

我是8位的

I am 8 bits, what about you?

随笔 - 205, 文章 - 0, 评论 - 103, 阅读 - 101万

导航

博客园

首页

新随笔

联系

订阅 RSS

管理

2022年3月						
日	一	二	三	四	五	六
27	28	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

公告

你的支持是我的动力

欢迎关注微信公众号“我是8位的”



昵称：我是8位的

园龄：4年7个月

粉丝：288

关注：5

+加关注

盖楼抽奖

#她的梦想在发光#

HWD科技女性故事有奖征集

分享最打动你的科技女性故事

活动时间：2022年3月8日-3月18日

马上参与



搜索

找找看

谷歌搜索

常用链接

我的随笔

我的评论

我的参与

最新评论

我的标签

积分与排名

积分 - 457097

排名 - 1198

概率笔记12——多维正态分布的最大似然估计

我们在前面的章节中见识过二维正态分布， (X, Y) 服从参数为 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 的二维正态分布，记作 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ ，它的密度函数：

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{2\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right)$$

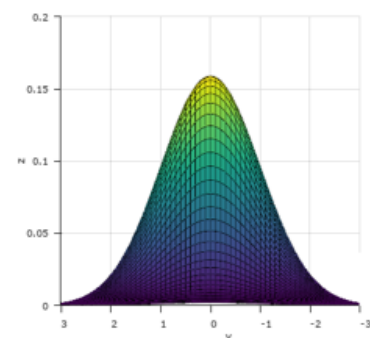
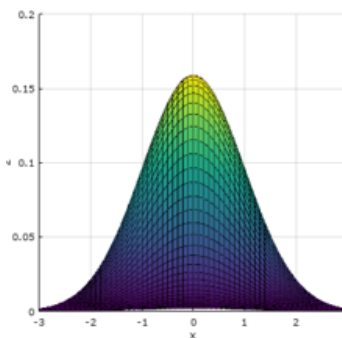
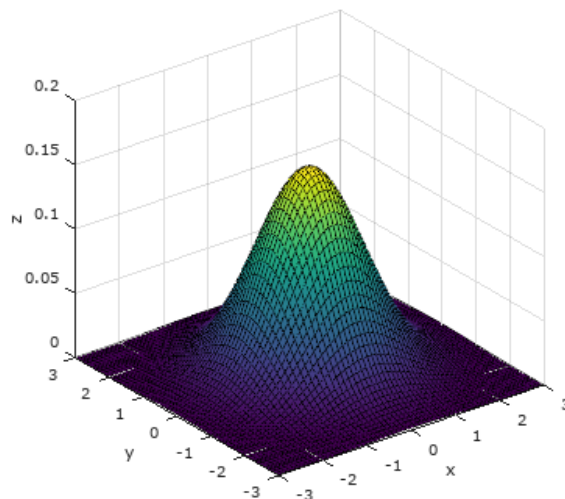
$$= \frac{1}{(\sqrt{2\pi})^2\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{2\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right)$$

其中 μ_1 是第1维度的均值， σ_1^2 是第1维度的方差， ρ 是将两个维度的相关性规范到-1到+1之间的统计量，称为样本的相关系数，定义为：

$$\rho = \frac{\text{COV}(X, Y)}{\sigma_1\sigma_2}, \quad |\rho| < 1$$

对于二维正态随机变量 (X, Y) ， X 和 Y 相互独立的充要条件是二者的协方差为0，也就是参数 $\rho=0$ 。由于一维随机变量没有是否独立一说， ρ 一定是0，因此没有在一维随机变量的正态分布中体现 ρ 。

下图是一个标准二维正态分布和其在 x - z ， y - z 平面的投影：



多维正态分布

随笔分类 (211)

- ★★资源下载★★(1)
- Java并发编程(1)
- 程序员的数学(24)
- 单变量微积分(31)
- 多变量微积分(24)
- 概率(24)
- 机器学习(27)
- 软件设计(1)
- 数据分析(6)
- 数据结构与算法(27)
- 随笔(5)
- 线性代数(34)
- 项目管理(2)
- 转载(4)

随笔档案 (205)

