

null的专栏

Keep your eyes open and your feet moving forward. You'll find what you need.

目录视图摘要视图RSS 订阅

个人声明

欢迎大家加群，探讨与机器学习相关技术相关的话题：

101620539

学习网站：www.wanwenonline.cn
博客的主要内容主要是自己的学习
笔记，并结合个人的理解，供各位在
学习过程中参考，如有疑问，欢迎提出
；若有侵权，请告知博主删除，原创
文章转载还请注明出处。

我写的书：



购买链接：
[京东-Python机器学习算法](#)

个人资料



zhiyong_will
关注发私信



访问：1292795次
积分：10841
等级：7
排名：第1697名

图灵赠书——程序员11月书单【思考】Python这么厉害的原因竟然是！感恩节赠书：《深度学习》等异步社区优秀图书和作译者评选启动！每周荐书：京东架构、Linux内核、Python全栈

简单易学的机器学习算法——马尔可夫链蒙特卡罗方法MCMC

标签：采样 MCMC

2016-06-29 18:4926152人阅读评论(0)收藏举报

分类：
Machine Learning (57)

版权声明：本文为博主原创文章，未经博主允许不得转载。

目录(?)

对于一般的分布的采样，在很多的编程语言中都有实现，如最基本的满足均匀分布的随机数，但是对于复杂的分布，要想对其采样，却没有实现好的函数，在这里，可以使用马尔可夫链蒙特卡罗(Markov Chain Monte Carlo, MCMC)方法，其中Metropolis-Hastings采样和Gibbs采样是MCMC中使用较为广泛的两种形式。

MCMC的基础理论为马尔可夫过程，在MCMC算法中，为了在一个指定的分布上采样，根据马尔可夫过程，首先从任一状态出发，模拟马尔可夫过程，不断进行状态转移，最终收敛到平稳分布。

一、马尔可夫链

1、马尔可夫链

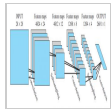
设 X_t 表示随机变量 X 在离散时间 t 时刻的取值。若该变量随时间变化的转移概率仅仅依赖于它的当前取值，即

$$P\left(X_{t+1}=s_j\mid X_0=s_0,X_1=s_1,\cdots,X_t=s_i\right)=P\left(X_{t+1}=s_j\mid X_t=s_i\right)$$

也就是说状态转移的概率只依赖于前一个状态。称这个变量为马尔可夫变量，其中 $s_0,s_1,\cdots,s_i,s_j\in\Omega$ 为随机变量 X 可能的状态。这个性质称为马尔可夫性质，具有马尔可夫性质的随机过程称为马尔可夫过程。

马尔可夫链指的是在一段时间内随机变量 X 的取值序列 (X_0,X_1,\cdots,X_m) ，它们满足如上的马尔可夫性质。

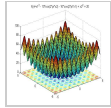
原创：155篇 转载：1篇 博客专栏
译文：1篇 评论：566条



深度学习Deep Learning

文章：10篇

阅读：45586



优化算法

文章：14篇

阅读：145260



机器学习，数据挖掘算法

文章：42篇

阅读：585551

文章分类

Machine Learning (58)

Deep Learning (11)

Optimization Algorithm (16)

论文与材料的学习笔记 (37)

Computational Advertising (3)

NLP (3)

Recommender System (6)

算法 (Algorithm) 算法题 (12)

Data Structure & Algorithm (6)

设计模式 (1)

Hadoop (4)

Python (13)

C/C++ (8)

Matlab技巧 (2)

Spark (1)

Linux (3)

PHP (2)

Practice in Job (1)

Web Spider (1)

联系我

Email: zhaozhong1989@126.com

2、转移概率

马尔可夫链是通过对应的转移概率定义的，转移概率指的是随机变量从一个时刻到下一个时刻，从状态 s_i 转移到另一个状态 s_j 的概率，即：

$$P(i \rightarrow j) := P_{i,j} = P(X_{t+1} = s_j | X_t = s_i)$$

记 $\pi_k^{(t)}$ 表示随机变量 X 在时刻 t 的取值为 s_k 的概率，则随机变量 X 在时刻 $t+1$ 的取值为 s_i 的概率为：

$$\begin{aligned}\pi_i^{(t+1)} &= P(X_{t+1} = s_i) \\ &= \sum_k P(X_{t+1} = s_i | X_t = s_k) \cdot P(X_t = s_k) \\ &= \sum_k P_{k,i} \cdot \pi_k^{(t)}\end{aligned}$$

假设状态的数目为 n ，则有：

$$\begin{pmatrix} \pi_1^{(t+1)} \\ \vdots \\ \pi_n^{(t+1)} \end{pmatrix} = \begin{pmatrix} \pi_1^{(t)} \\ \vdots \\ \pi_n^{(t)} \end{pmatrix} \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,n} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,n} \\ \vdots & \vdots & & \vdots \\ P_{n,1} & P_{n,2} & \cdots & P_{n,n} \end{bmatrix}$$

