VAE变分自编码器公式推导

VAE变分推导依赖数学公式

(1) 贝叶斯公式：$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$

(2) 边缘概率公式：$p(x) = \int p(x,z)dz$

(3) KL 散度公式：$D_{KL}(p||q) = \int p(x)log\frac{p(x)}{q(x)}dx$

推导方式一

注：一般随机变量是用大写字母表示，随机变量的取值用小写字母表示，随机变量的概率密度函数是用小写字母，而随机变量的分布函数是用大写字母，此处忽略字母的大小区别统一用小写字母表示。

贝叶斯变分自编码器，参考【1】中描述，公式推导是用一个分布q(z|x)去近似p(z|x)，即从 $D_{KL}(q(z|x)||p(z|x))$出发推导

$D_{KL}(q(z|x)||p(z|x))$ = $\int q(z|x)\cdot log\frac{q(z|x)}{p(z|x)}dz$ = $\int q(z|x)\cdot log\frac{q(z|x)}{\frac{p(z,x)}{p(x)}}dz$ = $\int q(z|x)\cdot log\frac{q(z|x)}{\frac{p(x|z)\cdot p(z)}{p(x)}}dz$

= $\int q(z|x)\cdot log\frac{q(z|x)\cdot p(x)}{p(x|z)\cdot p(z)}dz$

=$\int q(z|x)\cdot logq(z|x)dz + \int q(z|x)\cdot logp(x)dz - \int q(z|x)\cdot logp(x|z)dz - \int q(z|x)\cdot logp(z)dz$

=$logp(x) + \int q(z|x)\cdot logq(z|x)dz - \int q(z|x)\cdot logp(z)dz - \int q(z|x)\cdot logp(x|z)dz$

=$logp(x) + \int q(z|x)\cdot log\frac{q(z|x)}{p(z)}dz - \int q(z|x)\cdot logp(x|z)dz$

=$logp(x) + KL(q(z|x)||p(z)) - E_{z\sim q(z|x)}[logp(x|z)]$

==>

$logp(x) = D_{KL}(q(z|x)||p(z|x)) - KL(q(z|x)||p(z)) + E_{z\sim q(z|x)}[logp(x|z)]$

令$L(q) = -KL(q(z|x)||p(z)) + E_{z\sim q(z|x)}[logp(x|z)]$，对应给定的x, log(x)是一个常量，最小化 $D_{KL}(q(z|x)||p(z|x))$， 也就最大化L(q)。又由于KL散度 $D_{KL}(q(z|x)||p(z|x))$ 是非负的，因此 $logp(x)\geq L(q)$ ，把L(q)称之为变分下界 Evidence Lower Bound。

推导方式二

$logp(x) = logp(x)\cdot\int q(z|x)dz = \int q(z|x)\cdot logp(x)dz$

$logp(x) = log\frac{p(x,z)}{p(z|x)} = logp(x,z) - logp(z|x) = log\frac{p(x,z)}{q(z|x)} - log\frac{p(z|x)}{q(z|x)}$

同时对上述公式两侧乘以q(z|x)并取积分得：

$logp(x) = \int q(z|x)\cdot log\frac{p(x,z)}{q(z|x)}dz - \int q(z|x)\cdot log\frac{p(z|x)}{q(z|x)}dz$

= $D_{KL}(q(z|x)||p(z|x)) + \int q(z|x)\cdot log\frac{p(x|z)p(z)}{q(z|x)}dz$

=$D_{KL}(q(z|x)||p(z|x)) + \int q(z|x)\cdot logp(x|z)dz + \int q(z|x)\cdot log\frac{p(z)}{q(z|x)}dz$

