KL散度的三种估计k1 k2 k3

欲壑难填 关注

来自专栏 · 强化学习/RLHF

> 原文是 John Schulman 的博客：joschu.net/blog/kl-appr...

John Schulman 在多处代码实现中采用的对 KL 散度+ $\mathbf{KL}[q,p]$ 的估计是 $\frac{1}{2}(\log p(x)-\log q(x))^2$，并非常规的 $\log \frac{q(x)}{p(x)}$。本文将介绍这种对 KL 散度的估计形式为什么更好（虽然是有偏的），以及如何进一步追求无偏且低方差的 KL 散度估计。

k1 估计

KL 散度定义为：

$$\mathbf{KL}[q,p]=\sum_x q(x)\log\frac{q(x)}{p(x)}=\mathbb{E}_{x\sim q}\left[\log\frac{q(x)}{p(x)}\right]$$

由于计算复杂度高或分布本身没有闭式解等原因，KL 散度一般是无法解析求解的，我们需要从 q 分布中采样样本 $x_1,x_2,\cdots\sim q$，然后用蒙特卡洛方法对 KL 散度进行估计。我们对估计的 KL 散度有两个要求，第一最好是无偏的，即估计值的期望与真实值相等，第二方差要尽可能小。

最常见的对 KL 散度的估计，是直接按 KL 散度的定义，取：

$$k1=\log\frac{q(x)}{p(x)}=-\log r$$

这里记 $r=\frac{p(x)}{q(x)}$，我们将这个估计记作 $k1$。$k1$ 的形式就是按定义来的，所以他显然是无偏的。但是，它其中有个 $\log$ 函数，当 $\frac{q(x)}{p(x)}<1$ 时，它的值是负的，而我们知道 KL 散度一定是正的，所以说它的方差很大。因此，$k1$ 这个估计不满足低方差的要求。

f 散度+与k2 估计

我们考虑第二种估计：

$$k2=\frac{1}{2}\left(\log\frac{p(x)}{q(x)}\right)^2=\frac{1}{2}(\log r)^2$$

这个估计看起来很不错，它能表征出 p,q 两分布之间的差异，而且它是恒正的，方差比 $k1$ 要小。

但是这个形式是哪里来的呢？实际上，这是通过更一般的 f 散度近似来的。f 散度可以看作是 KL 散度的一种推广，其定义为：

$$D_f(p,q)=\mathbb{E}_{x\sim q}\left[f(\frac{p(x)}{q(x)})\right]$$

其中函数 $f(\cdot)$ 需要是凸函数。可以看到，KL 散度，实际就是取了 $f(x)=-\log(x)$ 的 f 散度。而我们刚刚介绍的 $k2$，则相当于是取了 $f(x)=\frac{1}{2}(\log x)^2$，其期望也是一种 f 散度。

关于作者
欲壑难填
回答 8 文章 101 关注者 466
关注 发私信

有这样一个事实：当两个概率分布 p 和 q 非常接近时，所有（ f 可微的）f 散度在二阶近似上都会表现得非常相似，这当然也包括 KL 散度。所以，我们可以选择一个其他的 f 函数构建 f 散度，来近似 KL 散度，只要保证 p, q 比较接近时的表现相似。而要保证这一点，只需要考察二者的 $f''(1)$ ，很明显， 二者的 $f''(1)$ 都为 1。

所以，我们可以将 KL 散度近似为 $f(x) = \frac{1}{2}(\log x)^2$ 下的 f 散度。虽然这会带来一些偏差（后面实验显示，增加的偏差其实很小），但降低了估计值的方差。

> From Kimi：
> 当 p 和 q 接近时，我们可以将 $\frac{q(x)}{p(x)}$ 近似为 $1+\epsilon$，其中 ϵ 是一个小的偏差。使用泰勒展开，我们有：
>
> $$f(1+\epsilon) \approx f(1) + f'(1)\epsilon + \frac{1}{2}f''(1)\epsilon^2$$
>
> 由于 $f(1) = 0$ （根据 f-散度的定义），并且 $\mathbb{E}_{x\sim q}[\epsilon] = 0$ （因为 p 和 q 接近），所以 f-散度的二次近似主要取决于 $f''(1)$ 。

control variate与k3 估计

更进一步，我们能不能既要无偏，又要低方差呢？想要降低方差，通常的办法是控制变量[+]（control variate）法，即选用无偏的 $k1$ ，但是需要再加上一些期望为 0 ，且与 $k1$ 负相关的项，从而在保证无偏的同时，降低方差。很巧的是，在这里 $r - 1 = \frac{p(x)}{q(x)} - 1$ 就是一个期望为零的项（推导如下）。

$$\begin{aligned}\mathbb{E}_q[r-1] &= \mathbb{E}_q\left[\frac{p(x)}{q(x)} - 1\right] \\ &= \int \left[\frac{p(x)}{q(x)} - 1\right] q(x)dx \\ &= \int p(x)dx - \int q(x)dx \\ &= 1 - 1 = 0\end{aligned}$$

