔兹曼机用于模型的训练。(玻尔兹曼机在深度学习复兴初期有非常重要的地位，但是由于依赖的马尔科夫链方法难以规模化，以及需要低效的多步生成而不是现在流行的一步生成，所以现在已经很少使用了，本文也就不多做介绍了。)

缺点：马尔科夫链采样需要经过一定的时间才能收敛，而在此之前的样本都不是来自于目标分布的，在实际应用中，通常会有过早使用样本的问题。而且对于高维空间，这种方法比较低效。

b. 隐式概率密度

这个分支不直接定义密度函数 $p_{model}(x; \theta)$ ，不要求解出密度函数的具体形式，只要求能够采样或者生成一个来自目标分布的样本，但是这类方法其实本质上也是在间接地学密度分布(而且也有办法可以让它得到密度函数)，所以就称为隐式密度。

这个分支也有两类，一类是基于马尔科夫链采样的，不停地迭代采样，直到收敛于目标分布。从而达到从分布采样的目的，代表模型有generative stochastic network，马尔科夫链的缺点上文已经说过了。另一类就是GAN了，GAN学习如何一步生成符合目标分布的样本，下一节开始将详细地介绍GAN。

(终于来到激动人心的分支了，我十天前为了想要知道GAN和其他生成方法的异同，开始看资料学习，结果学习前面的就花了我十天时间，说了这么多最终其实是为了GAN。。。)

c. 一些补充

在上述所介绍的生成模型中，FVBN, VAE和GAN是最常用的三种生成模型，而其中VAE和GAN也是最常用来比较的两种深度生成模型。

对比上述各类模型的缺点，GAN有以下的一些特点：

- i. 对比一次只能生成一个样本的FVBN，GAN生成速度较快，可以并行批量生成样本
- ii. 与ICA一样，GAN也是要学习一个从隐空间到数据空间的映射，但是对比ICA，GAN对于生成器generator没有太多的要求，比较灵活，只要是可导的都可以作为GAN的生成器。
- iii. 可以一步生成，对比需要多步迭代的基于马尔科夫链的生成模型玻尔兹曼机和GSN
- iv. 对比VAE，不需要变分下界，GAN是渐近一致的(就是说样本量无穷多的情况下，估计误差可以无穷小，有猜想VAE也可以渐近一致，但是还没有证明)(好复杂，实际应用中也很难体现，不是我想要的不同，后面我会再自己总结和对比一遍)。其他不同：VAE是参数模型而GAN是非参数模型。VAE假设隐变量服从高斯分布的，用变分推断该高斯分布的参数，即均值和方差；而GAN，则是要学习一个从隐空间到数据空间的映射， $Y=f(X)$
- v. 实验证明GAN可以生成更好的样本