MCMC(四)Gibbs采样

MCMC(一)蒙特卡罗方法

MCMC(二)马尔科夫链

MCMC(三)MCMC采样和M-H采样

MCMC(四)Gibbs采样

在MCMC(三)MCMC采样和M-H采样中，我们讲到了M-H采样已经可以很好的解决蒙特卡罗方法需要的任意概率分布的样本集的问题。但是M-H采样有两个缺点：一是需要计算接受率，在高维时计算量大。并且由于接受率的原因导致算法收敛时间变长。二是有些高维数据，特征的条件概率分布好求，但是特征的联合分布不好求。因此需要一个好的方法来改进M-H采样，这就是我们下面讲到的Gibbs采样。

1. 重新寻找合适的细致平稳条件

在上一篇中，我们讲到了细致平稳条件：如果非周期马尔科夫链的状态转移矩阵P P和概率分布$\pi(x)$ π(x)对于所有的i,j i,j满足：

$$\pi(i)P(i,j) = \pi(j)P(j,i)$$

π(i)P(i,j)=π(j)P(j,i)

则称概率分布$\pi(x)$ π(x)是状态转移矩阵P P的平稳分布。

在M-H采样中我们通过引入接受率使细致平稳条件满足。现在我们换一个思路。

从二维的数据分布开始，假设$\pi(x_1,x_2)$ π(x1,x2)是一个二维联合数据分布，观察第一个特征维度相同的两个点$A(x_1^{(1)},x_2^{(1)})$ A(x1(1),x2(1))和$B(x_1^{(1)},x_2^{(2)})$ B(x1(1),x2(2))，容易发现下面两式成立：

$$\pi(x_1^{(1)},x_2^{(1)})\pi(x_2^{(2)}|x_1^{(1)}) = \pi(x_1^{(1)})\pi(x_2^{(1)}|x_1^{(1)})\pi(x_2^{(2)}|x_1^{(1)})$$

π(x1(1),x2(1))π(x2(2)|x1(1))=π(x1(1))π(x2(1)|x1(1))π(x2(2)|x1(1))

$$\pi(x_1^{(1)},x_2^{(2)})\pi(x_2^{(1)}|x_1^{(1)}) = \pi(x_1^{(1)})\pi(x_2^{(2)}|x_1^{(1)})\pi(x_2^{(1)}|x_1^{(1)})$$

π(x1(1),x2(2))π(x2(1)|x1(1))=π(x1(1))π(x2(2)|x1(1))π(x2(1)|x1(1))

由于两式的右边相等，因此我们有：

$$\pi(x_1^{(1)},x_2^{(1)})\pi(x_2^{(2)}|x_1^{(1)}) = \pi(x_1^{(1)},x_2^{(2)})\pi(x_2^{(1)}|x_1^{(1)})$$

π(x1(1),x2(1))π(x2(2)|x1(1))=π(x1(1),x2(2))π(x2(1)|x1(1))

也就是：

$$\pi(A)\pi(x_2^{(2)}|x_1^{(1)}) = \pi(B)\pi(x_2^{(1)}|x_1^{(1)})$$

π(A)π(x2(2)|x1(1))=π(B)π(x2(1)|x1(1))

观察上式再观察细致平稳条件的公式，我们发现在$x_1 = x_1^{(1)}$ x1=x1(1)这条直线上，如果用条件概率分布$\pi(x_2|x_1^{(1)})$ π(x2|x1(1))作为马尔科夫链的状态转移概率，则任意两个点之间的转移满足细致平稳条件！这真是一个开心的发现，同样的道理，在在$x_2 = x_2^{(1)}$ x2=x2(1)这条直线上，如果用条件概率分布$\pi(x_1|x_2^{(1)})$ π(x1|x2(1))作为马尔科夫链的状态转移概率，则任意两个点之间的转移也满足细致平稳条件。那是因为假如有一点$C(x_1^{(2)},x_2^{(1)})$ C(x1(2),x2(1)),我们可以得到：

$$\pi(A)\pi(x_1^{(2)}|x_2^{(1)}) = \pi(C)\pi(x_1^{(1)}|x_2^{(1)})$$
π(A)π(x1(2)|x2(1))=π(C)π(x1(1)|x2(1))