- 2021年2月(1)
- 2020年3月(2)
- 2020年2月(6)
- 2020年1月(4)
- 2019年12月(7)
- 2019年11月(15)
- 2019年9月(3)
- 2019年8月(6)
- 2019年7月(1)
- 2019年6月(8)
- 2019年5月(3)
- 2019年4月(5)
- 2019年3月(7)
- 2019年2月(3)
- 2019年1月(7)
- 更多

阅读排行榜

- 1. 使用Apriori进行关联分析 (一) (29768)
- 2. 线性代数笔记12——列空间和零空间 (28772)
- 3. FP-growth算法发现频繁项集 (一)——构建FP树(24430)
- 4. 寻找“最好” (2) ——欧拉-拉格朗日方程(23099)
- 5. 多变量微积分笔记3——二元函数的极值(22772)

评论排行榜

- 1. 隐马尔可夫模型 (一) (8)
- 2. 线性代数笔记12——列空间和零空间 (7)
- 3. 线性代数笔记3——向量2 (点积) (7)
- 4. FP-growth算法发现频繁项集 (一)——构建FP树(5)
- 5. 寻找“最好” (2) ——欧拉-拉格朗日方程(4)

推荐排行榜

- 1. 寻找“最好” (2) ——欧拉-拉格朗日方程(7)
- 2. FP-growth算法发现频繁项集 (一)——构建FP树(7)
- 3. 线性代数笔记3——向量2 (点积) (6)
- 4. FP-growth算法发现频繁项集 (二)——发现频繁项集(5)
- 5. 隐马尔可夫模型 (一) (5)

最新评论

- 1. Re:线性代数笔记3——向量2 (点积)
如果点积小于0，即夹角小于90°，这个写错了吧。应该是夹角大于90°
--猫猫猫猫猫大人

现在推广到多维，为了便于表达，我们用向量的形式表示随机变量和参数，对于n维随机变量：

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \quad \sigma = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_n \end{bmatrix}$$

这里只考虑所有维度变量互相独立的情况，即ρ=0的情况，此时密度函数可表示为：

$$f(x) = \frac{1}{(\sqrt{2\pi})^n \sigma_1 \sigma_2 \cdots \sigma_n} \exp\left(-\frac{1}{2}\left[\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 + \cdots + \left(\frac{x_n - \mu_n}{\sigma_n}\right)^2\right]\right) \quad ①$$
$$= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right) \cdots \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{1}{2}\left(\frac{x_n - \mu_n}{\sigma_n}\right)^2\right)$$

上面的结果告诉我们，在各维度相互独立的情况下，多维正态分布的概率密度其实就是各个维度的正态分布密度函数的乘积。

在①中：

$$\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 + \cdots + \left(\frac{x_n - \mu_n}{\sigma_n}\right)^2$$
$$= [x_1 - \mu_1 \quad x_2 - \mu_2 \quad \cdots \quad x_n - \mu_n] \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_n - \mu_n \end{bmatrix}$$
$$= (x - \mu)^T \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix} (x - \mu)$$

σ_i²表示x_i的方差，如此看来，中间那个矩阵实际上是协方差矩阵的逆矩阵：

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix}, \quad \Sigma \Sigma^{-1} = I$$
$$\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 + \cdots + \left(\frac{x_n - \mu_n}{\sigma_n}\right)^2 = (x - \mu)^T \Sigma^{-1} (x - \mu) \quad ②$$

根据行列式的性质，上三角矩阵的行列式等于主对角线所有元素的乘积，斜对角矩阵当然也是一个上三角矩阵，因此协方差矩阵的行列式是：

$$|\Sigma| = \sigma_1^2 \sigma_2^2 \cdots \sigma_n^2$$
$$\sqrt{|\Sigma|} = \sigma_1 \sigma_2 \cdots \sigma_n \quad ③$$

将②、③代入①中，得到最终结果：

2. Re:线性代数笔记10——矩阵的LU分解写的很好，不过LU分解的前提是错的，LU分解只需要第三个条件，如果允许行置换就是下面写到的PLU，可以分解所有矩阵

--wiki3D

3. Re:单变量微积分笔记20——三角替换1 (sin和cos) 很nice

--尹保棕

4. Re:线性代数笔记24——微分方程和exp(At)

有些图片挂了呢

--ccchendada

5. Re:寻找“最好” (2) ——欧拉-拉格朗日方程

提个issue，最速降线中

\$v = \{2gh\}^{1/2}\$ 与配图不一致，建议以起点为原点，向右伸出x轴，向下伸出y轴建立坐标系

--trustInU

$$\begin{aligned} f(x) &= \frac{1}{(\sqrt{2\pi})^n \sigma_1 \sigma_2 \cdots \sigma_n} \exp \left(-\frac{1}{2} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 + \cdots + \left(\frac{x_n - \mu_n}{\sigma_n} \right)^2 \right] \right) \\ &= \frac{1}{(\sqrt{2\pi})^n \sqrt{|\Sigma|}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \\ &= (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \\ &= f(x; \mu, \Sigma) \end{aligned}$$

