Math in Information Age

Course Review

Chen Gong, Mingdong Wu, Cheng Chen

EECS, PKU

Outline

1. SVD 复习
2. 作业讲解
3. MCMC 方法复习

SVD 复习

Power Method

在多项式时间内计算矩阵 A 的奇异值分解：*power method*

设 $A=\sum_i \sigma_i u_i v_i, B=A^TA=\sum_i \sigma_i^2 v_i v_i^T$

$B^k=\sum_{i=1}^{r} \sigma_i^{2k} v_i v_i^T$

若 $\sigma_1>\sigma_2$，有 $B^k \to \sigma_1^{2k} v_1 v_1^T$

更快的方法

计算 $B^k x$:

$x = \sum_{i=1}^{d} c_i v_i, B^k x \approx \sigma_1^{2k} c_1 v_1$

为了计算 k 个奇异向量：
子空间： $r, Ar, ..., A^{k-1} r$
重复将 A 作用于这个子空间，每次作用后计算正交基，避免这个子空间坍缩成由 v_1 张成的一维子空间

$\sigma_1 = \sigma_2?$

中心化数据

是否要中心化数据？

- 数据的统计特征，和均值的关系：中心化数据
- 找矩阵的低秩估计：不中心化

作业讲解

MCMC 方法复习

Monte Carlo 方法是一种用随机数来模拟复杂的物理系统或数学计算的方法。尽管得到的结果有些许误差，但使用方便，效率较高，有些情况下甚至是唯一的选择。

- 计算积分 $\int_a^b f(x)dx$：在 $[a,b]$ 区间上用均匀分布随机得到 N 个点 $\{x_i\}$，计算 $\frac{b-a}{N}\sum_{i=1}^N f(x_i)$。
- 光线追踪算法中使用 Monte Carlo 方法，来模拟“视窗”。

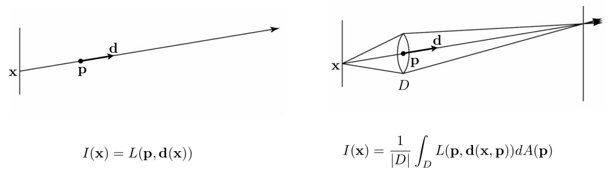

图 1: 光线追踪[1]

[1] Cornell CS4620 Spring 2013 · Lecture 22

马尔科夫链

在概率论研究的历史上，一度有观点认为，随机变量的独立性只是一种理想的情况，因而所谓的大数定律等仅仅适用于赌博游戏这种简单的问题。

俄国数学家 Андрей Андреевич Марков 为了反驳这种观点，构造了一组按条件概率相互依赖的随机变量，并证明了这组随机变量在一定条件下最终会收敛于一个稳定的向量。

简单来说，马尔科夫链由初始概率分布 $\mathbf{p}_0$ 和概率转移矩阵 P 构成，每一时刻的概率分布由 $\mathbf{p}_t = \mathbf{p}_{t-1}P$ 得到。马尔科夫链同样可以用无向图上的随机游走表示。

这里，P 实际上代表了一种条件概率，其中包含了无后效性。

非周期性

一个很自然的想法是，$\mathbf{p}_t$ 在 t 趋于无穷的时候，是否一定会收敛，对应于随机游走是否会达到一个平衡状态。这显然是不一定的：

- 比如两个状态的马尔科夫链：$\mathbf{p}_0 = (1, 0)$，$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

以上例子说明马尔科夫链在一定情况下会陷入一种循环往复的状态。
定义：若马氏链中所有有向环的长度的最大公倍数为 1，则称其是非周期性的。

平稳分布

令 $\mathbf{a}(t) = \frac{1}{t}(\mathbf{p}_0 + \mathbf{p}_1 + ...\mathbf{p}_t)$。

马尔科夫链基本定理：对于一个连通的马尔科夫链，存在一个唯一的概率分布 π_0 满足 $\pi P = \pi$。且 $\lim_{t\to\infty} \mathbf{a}(t) = \pi$。

π 被称为平稳分布。

马尔科夫链举例

- 小明去买彩票。从他家出发，往左走 k_1 步有一个彩票店，往右走 k_2 步有另一个彩票店。小明觉得选择过于困难，索性采用随机游走的办法，每一时刻有 $\frac{1}{2}$ 的概率往右走 1 步，或者 $\frac{1}{2}$ 往左走一步。求小明去两家店买彩票的概率。
- 盒子中有 3 个编号分别为 1，2，3 的球。小明从里面摸球，请问能否用马尔科夫链描述小明摸球的过程？