基于上面的发现，我们可以这样构造分布$\pi(x_1, x_2)$π(x1,x2)的马尔可夫链对应的状态转移矩阵P P：

$$P(A \to B) = \pi(x_2^{(B)}|x_1^{(1)}) \ \text{if} \ x_1^{(A)} = x_1^{(B)} = x_1^{(1)}$$
P(A→B)=π(x2(B)|x1(1))ifx1(A)=x1(B)=x1(1)
$$P(A \to C) = \pi(x_1^{(C)}|x_2^{(1)}) \ \text{if} \ x_2^{(A)} = x_2^{(C)} = x_2^{(1)}$$
P(A→C)=π(x1(C)|x2(1))ifx2(A)=x2(C)=x2(1)
$$P(A \to D) = 0 \ \text{else}$$
P(A→D)=0else

有了上面这个状态转移矩阵，我们很容易验证平面上的任意两点E, F E,F，满足细致平稳条件：

$$\pi(E)P(E \to F) = \pi(F)P(F \to E)$$
π(E)P(E→F)=π(F)P(F→E)

2. 二维Gibbs采样

利用上一节找到的状态转移矩阵，我们就得到了二维Gibbs采样，这个采样需要两个维度之间的条件概率。具体过程如下：

1）输入平稳分布$\pi(x_1, x_2)$π(x1,x2)，设定状态转移次数阈值n_1n1，需要的样本个数n_2n2

2）随机初始化初始状态值$x_1^{(0)}$x1(0)和$x_2^{(0)}$x2(0)

3）for $t = 0$t=0 to $n_1 + n_2 - 1$n1+n2−1:

a) 从条件概率分布$P(x_2|x_1^{(t)})$ P(x2|x1(t))中采样得到样本x_2^{t+1}x2t+1

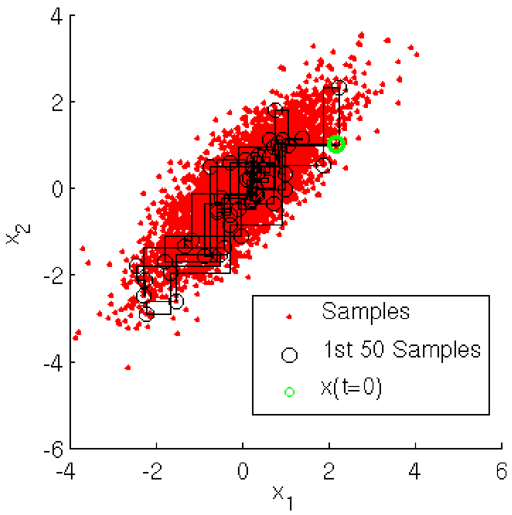
b) 从条件概率分布$P(x_1|x_2^{(t+1)})$ P(x1|x2(t+1))中采样得到样本x_1^{t+1}x1t+1

样本集$\{(x_1^{(n_1)}, x_2^{(n_1)}), (x_1^{(n_1+1)}, x_2^{(n_1+1)}), \ldots, (x_1^{(n_1+n_2-1)}, x_2^{(n_1+n_2-1)})\}$ {(x1(n1),x2(n1)),(x1(n1+1),x2(n1+1)),...,(x1(n1+n2−1),x2(n1+n2−1))}即为我们需要的平稳分布对应的样本集。

整个采样过程中，我们通过轮换坐标轴，采样的过程为：

$$(x_1^{(1)}, x_2^{(1)}) \to (x_1^{(1)}, x_2^{(2)}) \to (x_1^{(2)}, x_2^{(2)}) \to \ldots \to (x_1^{(n_1+n_2-1)}, x_2^{(n_1+n_2-1)})$$
(x1(1),x2(1))→(x1(1),x2(2))→(x1(2),x2(2))→...→(x1(n1+n2−1),x2(n1+n2−1))