最大似然估计量

n维相互独立的随机变量x服从正态分布：

$$x \sim N(\mu, \sigma^2), \quad \sigma_i \geq 0$$

在求最大似然估计量时和一维随机变量有所区别，根据上一节的最终结果：

$$f(x) = f(x; \mu, \Sigma)$$

假设有m个可观察样本，那么最大似然函数是：

$$\begin{aligned} L(\mu, \Sigma) &= \prod_{i=1}^m f(x^{(i)}; \mu, \Sigma) \\ &= \prod_{i=1}^m (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \right) \\ &= (2\pi)^{-\frac{mn}{2}} |\Sigma|^{-\frac{m}{2}} \exp \left(-\frac{1}{2} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \right) \end{aligned}$$

其对数似然函数是：

$$\begin{aligned} \ln L(\mu, \Sigma) &= \ln (2\pi)^{-\frac{mn}{2}} + \ln |\Sigma|^{-\frac{m}{2}} + \ln \exp \left(-\frac{1}{2} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \right) \\ &= -\frac{mn}{2} \ln 2\pi - \frac{m}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \\ &= C - \frac{m}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \end{aligned}$$

其中m和n是已知的，C是一个常数。

求极值需要对μ和Σ求偏导：

$$\begin{cases} \frac{\partial \ln L}{\partial \mu} = 0 \\ \frac{\partial \ln L}{\partial \Sigma} = 0 \end{cases}$$

μ和Σ是矩阵，涉及到矩阵的求导法则。先看对μ的求导，lnL由3个因子组成，只有一个因子含有μ，因此：

$$\frac{\partial \ln L}{\partial \mu} = \frac{\partial}{\partial \mu} \left(-\frac{1}{2} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \right)$$

其中：

$$\begin{aligned}
 (x - \mu)^T \Sigma^{-1} (x - \mu) &= (x^T - \mu^T) \Sigma^{-1} (x - \mu) \\
 &= (x^T \Sigma^{-1} - \mu^T \Sigma^{-1}) (x - \mu) \\
 &= x^T \Sigma^{-1} x - x^T \Sigma^{-1} \mu - \mu^T \Sigma^{-1} x + \mu^T \Sigma^{-1} \mu
 \end{aligned}$$

上式中:

$$x^T \Sigma^{-1} \mu = [x_1 \quad x_2 \quad \cdots \quad x_n] \Sigma^{-1} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_n] \Sigma^{-1} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mu^T \Sigma^{-1} x$$

因此:

$$\begin{aligned}
 (x - \mu)^T \Sigma^{-1} (x - \mu) &= x^T \Sigma^{-1} x - x^T \Sigma^{-1} \mu - \mu^T \Sigma^{-1} x + \mu^T \Sigma^{-1} \mu \\
 &= x^T \Sigma^{-1} x - 2x^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} \mu \\
 \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) &= \sum_{i=1}^m (x^{(i)T} \Sigma^{-1} x^{(i)} - 2x^{(i)T} \Sigma^{-1} \mu + \mu^T \Sigma^{-1} \mu) \\
 &= \sum_{i=1}^m x^{(i)T} \Sigma^{-1} x^{(i)} - 2 \sum_{i=1}^m x^{(i)T} \Sigma^{-1} \mu + m \mu^T \Sigma^{-1} \mu
 \end{aligned}$$