Markov chain Monte Carlo

Monte Carlo 方法：利用随机采样来完成某项数值的估计。

一个关键问题：如何随机采样？

直接得到所有状态的概率的困难包括但不限于：

- 状态数目太多，无法全部存储或一一读取。
- 单个状态的概率计算过于复杂，难以对所有状态都算一遍（常在贝叶斯统计中出现此问题）

MCMC 方法流程

- （将问题转换成）求随机变量函数的期望 $E(f(X))$，并有一个有限时间内可计算的概率函数 $p(x)$，或者知道 $p(x)$ 相关的一些性质。
- 隐式地构造转移规则，使得对应的马尔科夫链中，每个状态 x 的平稳分布恰好为 $p(x)$。
- 任取一个起始状态 x_0，一步步进行转移。
- 当 t 足够大时，计算 $\frac{1}{t}\sum_{i=1}^{t} f(x_t)$。

Metropolis-Hasting 算法

任取一个维度 d，将离散型随机变量 X 的取值（也叫状态）按一定的规则赋值到 d 维晶格上。注意到 d 维晶格也可以看作是度至多为 $2d$ 的无向图。

取 r 为 x_i 最大的度，也就是 $2d$，x_j 为 x_i 相邻的点。

$$P(i,j) = \frac{1}{r}\min(1, \frac{p_j}{p_i})$$

$$P(i,i) = 1 - \sum_{j \neq i} P(i,j)$$

值得注意的是这里我们只要求 $\frac{p_i}{p_j}$，而不需要确切地知道两个的概率。因此有时候我们在不知道概率的情况下，根据对称性使用赋值 p_i^* 代替概率。

Gibbs Sampling

Gibbs Sampling 同样应用在 d 维晶格模型上。其流程是这样的：

- 对于当前状态 $\mathbf{x} = (x_1, x_2, ..., x_d)$，随机选择一个坐标维度，这里不妨设为第 1 个维度。
- 取遍 $y_1 = 1, 2, 3, ..., n_1$， n_1 为第一维的能取到的最大值。以 $P(y_1|x_2, ...x_d)$ 的概率分布随机得到一个 $\mathbf{y} = (y_1, x_2, ..., x_d)$。$\mathbf{y}$ 即为下一个状态。

也就是说，当前状态为 $\mathbf{x} = (x_1, x_2, ..., x_n)$ 时，对于仅有一个维度不同的状态 $\mathbf{x}' = (x_1, x_2, ..., x_i', ..., x_n)$，

$$P_{\mathbf{x}\mathbf{x}'} = \frac{1}{d} P(x_i'|x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n)$$

利用 MCMC 方法计算高维空间问题

计算高维空间的技巧：

$$V(R) = \frac{V(S_k \cap R)}{V(S_{k-1} \cap R)} \cdot \frac{V(S_{k-1} \cap R)}{V(S_{k-2} \cap R)} \dots \frac{V(S_2 \cap R)}{V(S_1 \cap R)} \cdot V(S_1)$$

其中，$S_1 \subset R \subset S_k$，$S_1, S_2, ..., S_k$ 是一系列半径逐渐增长的高维立方体（或者球体）。

如何计算 $\frac{V(S_k \cap R)}{V(S_{k-1} \cap R)}$？我们计算它的倒数。

- 类似格点法，构造晶格，X 为晶格上的随机变量。
- 取函数 $f(X)$，$X \in S_{k-1} \cap R$，则 $f(X) = 1$，否则取零。
- 若 $X \in S_k \cap R$，则取 $p^*(X) = 1$，否则取零。然后使用 Metropolis-Hasting 算法模拟即可。

无向图上的随机游走问题

考虑带边权的无向图 G，ω_{xy} 就是边 xy 的权重。

定义 $\omega_x = \sum_y \omega_{xy}$。令 $p_{xy} = \frac{\omega_{xy}}{\omega_x}$。则令转移矩阵 $P = (p_{xy})$，可以得到 P 的平稳分布满足 $\pi_x = \frac{\omega_x}{\omega_{total}}$。

定义马尔科夫链的 ϵ 混合时间：若正整数 t 满足，对于任意初始分布 $\mathbf{p}_0$，$\|\mathbf{p}_t - \pi\|_1 \le \epsilon$，则称 t 为 ϵ 混合时间。

无向图上的随机游走

对于无向图 G 的顶点的一个子集合 S，令 $\pi(S) = \sum_{x \in S} \pi_x$。定义 S 的传导系数 $\Phi(S)$：

$$\Phi(S) = \frac{\sum_{(x,y) \in (S,\bar{S})} \pi_x p_{xy}}{\min(\pi(S), \pi(\bar{S}))}$$

定义马尔科夫链的标准化传导系数为

$$\Phi = min_{S \in V, S \neq \emptyset} \Phi(S)$$

则 ϵ 混合时间为

$$O(\frac{\ln(1/\pi_{min})}{\Phi^2 \epsilon^3})$$