用下图可以很直观的看出，采样是在两个坐标轴上不停的轮换的。当然，坐标轴轮换不是必须的，我们也可以每次随机选择一个坐标轴进行采样。不过常用的Gibbs采样的实现都是基于坐标轴轮换的。



3. 多维Gibbs采样

上面的这个算法推广到多维的时候也是成立的。比如一个n维的概率分布$\pi(x_1, x_2, ...x_n)$ π(x1,x2,...xn)，我们可以通过在n个坐标轴上轮换采样，来得到新的样本。对于轮换到的任意一个坐标轴$x_i$xi上的转移，马尔科夫链的状态转移概率为$P(x_i|x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n)$ P(xi|x1,x2,...,xi−1,xi+1,...,xn)，即固定$n-1$n−1个坐标轴，在某一个坐标轴上移动。

具体的算法过程如下：

1）输入平稳分布$\pi(x_1, x_2，..., x_n)$π(x1,x2，...,xn)或者对应的所有特征的条件概率分布，设定状态转移次数阈值n_1n1，需要的样本个数n_2n2

2）随机初始化初始状态值$(x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}$ (x1(0),x2(0),...,xn(0)

3）for $t=0$t=0 to n_1+n_2-1n1+n2−1:

a) 从条件概率分布$P(x_1|x_2^{(t)}, x_3^{(t)}, ..., x_n^{(t)})$ P(x1|x2(t),x3(t),...,xn(t))中采样得到样本x_1^{t+1}x1t+1

b) 从条件概率分布$P(x_2|x_1^{(t+1)}, x_3^{(t)}, x_4^{(t)}, ..., x_n^{(t)})$ P(x2|x1(t+1),x3(t),x4(t),...,xn(t))中采样得到样本x_2^{t+1}x2t+1

c)...

d) 从条件概率分布$P(x_j|x_1^{(t+1)}, x_2^{(t+1)}, ..., x_{j-1}^{(t+1)}, x_{j+1}^{(t)}..., x_n^{(t)})$ P(xj|x1(t+1),x2(t+1),...,xj−1(t+1),xj+1(t)...,xn(t))中采样得到样本x_j^{t+1}xjt+1

e)...

f) 从条件概率分布$P(x_n|x_1^{(t+1)}, x_2^{(t+1)}, ..., x_{n-1}^{(t+1)})$ P(xn|x1(t+1),x2(t+1),...,xn−1(t+1))中采样得到样本x_n^{t+1}xnt+1

样本集$\{(x_1^{(n_1)}, x_2^{(n_1)}, ..., x_n^{(n_1)}), ..., (x_1^{(n_1+n_2-1)}, x_2^{(n_1+n_2-1)}, ..., x_n^{(n_1+n_2-1)})\}$ {(x1(n1),x2(n1),...,xn(n1)),..., (x1(n1+n2−1),x2(n1+n2−1),...,xn(n1+n2−1))}即为我们需要的平稳分布对应的样本集。

整个采样过程和Lasso回归的坐标轴下降法算法非常类似，只不过Lasso回归是固定$n-1$n−1个特征，对某一个特征求极值。而Gibbs采样是固定$n-1$n−1个特征在某一个特征采样。

同样的，轮换坐标轴不是必须的，我们可以随机选择某一个坐标轴进行状态转移，只不过常用的Gibbs采样的实现都是基于坐标轴轮换的。

4. 二维Gibbs采样实例

这里给出一个Gibbs采样的例子。完整代码参见我的github: https://github.com/ljpzzz/machinelearning/blob/master/mathematics/mcmc_3_4.ipynb

假设我们要采样的是一个二维正态分布$Norm(\mu, \Sigma)$ Norm(μ,Σ),其中：

$$\mu = (\mu_1, \mu_2) = (5, -1)$$

μ=(μ1,μ2)=(5,−1)