=$D_{KL}(q(z|x)||p(z|x)) + E_{z\sim q(z|x)}[logp(x|z)] - D_{KL}(q(z|x)||p(z))$

这里就得到和上述推导一样的公式。令$L(q) = -KL(q(z|x)||p(z)) + E_{z\sim q(z|x)}[logp(x|z)]$

由于后验概率$p(z|x)$无法求解，于是希望找到一个$q(z|x)$近似$p(z|x)$

公告

昵称： 星辰大海,绿色星球
园龄： 7年5个月
粉丝： 2
关注： 0
+加关注

< 2024年2月 >

日	一	二	三	四	五	六
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	1	2
3	4	5	6	7	8	9

搜索

找找看

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

我的标签

深度学习(29)
多模态预训练(15)
机器学习(12)
高等代数(7)
语音识别(4)
卷积神经网络CNN(4)
大模型(4)
小故事(3)
MPI(2)
强化学习(1)

随笔档案

2024年2月(4)
2024年1月(1)
2023年11月(5)
2023年10月(3)
2023年9月(1)
2023年8月(2)
2023年7月(2)

由于任意KL散度$D_{KL}(q(z|x)||p(z|x))$ 都是非负的，所以$logp(x) \geq L(q)$，L(q)是关于分布函数q的泛函。从似然函数观点出发，最大化L(q)可以导出最大化logp(x)。

理解公式含义

上面介绍了公式推导，按照【1】中描述。

（1）Encoder-Decoder模型中中间的隐变量是具体的某个值，而VAE不是研究隐变量具体某个值，而是学习隐变量的某个分布，使得在这个分布上取样时，Decoder仍可以得到相似的输出。

（2）贝叶斯公式：$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$，其中p(z)可以假定服从某个分布如高斯分布，p(x|z) decoder可以是用神经网络来表示，但是要求后验概率还需要求解p(x)，由于$p(x) = \int p(x|z)p(z)dz$，z一般是维度很高的变量，求上述积分是很困难的，所以求后验概率是很困难。

（3）求上述积分大致有两种解法：蒙特卡罗和变分推断

变分推断：

既然p(z|x)求解很困难，那么尝试用一个方便求解的分布q(z|x)去近似p(z|x)，于是就想到最小化二者的KL散度即$minKL(q(z|x)||p(z|x))$，这也是上述推导，从$D_{KL}(q(z|x)||p(z|x))$开始的原因。通过公式推导变换，最小化$D_{KL}(q(z|x)||p(z|x))$等价于最大化L。

现在来看最大化L对应的含义：随机变量z, 其分布是q(z|x)，$E_{z\sim q(z|x)}[logp(x|z)]$最大化表明，在这个分布上不断对随机变量z进行取样，使得重建出x的几率最大；对于 $-KL(q(z|x)||p(z))$ 最大化即最小化 $KL(q(z|x)||p(z))$，使得求解的后验概率q(z|x)和先验分布p(z)尽量接近。

变分下界求导求解

依赖数学公式：

（1）**期望定义**：设离散随机变量X的分布列为$p(x_i) = P(X = x_i), i = 1, 2, \ldots, n, \ldots$, 如果级数$\sum_{i=1}^{+\infty} |x_i|p(x_i)$收敛，则称

$$E(X) = \sum_{i=1}^{+\infty} x_i p(x_i)$$

期望定义： 设连续随机变量X的密度函数为p(x)，如果无穷积分$\int_{-\infty}^{+\infty} |x|p(x)dx$ 存在，则称

$$E(X) = \int_{-\infty}^{+\infty} xp(x)dx$$

（2）**方差定义**： 如随机变量X^2的数学期望$E[X^2]$存在，则称偏差平方$(X - E[X])^2$的数学期望$E[(X - E[X])^2]$为随机变量X的方差，记为

$$Var(X) = E[(X - E[X])^2] = \begin{cases} \sum_i [x_i - E(x)]^2 p(x_i), & \text{在离散场合} \\ \int_{-\infty}^{+\infty} [x - E(X)]^2 p(x)dx, & \text{在连续场合} \end{cases}$$

（3）**方差性质**： $Var(X) = E(X^2) - [E(X)]^2$

（4）**正态分布**： 若随机变量X的密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-u)^2}, -\infty < x < +\infty$$

