

📤 🍆 (-)阿里云 游戏专享 性能强劲不卡顿 4核16G 10M 24,83元/月/尝鲜体验 8核32G 10M 90.6元/月/火力全开





会员

周边

新闻

博问

AI培训

云市场

代码改变世界

Q

注册



中容回 自原 新随管 ST. A 订道

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: $p(z|x)=rac{p(x|z)p(z)}{p(x)}$

(2) 边缘概率公式: $p(x) = \int p(x,z)dz$

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

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

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

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

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

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

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

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

==>

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

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) > L(q) ,把L(q)称之为变分下界 Evidence Lower Bound。

推导方式二

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

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

同时对上述公式两侧乘以q(z|x)并取积分得:

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

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

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

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

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

<			202	24年2月			>
	日	_	=	Ξ	四	五	六
2	28	29	30	31	1	2	<u>3</u>
	4	5	6	7	8	9	10
	11	12	13	<u>14</u>	<u>15</u>	<u>16</u>	17
1	18	19	20	21	22	23	24
2	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)

2024/2/26 16:52

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

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

- (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

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

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 $\phi, heta$,接下来推导中添加上这些。即

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

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 **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**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) $p_{ heta}(z)$ 是一个正态分布的概率密度函数,即 $p_{ heta}(z)=N(z;0,1)$
- (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

 $\int q_{\phi}(z|x) \cdot log p_{\theta}(z) dz$

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

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

阅读排行榜

- 1. DeepSpeech2(4844)
- 2. 如何计算卷积神经网络中接受野尺寸(2050)
- 3. StyleGan(2032)
- 4. Attention is all you need(1783)
- 5. 分布式训练——Ring-AllReduce(1655)

评论排行榜

1. DDPM [diffusers] 保姆级代码解释 (1)(2)

推荐排行榜

- 1. DDPM [diffusers] 保姆级代码解释 (1)(1)
- 2. 如何计算卷积神经网络中接受野尺寸(1)

最新评论

1. Re:DDPM [diffusers] 保姆级代码解释 (1) timestep特征维度是block_out_channels[0]的4倍, 为什么?

--一只楚楚猫

2. Re:DDPM [diffusers] 保姆级代码解释 (1) 牛腳

--苏从JS

$$egin{align*} &= E_{z \sim q_{\phi}(z|x)}[logN(z;0,1)] \ &= E_{z \sim q_{\phi}(z|x)}[lograc{1}{\sqrt{(2\pi)}}e^{-rac{1}{2}z^2}] \ &= E_{z \sim q_{\phi}(z|x)}[lograc{1}{\sqrt{(2\pi)}}-rac{1}{2}z^2] \ &= lograc{1}{\sqrt{(2\pi)}}-rac{1}{2}E_{z \sim q_{\phi}(z|x)}[z^2] \ &= -rac{1}{2}log2\pi - rac{1}{2}(\mu^2 + \sigma^2) \ egin{align*} 注: 最后一步推导可以由方差性质得到 \end{align*}$$

同样地:

$$\begin{split} &\int q_{\phi}(z|x) \cdot logq_{\phi}(z|x)dz \\ &= E_{z \sim q_{\phi}(z|x)}[logN(z;u,\sigma^2)] \\ &= E_{z \sim q_{\phi}(z|x)}[log\frac{1}{\sqrt{(2\pi\sigma^2)}}e^{-\frac{1}{2\sigma^2}(z-\mu)^2}] \\ &= E_{z \sim q_{\phi}(z|x)}[log\frac{1}{\sqrt{(2\pi\sigma^2)}} - \frac{1}{2\sigma^2}(z-\mu)^2] \\ &= log\frac{1}{\sqrt{(2\pi\sigma^2)}} - \frac{1}{2\sigma^2}E_{z \sim q_{\phi}(z|x)}[(z-\mu)^2] \\ &= -\frac{1}{2}log2\pi - log\sigma - \frac{1}{2\sigma^2}Var(z) \\ &= -\frac{1}{2}log2\pi - log\sigma - \frac{1}{2\sigma^2}\sigma^2 \\ &= -\frac{1}{2}log2\pi - log\sigma - \frac{1}{2} \end{split}$$