将该结论代入 $\partial \ln L / \partial \mu$ 中:

$$\begin{aligned}
 \frac{\partial \ln L}{\partial \mu} &= \frac{\partial}{\partial \mu} \left(-\frac{1}{2} \left(\sum_{i=1}^m x^{(i)T} \Sigma^{-1} x^{(i)} - 2 \sum_{i=1}^m x^{(i)T} \Sigma^{-1} \mu + m \mu^T \Sigma^{-1} \mu \right) \right) \\
 &= \frac{\partial}{\partial \mu} \left(-\frac{1}{2} \sum_{i=1}^m x^{(i)T} \Sigma^{-1} x^{(i)} \right) + \frac{\partial}{\partial \mu} \left(\sum_{i=1}^m x^{(i)T} \Sigma^{-1} \mu \right) - \frac{1}{2} \frac{\partial}{\partial \mu} m \mu^T \Sigma^{-1} \mu \\
 &= \underbrace{\frac{\partial}{\partial \mu} \left(\sum_{i=1}^m x^{(i)T} \Sigma^{-1} \mu \right)}_{a_1} - \underbrace{\frac{1}{2} \frac{\partial}{\partial \mu} m \mu^T \Sigma^{-1} \mu}_{a_2}
 \end{aligned}$$

μ 和 Σ 是矩阵, 根据矩阵的求导法则:

$$\begin{aligned}
 \text{if } f(X) &= A^T X, \quad \text{then } \frac{df}{dX} = A \\
 \Rightarrow a_1 &= \sum_{i=1}^m (x^{(i)T} \Sigma^{-1})^T = \sum_{i=1}^m \Sigma^{-1T} x^{(i)}
 \end{aligned}$$

因为 Σ^{-1} 是一个对称矩阵, 因此:

$$\Sigma^{-1T} = \Sigma^{-1}, \quad a_1 = \sum_{i=1}^m \Sigma^{-1T} x^{(i)} = \sum_{i=1}^m \Sigma^{-1} x^{(i)} = \Sigma^{-1} \sum_{i=1}^m x^{(i)}$$

根据矩阵的求导法则:

$$\begin{aligned}
 \text{if } f(X) &= X^T A X, \quad \text{then } \frac{df}{dX} = A X + A^T X \\
 \text{when } A &= A^T, \quad \text{then } \frac{df}{dX} = A X + A^T X = 2 A X \\
 \Rightarrow a_2 &= \frac{1}{2} \frac{\partial}{\partial \mu} m \mu^T \Sigma^{-1} \mu = m \Sigma^{-1} \mu
 \end{aligned}$$

将 a_1, a_2 代入 $\partial \ln L / \partial \mu$ 中:

$$\frac{\partial \ln L}{\partial \mu} = a_1 + a_2 = \sum_{i=1}^m \Sigma^{-1} x^{(i)} - m \Sigma^{-1} \mu = 0$$

$$\hat{\mu} = \frac{\Sigma^{-1} \sum_{i=1}^m x^{(i)}}{m \Sigma^{-1}} = \frac{1}{m} \sum_{i=1}^m x^{(i)} = \bar{x}$$

再看对 Σ 求偏导:

$$\begin{aligned} \frac{\partial \ln L}{\partial \Sigma} &= \frac{\partial}{\partial \Sigma} \left(c - \frac{m}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \right) \\ &= -\frac{m}{2} \underbrace{\frac{\partial}{\partial \Sigma} \ln |\Sigma|}_{b_1} - \frac{1}{2} \underbrace{\frac{\partial}{\partial \Sigma} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu)}_{b_2} \end{aligned}$$

Σ 和 Σ^{-1} 都是实对称矩阵, 根据矩阵的求导法则, 当 $\mathbf{A}$ 是实对称矩阵时:

$$\begin{aligned} \frac{\partial \ln |\mathbf{A}|}{\partial \mathbf{A}} &= \mathbf{A}^{-1} \\ \Rightarrow b_1 &= \frac{\partial}{\partial \Sigma} \ln |\Sigma| = \Sigma^{-1} \end{aligned}$$