则称X服从正态分布，称X为正态变量，记作$X \sim N(\mu, \sigma^2)$

其期望：$E[X] = \mu$，方差：$Var(X) = \sigma^2$

现在就开始L(q)的求导

$L(q) = -KL(q(z|x)||p(z)) + E_{z\sim q(z|x)}[logp(x|z)]$

变分下界公式第一项记为$L_{1st}(q) = -KL(q(z|x)||p(z))$，前期为了推导过程简洁，省去了概率密度函数涉及的参数ϕ, θ，接下来推导中添加上这些。即

$L_{1st}(q) = -KL(q_\phi(z|x)||p_\theta(z)) = -\int q_\phi(z|x) \cdot log\frac{q_\phi(z|x)}{q_\theta(z)}dz = -\int q_\phi(z|x) \cdot logq_\phi(z|x)dz + \int q_\phi(z|x) \cdot logp_\theta(z)dz$

注：上文为了简化推导过程，尤其方便手写推导过程，上述公式中的变量z是高维的，应该用黑体 $\mathbf{z}$ 表示。后续推导过程仍延续小写字母，但是最终结论中要将变量z看做是高维的，或最后展示时采取黑体加粗表示。

为了后续求导解决方便做出如下几点假设：

（1）其变量$\mathbf{z}$各个分量/维度相互独立，此处仅将z是一维进行推导，高维的最终由于独立性可以拆分

（2）$p_\theta(z)$是一个正态分布的概率密度函数，即 $p_\theta(z) = N(z; 0, 1)$

（3）$q_\phi(z|x)$也是一个正态分布的概率密度函数，即 $q_\phi(z|x) = N(z; u(x, \phi), \sigma^2(x, \phi))$

由期望的定义得

$\int q_\phi(z|x) \cdot logp_\theta(z)dz$

$= E_{z\sim q_\phi(z|x)}[logp_\theta(z)]$

2023年5月(3)
2023年3月(2)
2023年1月(3)
2022年11月(5)
2022年10月(2)
2022年9月(1)
2022年8月(1)
2022年5月(5)
更多

$= E_{z\sim q_\phi(z|x)}[logN(z;0,1)]$

$= E_{z\sim q_\phi(z|x)}[log\frac{1}{\sqrt(2\pi)}e^{-\frac{1}{2}z^2}]$

$= E_{z\sim q_\phi(z|x)}[log\frac{1}{\sqrt(2\pi)} - \frac{1}{2}z^2]$

$= log\frac{1}{\sqrt(2\pi)} - \frac{1}{2}E_{z\sim q_\phi(z|x)}[z^2]$

$= -\frac{1}{2}log2\pi - \frac{1}{2}(\mu^2+\sigma^2)$

注：最后一步推导可以由方差性质得到

同样地：

$\int q_\phi(z|x)\cdot logq_\phi(z|x)dz$

$= E_{z\sim q_\phi(z|x)}[logN(z;u,\sigma^2)]$

$= E_{z\sim q_\phi(z|x)}[log\frac{1}{\sqrt(2\pi\sigma^2)}e^{-\frac{1}{2\sigma^2}(z-\mu)^2}]$

$= E_{z\sim q_\phi(z|x)}[log\frac{1}{\sqrt(2\pi\sigma^2)} - \frac{1}{2\sigma^2}(z-\mu)^2]$

$= log\frac{1}{\sqrt(2\pi\sigma^2)} - \frac{1}{2\sigma^2}E_{z\sim q_\phi(z|x)}[(z-\mu)^2]$