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$$

Σ=(σ12ρσ1σ2ρσ1σ2σ22)=(1114)

而采样过程中的需要的状态转移条件分布为：

$$P(x_1|x_2) = Norm\left(\mu_1 + \rho\sigma_1/\sigma_2 (x_2 - \mu_2), (1 - \rho^2)\sigma_1^2\right)$$

P(x1|x2)=Norm(μ1+ρσ1/σ2(x2−μ2),(1−ρ2)σ12)

$$P(x_2|x_1) = Norm\left(\mu_2 + \rho\sigma_2/\sigma_1 (x_1 - \mu_1), (1 - \rho^2)\sigma_2^2\right)$$

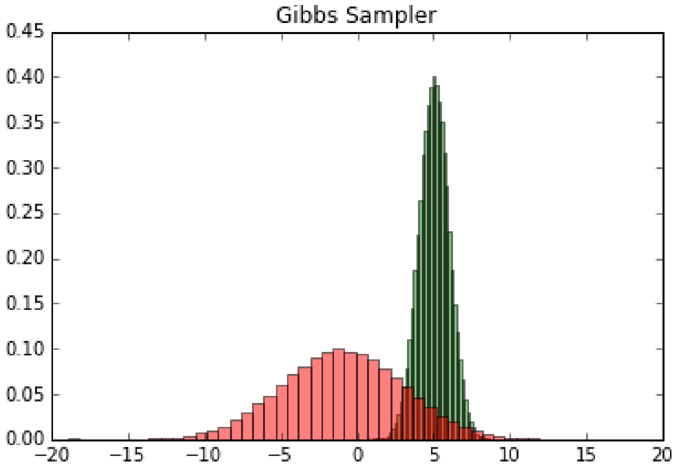
P(x2|x1)=Norm(μ2+ρσ2/σ1(x1−μ1),(1−ρ2)σ22)

具体的代码如下：

```
from mpl_toolkits.mplot3d import Axes3D
from scipy.stats import multivariate_normal
samplesource = multivariate_normal(mean=[5,-1], cov=[[1,0.5],[0.5,2]])

def p_ygivenx(x, m1, m2, s1, s2):
    return (random.normalvariate(m2 + rho * s2 / s1 * (x - m1), math.sqrt(1 - rho ** 2) * s2))

def p_xgiveny(y, m1, m2, s1, s2):
    return (random.normalvariate(m1 + rho * s1 / s2 * (y - m2), math.sqrt(1 - rho ** 2) * s1))

N = 5000
K = 20
x_res = []
y_res = []
z_res = []
m1 = 5
m2 = -1
s1 = 1
s2 = 2

rho = 0.5
y = m2

for i in xrange(N):
    for j in xrange(K):
        x = p_xgiveny(y, m1, m2, s1, s2)
        y = p_ygivenx(x, m1, m2, s1, s2)
        z = samplesource.pdf([x,y])
        x_res.append(x)
        y_res.append(y)
        z_res.append(z)
```

```
num_bins = 50
plt.hist(x_res, num_bins, normed=1, facecolor='green', alpha=0.5)
plt.hist(y_res, num_bins, normed=1, facecolor='red', alpha=0.5)
plt.title('Histogram')
plt.show()
```

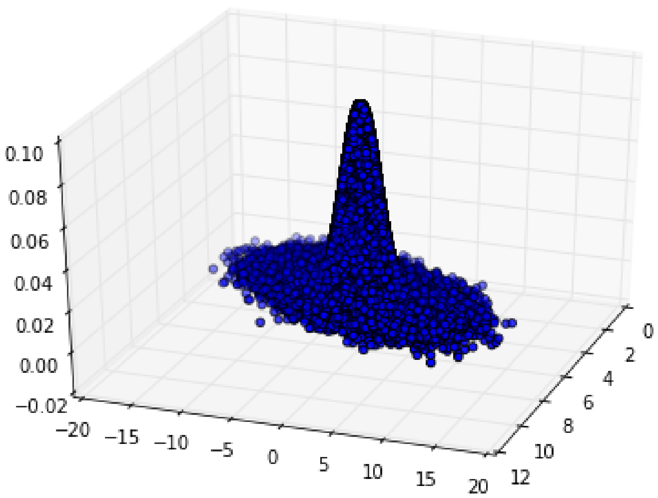
输出的两个特征各自的分布如下:



然后我们看看样本集生成的二维正态分布，代码如下:

```
fig = plt.figure()
ax = Axes3D(fig, rect=[0, 0, 1, 1], elev=30, azim=20)
ax.scatter(x_res, y_res, z_res,marker='o')
plt.show()
```

输出的正态分布图如下:



5. Gibbs采样小结

由于Gibbs采样在高维特征时的优势，目前我们通常意义上的MCMC采样都是用的Gibbs采样。当然Gibbs采样是从M-H采样的基础上的进化而来的，同时Gibbs采样要求数据至少有两个维度，一维概率分布的采样是没法用Gibbs采样的,这时M-H采样仍然成立。

有了Gibbs采样来获取概率分布的样本集，有了蒙特卡罗方法来用样本集模拟求和，他们一起就奠定了MCMC算法在大数据时代高维数据模拟求和时的作用。MCMC系列就在这里结束吧。

（欢迎转载，转载请注明出处。欢迎沟通交流： liujianping-ok@163.com）

« 上一篇：MCMC(三)MCMC采样和M-H采样
» 下一篇：文本挖掘的分词原理

评论列表

#1楼 2017-07-31 15:58 jingchunzhen
请问博主在二维Gibbs采样实例中，为什么设置了双重for循环for i in xrange(N):
for j in xrange(K):，这样做的意义是什么呢？参照伪代码是for i in xrange(n1+n2):
支持(0)反对(0)

#2楼[楼主] 2017-07-31 16:19 刘建平Pinard
@ jingchunzhen

你好，这里代码里的用法的确和算法描述稍有不同：

在代码里外层循环是得到N个样本，对于每个样本，我们要在内层循环中分别进行马尔科夫状态转移K次得到对应的样本。这样的好处是得到的样本独立性好一点，毕竟我们的条件概率计算量不大。

在算法里描述的的确如你所说，先进行若干次马尔科夫状态转移，到了我们认为已经收敛到目标分布时，就进行连续的采样，得到N个样本。这样的好处是采样的计算量小很多，尤其是对于复杂分布的计算，一次马尔科夫过程后就可以得到所有想要个数的样本，但是样本之间的独立性会稍差。
支持(5)反对(0)

#3楼 2017-12-29 18:35 忆梦涟
你好，请问状态转移次数阈值，也就是马尔科夫链达到稳定的转移次数n1，这个值如何确定，
支持(0)反对(0)

#4楼[楼主] 2018-01-02 11:03 刘建平Pinard
@ 忆梦涟
你好，实际应用中都是排脑袋拍出来的，如果计算能力足够，n1可以比较大一些，这样保险，否则就没有办法了，甚至可能只能取n1为个位数，当然这样有可能马尔科夫链并没有真正得到稳态，但是也算是一种近似了。
支持(0)反对(0)

#5楼 2018-01-11 09:25 CuiPeng
你好，我看到LDA中有用gibbs采样训练参数，我看文章中只讲到gibbs采样是用来生成稳定的指定分布的数据，请问怎样用gibbs采样去更新参数呢？比如我给定了一组参数，如何根据样本去训练参数呢？
支持(0)反对(0)

#6楼[楼主] 2018-01-11 11:45 刘建平Pinard
@ CuiPeng
你好，Gibbs采样只是一个数学方法，不是一个机器学习算法，所以本身不能帮你训练模型参数，更新模型参数。
在LDA中，Gibbs采样也只是为了采样分布数据然后做近似计算的，也就是为LDA求解服务的。
支持(1)反对(0)