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

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

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

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

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

简写为:

$$egin{align*} &=E_{z_1\sim N(z_1),z_2\sim N(z_2),...,z_J\sim N(z_J)}[logN(\mathbf{z};\mathbf{0},\mathbf{I})] \ &=E_{z_1\sim N(z_1)}\cdot E_{z_2\sim N(z_2)},\dots,\cdot E_{z_J\sim N(z_J)}[logN(\mathbf{z};\mathbf{0},\mathbf{I})] \ &=E_{z_1\sim N(z_1)}\cdot E_{z_2\sim N(z_2)},\dots,\cdot E_{z_J\sim N(z_J)}[logN(z_J;0,1)+logN(z_{J-1};0,1)+\dots \ &+logN(z_1;0,1)] \ &=E_{z_1\sim N(z_1)}\cdot E_{z_2\sim N(z_2)},\dots,\cdot E_{z_{J-1}\sim N(z_{J-1})}[E_{z_J\sim N(z_J)}[logN(z_J;0,1)]+logN(z_{J-1};0,1)+\dots \ &+logN(z_1;0,1)] \ \end{split}$$

 $E_{z_J \sim N(z_J)}[logN(z_J;0,1)]$ 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

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

依次类推求得

$$egin{aligned} &= E_{z_{J} \sim N(z_{J})}[logN(z_{J};0,1)] + E_{z_{J-1} \sim N(z_{J-1})}[logN(z_{J-1};0,1)] + \ldots + E_{z_{1} \sim N(z_{1})}[logN(z_{1};0,1)] \ &= -rac{J}{2}log2\pi - rac{1}{2}\sum_{j=1}^{J}\left(\mu_{j}^{2} + \sigma_{j}^{2}
ight) \end{aligned}$$

即
$$\int q_{\phi}(\mathbf{z}|x) \cdot logp_{ heta}(\mathbf{z})d\mathbf{z} = -rac{J}{2}log2\pi - rac{1}{2}\sum_{i=1}^{J}\left(\mu_{i}^{2} + \sigma_{i}^{2}
ight)$$

同理可以得出

$$egin{aligned} \int q_\phi(z|x) \cdot log q_\phi(z|x) dz \ &= -rac{J}{2} log 2\pi - rac{1}{2} \sum_{j=1}^J \left(1 + log \sigma_j^2
ight) \ &L_{1st}(q) = rac{1}{2} \sum_{j=1}^J \left(1 + log \sigma_j^2 - u_j^2 - \sigma_j^2
ight) \end{aligned}$$

- [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





<u>星辰大海,绿色星球</u> 粉丝 - 2 <u>关注 - 0</u>

0 0

升级成为会员

« 上一篇: <u>机器学习——期望与方差</u>

»下一篇: 回归

posted @ 2022-08-23 21:24 星辰大海,绿色星球 阅读(944) 评论(0) 编辑 收藏 举报

会员力量, 点亮园子希望

刷新页面 返回顶部

登录后才能查看或发表评论,立即 登录 或者 逛逛 博客园首页

【推荐】发个阿里云广告,对园子很重要: 阿里云上部署幻兽帕鲁

【推荐】园子的第一款简陋鼠标垫,是否是您值得拥有的周边 【推荐】编程路上的催化剂:大道至简,给所有人看的编程书

【推荐】会员力量,点亮园子希望,期待您升级成为园子会员



编辑推荐:

- · 都说了能不动就别动,非要去调整,出生产事故了吧
- · Redis 分布式锁的正确使用姿势
- ・聊一聊程序员沟通相关的问题
- · 优化接口设计的思路系列: 分页接口的设计和优化
- ·记一次 .NET某列控连锁系统 崩溃分析

阅读排行:

- ·在做程序员的道路上,你掌握了什么概念或技术使你感觉自我提升突飞猛进?
- ·都说了能不动就别动,非要去调整,出生产事故了吧 ightarrow 补充
- · 前端树形Tree数据结构使用- \$ 各种姿势总结
- · ASP.NET Core MVC应用模型的构建[1]: 应用的蓝图
- ·美团面试:说说OOM三大场景和解决方案? (绝对史上最全)

Copyright © 2024 星辰大海,绿色星球 Powered by .NET 8.0 on Kubernetes