$= -\frac{1}{2}log2\pi - log\sigma - \frac{1}{2\sigma^2}Var(z)$

$= -\frac{1}{2}log2\pi - log\sigma - \frac{1}{2\sigma^2}\sigma^2$

$= -\frac{1}{2}log2\pi - log\sigma - \frac{1}{2}$

下面将变量z使用黑体$\mathbf{z}$表示来推导多维变量情况下公式，假设$\mathbf{z}$的维度是J，由于各个维度独立无关

$\int q_\phi(\mathbf{z}|x)\cdot logp_\theta(\mathbf{z})d\mathbf{z}$

$= E_{\mathbf{z}\sim q_\phi(\mathbf{z}|x)}[logp_\theta(\mathbf{z})]$

$= E_{\mathbf{z}\sim N(\mathbf{z};\mathbf{u},\Sigma)}[logN(\mathbf{z};\mathbf{0},\mathbf{I})]$

$= E_{z_1\sim N(z_1;u_1,\sigma_1^2),z_2\sim N(z_2;u_2,\sigma_2^2),\ldots,z_J\sim N(z_J;u_J,\sigma_J^2)}[logN(\mathbf{z};\mathbf{0},\mathbf{I})]$

简写为：

$= E_{z_1\sim N(z_1),z_2\sim N(z_2),\ldots,z_J\sim N(z_J)}[logN(\mathbf{z};\mathbf{0},\mathbf{I})]$

$= E_{z_1\sim N(z_1)}\cdot E_{z_2\sim N(z_2)},\ldots,\cdot E_{z_J\sim N(z_J)}[logN(\mathbf{z};\mathbf{0},\mathbf{I})]$

$= E_{z_1\sim N(z_1)}\cdot E_{z_2\sim N(z_2)},\ldots,\cdot E_{z_J\sim N(z_J)}[logN(z_J;0,1)+logN(z_{J-1};0,1)+\ldots$
$+logN(z_1;0,1)]$

$= E_{z_1\sim N(z_1)}\cdot E_{z_2\sim N(z_2)},\ldots,\cdot E_{z_{J-1}\sim N(z_{J-1})}[E_{z_J\sim N(z_J)}[logN(z_J;0,1)]+logN(z_{J-1};0,1)+\ldots$
$+logN(z_1;0,1)]$

$E_{z_J\sim N(z_J)}[logN(z_J;0,1)]$ 对于其它变量求期望相当于是一个常量，可以移出到外面，于是有

$= E_{z_J\sim N(z_J)}[logN(z_J;0,1)] + E_{z_1\sim N(z_1)}\cdot E_{z_2\sim N(z_2)},\ldots,$
$\cdot E_{z_{J-1}\sim N(z_{J-1})}[logN(z_{J-1};0,1)+\ldots+logN(z_1;0,1)]$

依次类推求得

$= E_{z_J\sim N(z_J)}[logN(z_J;0,1)] + E_{z_{J-1}\sim N(z_{J-1})}[logN(z_{J-1};0,1)]+\ldots+E_{z_1\sim N(z_1)}[logN(z_1;0,1)]$

$= -\frac{J}{2}log2\pi - \frac{1}{2}\sum_{j=1}^J(\mu_j^2+\sigma_j^2)$

即$\int q_\phi(\mathbf{z}|x)\cdot logp_\theta(\mathbf{z})d\mathbf{z}$
$= -\frac{J}{2}log2\pi - \frac{1}{2}\sum_{j=1}^J(\mu_j^2+\sigma_j^2)$

同理可以得出

$\int q_\phi(z|x)\cdot logq_\phi(z|x)dz$

$= -\frac{J}{2}log2\pi - \frac{1}{2}\sum_{j=1}^J(1+log\sigma_j^2)$

$L_{1st}(q) = \frac{1}{2}\sum_{j=1}^J(1+log\sigma_j^2-u_j^2-\sigma_j^2)$

【1】https://zhuanlan.zhihu.com/p/83865427

【2】https://www.cnblogs.com/hugh2006/p/9693891.html

【3】【机器学习】【白板推导系列】bilibili视频

【4】https://zhuanlan.zhihu.com/p/25429486

【5】https://blog.csdn.net/tsinghuahui/article/details/80530750

【6】Auto-Encoding Variational Variational Bayes | AISC Foundational

【7】An Introduction to Variational Autoencoders

标签: 多模态预训练