3、马尔可夫链的平稳分布

对于马尔可夫链，需要注意以下的两点：

- 1、周期性：即经过有限次的状态转移，又回到了自身；
- 2、不可约：即两个状态之间相互转移；

如果一个马尔可夫过程既没有周期性，又不可约，则称为各态遍历的。

对于一个各态遍历的马尔可夫过程，无论初始值 $\pi^{(0)}$ 取何值，随着转移次数的增多，随机变量的取值分布最终都会收敛到唯一的平稳分布 π^* ，即：

$$\lim_{t \rightarrow \infty} \pi^{(0)} \mathbf{P}^t = \pi^*$$

且这个平稳分布 π^* 满足：

$$\pi^* \mathbf{P} = \pi^*$$

其中， $\mathbf{P} = (p_{i,j})_{n \times n}$ 为转移概率矩阵。

二、马尔可夫链蒙特卡罗方法

微博

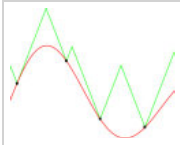


赵志勇 北京 海淀区

加关注

#松鼠存储# // @LR机器学习计算机视觉:// @bit_yangxm: Repost

爱可可-爱生活 : 【新版dlib全局优化算法简介】《A Global Optimization Algorithm Worth Using》<http://t.cn/RH0sQY5>



今天 05:19

转发 | 评论

60分钟前

转发(1) | 评论

<http://t.cn/RMvbyPf>

今天 03:21

转发 | 评论

#松鼠存储#

龙星镖局 : 这个页面有计算机领域历届会议的best paper汇总, 假期有兴趣的同学可以读一读, 没空的也可以存下来收藏一下。👍
<http://t.cn/h4nrwN>



文章存档

2017年10月 (1)

2017年09月 (1)

2017年08月 (5)

2017年07月 (1)

2017年06月 (4)

展开

阅读排行

简单易学的机器学习算法——... (60233)

python——时间与时间戳之间... (54045)

简单易学的机器学习算法——... (51015)

简单易学的机器学习算法——... (47421)

数据处理——One-Hot Encodi... (44023)

Python技巧——list与字符串... (35343)

简单易学的机器学习算法——... (32878)

1、基本思想

对于一个给定的概率分布 $P(X)$, 若是要得到其样本, 通过上述的马尔可夫链的概念, 我们可以构造一个转移矩阵为 \mathbf{P} 的马尔可夫链, 使得该马尔可夫链的平稳分布为 $P(X)$, 这样, 无论其初始状态为何值, 假设记为 x_0 , 那么随着马尔可夫过程的转移, 得到了一系列的状态值, 如: $x_0, x_1, x_2, \dots, x_n, x_{n+1}, \dots$, 如果这个马尔可夫过程在第 n 步时已经收敛, 那么分布 $P(X)$ 的样本即为 x_n, x_{n+1}, \dots 。

2、细致平稳条件

对于一个各态遍历的马尔可夫过程, 若其转移矩阵为 \mathbf{P} , 分布为 $\pi(x)$, 若满

$$\pi(i) P_{i,j} = \pi(j) P_{j,i}$$

则 $\pi(x)$ 是马尔可夫链的平稳分布, 上式称为细致平稳条件。

3、Metropolis采样算法

Metropolis采样算法是最基本的基于MCMC的采样算法。

3.1、Metropolis采样算法的基本原理

假设需要从目标概率密度函数 $p(\theta)$ 中进行采样, 同时, θ 满足 $-\infty < \theta < \infty$ 。Metropolis采样算法根据马尔可夫链去生成一个序列:

$$\theta^{(1)} \rightarrow \theta^{(2)} \rightarrow \dots \theta^{(t)} \rightarrow$$

其中, $\theta^{(t)}$ 表示的是马尔可夫链在第 t 代时的状态。

在Metropolis采样算法的过程中, 首先初始化状态值 $\theta^{(1)}$, 然后利用一个已知的分布 $q(\theta | \theta^{(t-1)})$ 生成一个新的候选状态 $\theta^{(*)}$, 随后根据一定的概率选择接受这个新值, 或者拒绝这个新值, 在Metropolis采样算法中, 概率为:

$$\alpha = \min \left(1, \frac{p(\theta^{(*)})}{p(\theta^{(t-1)})} \right)$$

这样的过程一直持续到采样过程的收敛, 当收敛以后, 样本 $\theta^{(t)}$ 即为目标分布 $p(\theta)$ 中的样本。

3.2、Metropolis采样算法的流程

基于以上的分析, 可以总结出如下的Metropolis采样算法的流程:

- 初始化时间 $t = 1$
- 设置 u 的值, 并初始化初始状态 $\theta^{(t)} = u$

简单易学的机器学习算法——...	(28058)
简单易学的机器学习算法——...	(26135)
简单易学的机器学习算法——...	(24586)

评论排行

简单易学的机器学习算法——...	(65)
简单易学的机器学习算法——...	(33)
推荐算法——基于矩阵分解的...	(29)
优化算法——人工蜂群算法(A...	(28)
简单易学的机器学习算法——...	(26)
社团划分——Fast Unfolding...	(23)
简单易学的机器学习算法——...	(17)
简单易学的机器学习算法——...	(16)
简单易学的机器学习算法——...	(15)
机器学习算法实践——K-Mea...	(15)

最新评论

简单易学的机器学习算法——Softmax Re...
火星十一郎 :@szp686886:阿里

《Python机器学习算法》的写作历程
drxueer :正在仔细专研您的教材,很是受用。感谢!

简单易学的机器学习算法——岭回归(Ridg...
weixin_38291590 :你好,请问未能显示的图片能更新一下吗?

简单易学的机器学习算法——集成方法(E...
微风好飞行 :好文章,但是公式都显示为"E quation Quota Exceed!"了,也许直接换成图片更好。

优化算法——粒子群算法(PSO)
zhiyong_will :@Youyusi12138:对于不同的应用函数,设置不同的参数能够得到不同的效果

优化算法——粒子群算法(PSO)
weixin_41302024 :感谢po主分享实例,非常有用

简单易学的机器学习算法——梯度提升决...
iszhuangsha :您好。梯度下降法每一步的不长不是固定的。事实上,步长是逐渐减小的。可以由wolfe conditio...

优化算法——人工蜂群算法(ABC)
Awersome :楼主您好,最近在看这个人工蜂群算法,从网上找了一个人工蜂群的源码,java版的。因为学姐用的是jav...

剑指Offer——编程题的Java实现
shihengli2010 :第11题有问题啊...

深度学习算法原理——稀疏自编码器
zhiyong_will :@zhangyukun230:先去理解其基本原理,谢谢

其他人的博客

Rachel Zhang的专栏

结构之法 算法之道

美团点评技术团队

• 重复一下的过程：

- 令 $t = t + 1$
- 从已知分布 $q(\theta | \theta^{(t-1)})$ 中生成一个候选状态 $\theta^{(*)}$
- 计算接受的概率： $\alpha = \min \left(1, \frac{p(\theta^{(*)})}{p(\theta^{(t-1)})} \right)$
- 从均匀分布 $Uniform(0, 1)$ 生成一个随机值 a
- 如果 $a \leq \alpha$, 接受新生成的值： $\theta^{(t)} = \theta^{(*)}$; 否则： $\theta^{(t)} = \theta^{(t-1)}$

• 直到 $t = T$

3.3、Metropolis算法的解释

要证明Metropolis采样算法的正确性,最重要的是要证明构造的马尔可夫过程满足如上的细致平稳条件,即:

$$\pi(i) P_{i,j} = \pi(j) P_{j,i}$$

对于上面所述的过程,分布为 $p(\theta)$, 从状态 i 转移到状态 j 的转移概率为:

$$P_{i,j} = \alpha_{i,j} \cdot Q_{i,j}$$

其中, $Q_{i,j}$ 为上述已知的分布。

对于选择该已知的分布,在Metropolis采样算法中,要求该已知的分布必须是对称的,即 $Q_{i,j} = Q_{j,i}$, 即

$$q(\theta = \theta^{(t)} | \theta^{(t-1)}) = q(\theta = \theta^{(t-1)} | \theta^{(t)})$$

常用的符合对称的分布主要有:正态分布,柯西分布以及均匀分布等。

接下来,需要证明在Metropolis采样算法中构造的马尔可夫链满足细致平稳条件。

$$\begin{aligned} p(\theta^{(i)}) P_{i,j} &= p(\theta^{(i)}) \cdot \alpha_{i,j} \cdot Q_{i,j} \\ &= p(\theta^{(i)}) \cdot \min \left\{ 1, \frac{p(\theta^{(j)})}{p(\theta^{(i)})} \right\} \cdot Q_{i,j} \\ &= \min \left\{ p(\theta^{(i)}) Q_{i,j}, p(\theta^{(j)}) Q_{i,j} \right\} \\ &= p(\theta^{(j)}) \cdot \min \left\{ \frac{p(\theta^{(i)})}{p(\theta^{(j)})}, 1 \right\} \cdot Q_{j,i} \\ &= p(\theta^{(j)}) \cdot \alpha_{j,i} \cdot Q_{j,i} \\ &= p(\theta^{(j)}) P_{j,i} \end{aligned}$$