#7楼 2018-04-15 11:21 小小老实人
楼主你好，文章写得很好，对我帮助很大。一个无关紧要得问题，请问算法里随机初始化状态值，角标是不是该为0呢?
支持(0)反对(0)

#8楼[楼主] 2018-04-16 20:35 刘建平Pinard
@ 小小老实人
你好，那里的确是我写错，已经修改，感谢指出错误。
支持(0)反对(0)

#9楼 2018-05-12 14:06 推荐算法萌新
博主您好！看了这篇文章我有一点疑问:
1、在每一节里面，采样过程的第一步都写了，输入平稳分布π(x)，那么这个平稳分布在采样过程中的作用是什么?
支持(0)反对(0)

#10楼 2018-05-12 14:15 推荐算法萌新
或者我是这样理解，多维Gibbs采样时，不管初始状态如何，只要依据当前状态下计算的单轴条件概率进行转移，最终都会收敛，然后的得到一个分布P，而这个分布P，是满足那个细致平稳条件的，所以分布P是一个平稳分布。
支持(0)反对(0)

#11楼[楼主] 2018-05-12 22:11 刘建平Pinard
@ 推荐算法萌新
你好，是的，这个平稳分布就是我们需要采样的分布，而Gibbs采样就是帮助我们采样得到改分布样本的方法。
支持(0)反对(0)

#12楼 2018-06-09 22:29 zhuyunxiu
有个问题想请教一下，样本是根据任意的Q概率分布采样得到的，然后是Q开始改变了对吧，变成了Q（x*|xt），再根据这个概率进行采样的。那到底这个概率是什么呢？我看到您的实例中用高斯分布先采一个，再用采到的样品作为新的高斯分布的均值进行采样，那要是不是高斯分布，假设是多项分布作为起始采样概率，那这个Q（x*|xt）该怎么求?
还有，同样的，吉布斯中的条件概率如果不是高斯要怎么进行采样?
支持(0)反对(0)

#13楼[楼主] 2018-06-10 12:43 刘建平Pinard
@ zhuyunxiu
你好，我们选择Q的原则肯定是好采样的常用概率分布。Q的联合分布和条件分布函数也都是已知很好采样的概率分布。所以你说的Q（x*|xt）也是一个已知的好采样的概率分布。
回到你说的：“假设是多项分布作为起始采样概率，那这个Q（x*|xt）该怎么求？”，选择Q的原则是好采样的条件概率。所以如果你这个Q（x*|xt）不好求，那可以换一个Q了。
最后一个问题，“条件概率如果不是高斯要怎么进行采样？”，Gibbs采样比较特殊，它是基于坐标轴轮转迭代的，所以需要任意一个坐标轴基于其他所有坐标轴的条件概率，这个条件概率如果不是常见的好采样的分布，那么也是不好采样求解的。也就是说Gibbs采样也不是万能的。
支持(2)反对(0)

#14楼 2018-08-19 12:45 Lszx
老师您好，看了您的文章受益匪浅
请问:
a) 从条件概率分布 P(x2|x1(t))中采样得到样本 x2 (t+1)
b) 从条件概率分布P(x1|x2(t+1))中采样得到样本 x1 (t+1)
根据原理 任意两点E F之间的转移 都服从细致平稳分布
这里的第二步采样 为什么不是从P(x1|x2(t))中采样得到样本 x1 (t+1)
支持(0)反对(0)

#15楼[楼主] 2018-08-19 23:02 刘建平Pinard
@ Lszx
你好，你这样采样也是可以的。我采用坐标轴轮换只是为了符合大多数Gibbs采样算法使用的的思路。第三节的末尾也讲到了原因:
用下图可以很直观的看出，采样是在两个坐标轴上不停的轮换的。当然，坐标轴轮换不是必须的，我们也可以每次随机选择一个坐标轴进行采样。不过常用的Gibbs采样的实现都是基于坐标轴轮换的。
支持(3)反对(0)