再看 b_2 . 设 ω_{pq} 是 Σ 第 p 行第 q 列的元素, E_{pq} 是一个第 p 行第 q 列元素为1, 其它元素全为0的矩阵, E 与 Σ^{-1} 同阶. 根据矩阵的求导公式:

$$\begin{aligned} \frac{\partial \mathbf{X}^{-1}}{\partial x} &= -\mathbf{X}^{-1} \frac{\partial \mathbf{X}}{\partial x} \mathbf{X}^{-1} \\ \Rightarrow \frac{\partial \Sigma^{-1}}{\partial \omega_{pq}} &= -\Sigma^{-1} \frac{\partial \Sigma}{\partial \omega_{pq}} \Sigma^{-1} = -\Sigma^{-1} E_{pq} \Sigma^{-1} \\ \Rightarrow \frac{\partial (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu)}{\partial \omega_{pq}} &= (x^{(i)} - \mu)^T \frac{\partial \Sigma^{-1}}{\partial \omega_{pq}} (x^{(i)} - \mu) \\ &= (x^{(i)} - \mu)^T (-\Sigma^{-1} E_{pq} \Sigma^{-1}) (x^{(i)} - \mu) \\ &= -(x^{(i)} - \mu)^T (\Sigma^{-1} E_{pq} \Sigma^{-1}) (x^{(i)} - \mu) \end{aligned}$$

已经知道了 Σ^{-1} 是一个对称矩阵, 矩阵乘法满足结合律, 在不改变矩阵顺序的条件下可以任意加括号:

$$\begin{aligned} \frac{\partial (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu)}{\partial \omega_{pq}} &= -(x^{(i)} - \mu)^T (\Sigma^{-1} E_{pq} \Sigma^{-1}) (x^{(i)} - \mu) \\ &= -((x^{(i)} - \mu)^T \Sigma^{-1}) E_{pq} (\Sigma^{-1} (x^{(i)} - \mu)) \\ &= -((x^{(i)} - \mu)^T (\Sigma^{-1})^T) E_{pq} (\Sigma^{-1} (x^{(i)} - \mu)) \\ &= -\underbrace{(\Sigma^{-1} (x^{(i)} - \mu))^T}_{\mathbf{A}^T \mathbf{B}^T = (\mathbf{AB})^T} E_{pq} (\Sigma^{-1} (x^{(i)} - \mu)) \\ &= -(\Sigma^{-1} (x^{(i)} - \mu))^T_p (\Sigma^{-1} (x^{(i)} - \mu))_q \quad (4) \end{aligned}$$

其中 $(\Sigma^{-1} (x^{(i)} - \mu))^T$ 是一个 $1 \times n$ 的矩阵, $(\Sigma^{-1} (x^{(i)} - \mu))^T_p$ 表示矩阵中的第 p 个元素; $\Sigma^{-1} (x^{(i)} - \mu)$ 是一个 $n \times 1$ 的矩阵, $(\Sigma^{-1} (x^{(i)} - \mu))_q$ 表示矩阵中的第 q 个元素. 将该结论推广到矩阵对矩阵的求导, 根据矩阵对矩阵的求导公式:

$$\frac{dF}{dX} = \begin{bmatrix} \frac{\partial F}{\partial x_{11}} & \frac{\partial F}{\partial x_{12}} & \cdots & \frac{\partial F}{\partial x_{1s}} \\ \frac{\partial F}{\partial x_{21}} & \frac{\partial F}{\partial x_{22}} & \cdots & \frac{\partial F}{\partial x_{2s}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F}{\partial x_{r1}} & \frac{\partial F}{\partial x_{r2}} & \cdots & \frac{\partial F}{\partial x_{rs}} \end{bmatrix}$$

$$\begin{aligned} & \frac{\partial}{\partial \Sigma} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \\ &= - \begin{bmatrix} \frac{\partial}{\partial \omega_{11}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) & \frac{\partial}{\partial \omega_{12}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) & \cdots & \frac{\partial}{\partial \omega_{1n}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \\ \frac{\partial}{\partial \omega_{21}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) & \frac{\partial}{\partial \omega_{22}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) & \cdots & \frac{\partial}{\partial \omega_{2n}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \omega_{n1}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) & \frac{\partial}{\partial \omega_{n1}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) & \cdots & \frac{\partial}{\partial \omega_{nn}} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \end{bmatrix} \\ &= - \begin{bmatrix} (\Sigma^{-1} (x^{(i)} - \mu))_1^T (\Sigma^{-1} (x^{(i)} - \mu))_1 & \cdots & (\Sigma^{-1} (x^{(i)} - \mu))_1^T (\Sigma^{-1} (x^{(i)} - \mu))_n \\ (\Sigma^{-1} (x^{(i)} - \mu))_2^T (\Sigma^{-1} (x^{(i)} - \mu))_1 & \cdots & (\Sigma^{-1} (x^{(i)} - \mu))_2^T (\Sigma^{-1} (x^{(i)} - \mu))_n \\ \vdots & \ddots & \vdots \\ (\Sigma^{-1} (x^{(i)} - \mu))_n^T (\Sigma^{-1} (x^{(i)} - \mu))_1 & \cdots & (\Sigma^{-1} (x^{(i)} - \mu))_n^T (\Sigma^{-1} (x^{(i)} - \mu))_n \end{bmatrix} \\ &= - \underbrace{\begin{bmatrix} (\Sigma^{-1} (x^{(i)} - \mu))_1^T \\ (\Sigma^{-1} (x^{(i)} - \mu))_2^T \\ \vdots \\ (\Sigma^{-1} (x^{(i)} - \mu))_n^T \end{bmatrix}}_{A_1} \underbrace{\begin{bmatrix} (\Sigma^{-1} (x^{(i)} - \mu))_1 & (\Sigma^{-1} (x^{(i)} - \mu))_2 & \cdots & (\Sigma^{-1} (x^{(i)} - \mu))_n \end{bmatrix}}_{A_2} \end{aligned}$$