因此，通过以上的方法构造出来的马尔可夫链是满足细致平稳条件的。

3.4、实验

假设需要从柯西分布中采样数据，我们利用Metropolis采样算法来生成样本，其中，柯西分布的概率密度函数为：

$$f(\theta) = \frac{1}{\pi(1 + \theta^2)}$$

那么，根据上述的Metropolis采样算法的流程，接受概率 α 的值为：

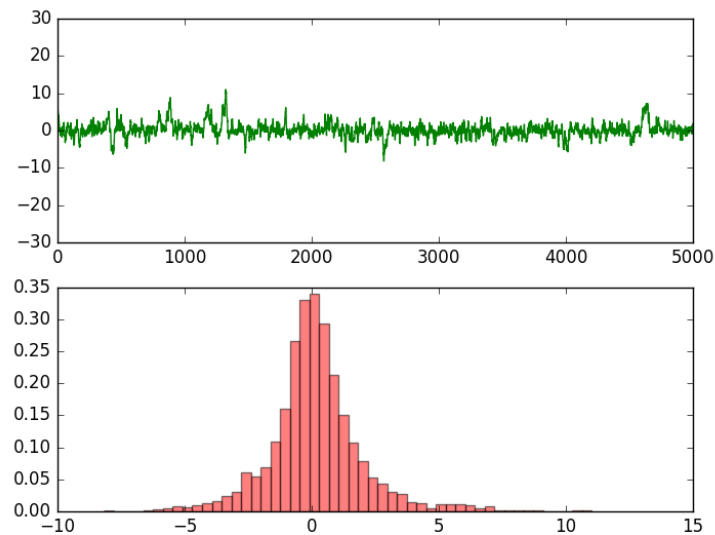
$$\alpha = \min \left(1, \frac{1 + [\theta^{(t)}]^2}{1 + [\theta^{(*)}]^2} \right)$$

代码如下：

```
1  '''
2  Date:20160629
3  @author: zhaozhiyong
4  '''
5  import random
6  from scipy.stats import norm
7  import matplotlib.pyplot as plt
8
9  def cauchy(theta):
10     y = 1.0 / (1.0 + theta ** 2)
11     return y
12
13  T = 5000
14  sigma = 1
15  thetamin = -30
16  thetamax = 30
17  theta = [0.0] * (T+1)
18  theta[0] = random.uniform(thetamin, thetamax)
19
20  t = 0
21  while t < T:
22     t = t + 1
23     theta_star = norm.rvs(loc=theta[t - 1], scale=sigma, size=1, random_state=None)
24     #print theta_star
25     alpha = min(1, (cauchy(theta_star[0]) / cauchy(theta[t - 1])))
26
27     u = random.uniform(0, 1)
28     if u <= alpha:
29         theta[t] = theta_star[0]
30     else:
31         theta[t] = theta[t - 1]
32
```

```
33 ax1 = plt.subplot(211)
34 ax2 = plt.subplot(212)
35 plt.sca(ax1)
36 plt.ylim(thetamin, thetamax)
37 plt.plot(range(T+1), theta, 'g-')
38 plt.sca(ax2)
39 num_bins = 50
40 plt.hist(theta, num_bins, normed=1, facecolor='red', alpha=0.5)
41 plt.show()
```

实验的结果：



对于Metropolis采样算法，其要求选定的分布必须是对称的，为了弥补这样的缺陷，在下一篇中，介绍一下Metropolis-Hastings采样算法，其是Metropolis采样算法的推广形式。

参考文献

- 1、[马尔可夫链蒙特卡罗算法](#)
- 2、[受限玻尔兹曼机（RBM）学习笔记（一）预备知识](#)
- 3、[LDA数学八卦](#)

顶 踩
7 0

- [上一篇](#) C/C++——set的基本操作总结
- [下一篇](#) PHP基础——字符串的常用操作

相关文章推荐

- 马尔科夫蒙特卡洛算法(MCMC)
- MySQL在微信支付下的高可用运营-莫晓东
- MCMC抽样算法要点总结
- 容器技术在58同城的实践-姚远
- MCMC算法学习总结
- SDCC 2017之容器技术实战线上峰会
- 简单易学的机器学习算法——马尔可夫链蒙特卡罗...
- SDCC 2017之数据库技术实战线上峰会

- R语言 MCMC算法及其实现
- 腾讯云容器服务架构实现介绍-董晓杰
- MCMC算法介绍
- 微博热点事件背后的数据库运维心得-张冬洪
- Metropolis 采样与蒙特卡洛算法
- 蒙特卡洛算法
- 强化学习中的蒙特卡洛 (monte-carlo)
- MCMC (马尔科夫蒙特卡洛) ——q

查看评论

暂无评论

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场