所以，对于任意的 λ ， $-\log r + \lambda(r-1)$ 都是一个无偏估计。这样我们就可以选择一个 λ ，使得该式的方差最小。但是由于该式依赖于 p 和 q ，所以没法直接解析求解。我们就直接考虑一个简单的选择，取 $\lambda = 1$ ，由于有 $\log(x) \le x - 1$ ，所以该式能保证是正的，已经能够尽量减小方差了。所以，我们有对 KL 散度的第三种估计：

$$k3 = (r-1) - \log r$$

我们可以将这个思想扩展到任意的 f 散度估计上，就比如 $\mathbf{KL}[p, q]$ （注意 p, q 反过来了），对它的无偏低方差估计，就可以取 $r\log r - (r-1)$ 。

实验

我们进行一个简单的实验来对比这三种 KL 散度的估计。假设 $q = N(0,1)$ ，$p = N(0.1, 1)$ ，它们真实的 KL 散度为 0.005。

```
import torch.distributions as dis
p = dis.Normal(loc=0, scale=1)
q = dis.Normal(loc=0.1, scale=1)
x = q.sample(sample_shape=(10_000_000,))
truekl = dis.kl_divergence(p, q)
print("true", truekl)
logr = p.log_prob(x) - q.log_prob(x)
k1 = -logr
k2 = logr ** 2 / 2
k3 = (logr.exp() - 1) - logr
for k in (k1, k2, k3):
    print((k.mean() -
```

	bias/true	stdev/true
k1	0	20
k2	0.002	1.42
k3	0	1.42

可以看到，k2 虽然不是无偏的，但是偏差非常小，只有 0.2%，因为此时的 p, q 两分布很接近。

我们再将 p 改为 $N(1,1)$，此时 KL 散度的真实值为 0.5。

	bias/true	stdev/true
k1	0	2
k2	0.25	1.73
k3	0	1.7

此时，k2 的偏差就非常大了，因为此时的 p, q 两分布已经不是那么接近了。而 k3 甚至比 k2 的方差还要低，所以看起来 k3 是一个全面更优的估计。

OpenRLHF 中也实现了 k1k2k3 估计方法：link。

总结

本文中我们首先介绍了 KL 散度最常用的估计 k1，但是发现它方差非常大，然后我们介绍 f 散度并设计了对 KL 散度近似的 k2 估计，k2 降低了方差但是是有偏的。为了得到无偏且低方差的估计，我们又考虑通过 control variate 构造了 k3 估计，达到了比较理想的对 KL 散度的估计。在 RL (for LLM) 中，k2、k3 都有被选用，我们需要根据实际场景分析和实验来决定选用哪种估计（比如 k2 估计要求两分布是比较接近的，才能有较低的偏差）。

编辑于 2025-05-01 16:04 · 北京

人工智能 机器学习 强化学习

欲壑难填 作者

感谢反馈。RLHF 中的 kl 我还没有仔细研究，后面会学习下。请问本文中哪一段表述有问题，我修改下。是最后一段吗?

04-30 · 北京 回复 喜欢

推荐阅读

GRPO和K1.5中的KL散度计算方式不一样?

KL散度的三种计算形式：原始的KL散度计算公式：log\frac{{\pi}_{\theta}(y_{j},z_{j}|x)}{{\pi}_{{\theta}_{i}}(y_{j},z_{j}|x)} Kimi K1.5中使用的KL散度计算公式：...

沫成

【解构云原生】K8s踩坑：Ingress四层负载均衡端口不...

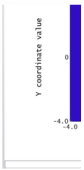
背景知识ingress原生是仅支持七层负载均衡（基于路径）的，其中ingress-nginx通过configmap的方式也能做到四层（基于端口）的负载均衡。在ingress-nginx 0.21.0版本中，作者原计划要移除对...

网易数帆 发表于网易云基础...

问题：K3曲面可被曲线的乘积支配吗?

编号：1 在问题系列里，我会列举代数几何方面一些未解决的问题，太有名的问题除外。 这些大多是我关心的问题，难度不一。 ————————————————...

编织者



ZEMAX 拟表面

达摩寺扫