其中：

$$A_2 = [(\Sigma^{-1} (x^{(i)} - \mu))_1 \quad (\Sigma^{-1} (x^{(i)} - \mu))_2 \quad \cdots \quad (\Sigma^{-1} (x^{(i)} - \mu))_n] = (\Sigma^{-1} (x^{(i)} - \mu))^T$$

在 A_1 中， $(\Sigma^{-1} (x^{(i)} - \mu))^T$ 是一个 $1 \times n$ 的矩阵， $(\Sigma^{-1} (x^{(i)} - \mu))_i^T$ 表示矩阵中的第 i 个元素，是一个标量； $\Sigma^{-1} (x^{(i)} - \mu)$ 是一个 $n \times 1$ 的矩阵， $(\Sigma^{-1} (x^{(i)} - \mu))_i$ 表示矩阵中的第 i 个元素，也是一个标量，因此：

$$(\Sigma^{-1} (x^{(i)} - \mu))^T_i = (\Sigma^{-1} (x^{(i)} - \mu))_i$$

$$A_1 = \begin{bmatrix} (\Sigma^{-1} (x^{(i)} - \mu))_1^T \\ (\Sigma^{-1} (x^{(i)} - \mu))_2^T \\ \vdots \\ (\Sigma^{-1} (x^{(i)} - \mu))_n^T \end{bmatrix} = \begin{bmatrix} (\Sigma^{-1} (x^{(i)} - \mu))_1 \\ (\Sigma^{-1} (x^{(i)} - \mu))_2 \\ \vdots \\ (\Sigma^{-1} (x^{(i)} - \mu))_n \end{bmatrix} = \Sigma^{-1} (x^{(i)} - \mu)$$

$$\frac{\partial}{\partial \Sigma} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) = -A_1 A_2$$

$$= -\Sigma^{-1} (x^{(i)} - \mu) (\Sigma^{-1} (x^{(i)} - \mu))^T$$

$$= -\Sigma^{-1} (x^{(i)} - \mu) (x^{(i)} - \mu)^T (\Sigma^{-1})^T$$

$$= -\Sigma^{-1} (x^{(i)} - \mu) (x^{(i)} - \mu)^T \Sigma^{-1}$$

终于可以求得 b_2 了：

$$b_2 = \frac{\partial}{\partial \Sigma} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) = \sum_{i=1}^m (-\Sigma^{-1} (x^{(i)} - \mu) (x^{(i)} - \mu)^T \Sigma^{-1})$$

现在可以看看最终的似然函数：

$$\begin{aligned} \frac{\partial \ln L}{\partial \Sigma} &= - \underbrace{\frac{m}{2} \frac{\partial}{\partial \Sigma} \ln |\Sigma|}_{b_1} - \underbrace{\frac{1}{2} \frac{\partial}{\partial \Sigma} \sum_{i=1}^m (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu)}_{b_2} \\ &= -\frac{m}{2} \Sigma^{-1} - \frac{1}{2} \sum_{i=1}^m (-\Sigma^{-1} (x^{(i)} - \mu) (x^{(i)} - \mu)^T \Sigma^{-1}) \end{aligned}$$