#16楼 2018-08-21 11:12 Lszx
@ 刘建平Pinard
老师 我有个问题想请您指导一下 具体发您163邮箱了
期待您百忙之中的回复
支持(0)反对(0)

#17楼 2018-09-13 10:03 Process520
老师，您好，关于下面两步算法描述中，有三个疑问:
a) 从条件概率分布 P(x2|x1(t))中采样得到样本 x2 (t+1)
b) 从条件概率分布P(x1|x2(t+1))中采样得到样本 x1 (t+1)
1. 这里说“得到样本x2(t+1)”,我理解是得到样本为(x1(t), x2(t+1)),
x2(t+1)是它在属性x2处的取值，这样理解对吗?
2.如果上面对的话，那我们取的样本(x1(t+1), x2(t+1))是由(x1(t), x2(t+1))转移来的吗?
3.如果第二问对的话，能把(x1(t), x2(t+1))作为一个采样样本吗？不能或不好的原因是因为样本(x1(t+1), x2(t+1))与(x1(t), x2(t))更加独立吗?
支持(0)反对(0)

#18楼[楼主] 2018-09-13 10:25 刘建平Pinard
@ Process520
你好!
1） 的确这里采样得到的是某一个维度的数据，而不是整个数据。
2） 是的，这就是坐标轴轮转。参看第二节图中的黑色采样路径。
3） 也是可以的，但是如你所说，我们期望两个样本的独立性稍微强一些，所以一般不会把相邻采样的数据的弄的某个维度总是一样。
支持(0)反对(0)

#19楼 2018-09-13 11:25 Process520
@ 刘建平Pinard
明白了，感谢您的解答。
祝您生活愉快，身体健康。
支持(0)反对(0)

#20楼 2018-09-19 16:38 aaronwang123
你好，博主:

“MCMC算法在大数据时代高维数据模拟求和时的作用。"

看了这个mcmc系列后，想问下，我的理解是否是对的?

1.整个系列从求不规则面积出发。引出了不同概率分布如何采样问题。
2.那么我们最后有价值的东西是，得到了如何正确采样的方法论。而不是求和问题的解决方法。
3.请问我们大数据处理，为什么会有采样问题
支持(0)反对(0)

#21楼[楼主] 2018-09-20 13:41 刘建平Pinard
@ aaronwang123
你好!

1. 这个没问题，其实最初的原理就是第一篇的接受-拒绝采样。

2. 有价值的东西的确是正确采样的方法论，当有了这个方法后，一些复杂的问题比如基于概率分布的积分求和问题可以使用采样的方法近似求解。

3. 主要是有些计算式太复杂，比如包含复杂的概率分布积分等，没有办法直接求得代数解，但是我们又需要这个结果，那么可以通过采样的方法得到对应分布样本，然后求对应的近似解。

比如在LDA主题模型中，我们可以用MCMC来采样求主题分布和词分布的近似解。
文本主题模型之LDA(二) LDA求解之Gibbs采样算法
https://www.cnblogs.com/pinard/p/6867828.html

此外还有分解机模型里的积分求解。
分解机(Factorization Machines)推荐算法原理
https://www.cnblogs.com/pinard/p/6370127.html
支持(0)反对(0)

#22楼 2018-10-06 18:05 hui0618
博主你好, 请问状态转移条件是如何得到的? 看表达式我的理解是, 当前采集到的x1如果偏移其均值较大, 那么当前采集到的x1对应其小概率分布, 于是调整x2所属的正态分布的均值和方差, 使得采样x2时也会偏移x2的均值更可能得到概率密度较小的x2样本.
支持(0)反对(0)
刷新评论刷新页面返回顶部
注册用户登录后才能发表评论，请 登录 或 注册，访问网站首页。

最新IT新闻:
· SwiftKey Android版即将支持Windows 10云剪贴板
· 福布斯专访贝索斯：他的字典里没有“边界”这个词