I 是单位矩阵， $\Sigma^{-1} I = \Sigma^{-1}$ ：

$$\begin{aligned}
\frac{\partial \ln L}{\partial \Sigma} &= -\frac{m}{2} \Sigma^{-1} I - \frac{1}{2} \sum_{i=1}^m \left(-\Sigma^{-1} (x^{(i)} - \mu) (x^{(i)} - \mu)^T \Sigma^{-1} \right) \\
&= -\frac{m}{2} \Sigma^{-1} I + \frac{1}{2} \sum_{i=1}^m \left(\Sigma^{-1} (x^{(i)} - \mu) (x^{(i)} - \mu)^T \Sigma^{-1} \right) \\
&= -\frac{1}{2} \Sigma^{-1} \left(mI - \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T \Sigma^{-1} \right) \\
&= -\frac{1}{2} \Sigma^{-1} \left(m \Sigma \Sigma^{-1} - \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T \Sigma^{-1} \right) \\
&= -\frac{1}{2} \Sigma^{-1} \left(m \Sigma - \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T \right) \Sigma^{-1} \\
&= 0
\end{aligned}$$

等号两侧同时左乘 Σ ：

$$\begin{aligned}
\Sigma \left(-\frac{1}{2} \Sigma^{-1} \right) \left(m \Sigma - \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T \right) \Sigma^{-1} &= \Sigma 0 \\
-\frac{1}{2} I \left(m \Sigma - \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T \right) \Sigma^{-1} &= 0 \\
\left(m \Sigma - \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T \right) \Sigma^{-1} &= 0
\end{aligned}$$

两侧同时右乘 Σ ：

$$\begin{aligned}
\left(m \Sigma - \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T \right) \Sigma^{-1} \Sigma &= 0 \Sigma \\
m \Sigma - \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T &= 0
\end{aligned}$$

最终解得：

$$\Sigma = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T$$

最终结论，多维正态分布的最大似然估计量是：

$$\begin{aligned}
\hat{\mu} &= \frac{1}{m} \sum_{i=1}^m x^{(i)} = \bar{x} \\
\Sigma &= \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu) (x^{(i)} - \mu)^T
\end{aligned}$$

作者：我是8位的

出处：<http://www.cnblogs.com/bigmonkey>

本文以学习、研究和分享为主，如需转载，请联系本人，标明作者和出处，非商业用途！

扫描二维码关注公众号“我是8位的”



随笔

分类: 概率

标签: 多维正态分布, 最大似然估计, 详细推导过程

好文要顶

关注我

收藏该文



我是8位的

关注 - 5

粉丝 - 288

+加关注

0

推荐

0

反对

« 上一篇: 概率笔记11——一维正态分布的最大似然估计

» 下一篇: Java 并发 (1) ——线程安全

posted on 2019-08-19 19:34 我是8位的 阅读(8711) 评论(2) 编辑 收藏 举报

刷新评论 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】华为 HWD 2022 故事征集, 分享最打动你的科技女性故事

【推荐】华为开发者专区, 与开发者一起构建万物互联的智能世界

广告 X

JAVA Office文档在线编辑APIs

简单易用的Word, Excel, PowerPoint在线编辑接口

cloud.e-iceblue.cn

打开

编辑推荐:

- 革命性创新, 动画杀手铜 @scroll-timeline
- 戏说领域驱动设计 (十二) —— 服务
- ASP.NET Core 6框架揭秘实例演示[16]: 内存缓存与分布式缓存的使用
- .Net Core 中无处不在的 Async/Await 是如何提升性能的?
- 分布式系统改造方案 —— 老旧系统改造篇

#她的梦想在发光#

HWD科技女性故事有奖征集

活动时间: 2022年3月8日-3月18日



最新新闻:

- 乔布斯的创业搭档: 他缺乏工程师才能, 不得不锻炼营销能力来弥补
- 美国大厂码农薪资曝光: 年薪18万美元, 够养家, 不够买海景房
- 两张照片就能转视频! Google提出FLIM帧插值模型

- [Android 再推“杀手级”功能，可回收 60% 存储空间](#)
- [溺在理财暴雷潮的投资人：本金63万，月兑25元不够卖菜](#)
- » [更多新闻...](#)

Powered by:
[博客园](#)

Copyright © 2022 我是8位的
Powered by .NET 6 on Kubernetes