机器学习中的高斯分布

原创 Zicon in广工
于 2020-12-22 11:35:56 发布 470 收藏 10 版权
分类专栏： 笔记 深度学习

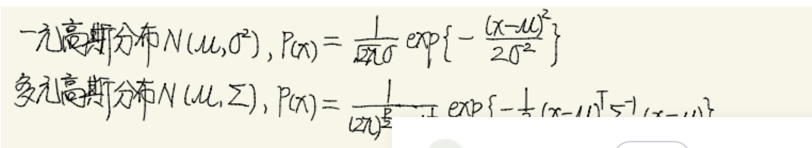
文章目录

- 一、高斯分布的概率密度函数
- 二、一元高斯分布的极大似然估计
  - 2.1 μ_{MLE}, σ_{MLE} 的求解
  - 2.2 验证 μ_{MLE}, σ_{MLE} 的无偏性
- 三、多元高斯分布
  - 3.1 马氏距离
  - 3.2 $(x-\mu)^T\Sigma^{-1}(x-\mu)$ 的求解
  - 3.3 实际应用过程的问题
- 四、联合分布 -> 边缘分布 + 条件分布
  - 4.1 数据说明
  - 4.2 关键推论
  - 4.3 求解边缘分布 $P(x_a)$
  - 4.4 求解条件分布 $P(x_b|x_a)$
- 五、线性高斯系统
  - 5.1 问题介绍
  - 5.2 求解 $P(y)$
  - 5.3 求解 $P(x|y)$
- 六、参考资料

本文主要对机器学习中的高斯分布进行总结：

- 第一章先总结一元高斯分布和多元高斯分布的概率密度函数；
- 第二章以一元高斯分布为例，通过极大似然估计推导一元高斯分布中均值和方差的求法，并验证均值是无偏估计，而方差是有偏估计（相对真实值偏小）；
- 第三章主要对多元高斯分布的二次型部分进行介绍，并通过吴恩达课程中二元高斯分布的图像，体验一下均值和协方差矩阵对分布图像的影响，最后点明高斯分布在实际应用中的两个问题和解决方法；
- 第四章主要推导在已知联合分布的情况下，如何求解边缘分布和条件分布；
- 第五章主要推导线性高斯系统中相关分布的求解，可用于卡尔曼滤波的推导等。

一、高斯分布的概率密度函数



二、一元高斯分布的极大似然估计

2.1 μ_{MLE}, σ_{MLE} 的求解

设Data为：$X_{n\times p}=\{x_1,x_2,\cdots,x_n\}^T$，$x_i \overset{i.i.d}{\sim} P(x|\theta)=P(x|\mu,\sigma^2)$

$$
\begin{aligned}
\theta_{MLE}=\arg\max_{\theta} P(X|\theta) &=\arg\max_{\theta}\sum_{i=1}^{n}\log P(x_i|\theta)\\
&=\arg\max_{\theta}\sum_{i=1}^{n}\log\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}\\
&=\arg\max_{\theta}\sum_{i=1}^{n}\left[\log\frac{1}{\sqrt{2\pi}}+\log\frac{1}{\sigma}-\frac{(x-\mu)^2}{2\sigma^2}\right]
\end{aligned}
$$

对于μ_{MLE}，$\mu_{MLE}=\arg\min_{\mu}\sum_{i=1}^{n}(x_i-\mu)^2$.

$$
\begin{aligned}
&\therefore \frac{\partial}{\partial\mu}\sum_{i=1}^{n}(x_i-\mu)^2=0\\
&\sum_{i=1}^{n}2(x_i-\mu)\cdot(-1)=0\\
&\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\mu=0
\end{aligned}
$$

有 $\mu_{MLE}=\frac{1}{N}\sum_{i=1}^{n}x_i$

对于σ_{MLE}，$\sigma_{MLE}=\arg\min_{\sigma}\sum_{i=1}^{n}\left[\log\sigma+\frac{1}{2\sigma^2}(x_i-\mu)^2\right]$

$$
\begin{aligned}
&\therefore \frac{\partial}{\partial\sigma}\sum_{i=1}^{n}\left[\log\sigma+\frac{1}{2\sigma^2}(x_i-\mu)^2\right]=0\\
&\sum_{i=1}^{n}\left[\frac{1}{\sigma}+\frac{(x_i-\mu)^2}{2}\cdot(-2)\cdot\sigma^{-3}\right]=0\\
&\sum_{i=1}^{n}\left[\sigma^2-(x_i-\mu)^2\right]=0
\end{aligned}
$$

有 $\sigma^2_{MLE}=\frac{1}{N}\sum_{i=1}^{n}(x_i-\mu)^2$

2.2 验证 μ_{MLE}, σ_{MLE} 的无偏性

实际上，通过极大似然估计得到的 μ_{MLE} 是无偏估计，而 σ_{MLE} 是相对真实方差偏小的有偏估计，原因如下：

对于μ_{MLE}：

$$E[\mu_{MLE}]=E[\frac{1}{N}\sum_{i=1}^{n}x_i]=\frac{1}{N}\sum_{i=1}^{n}E[x_i]=\frac{1}{N}\sum_{i=1}^{n}\mu=\mu$$

对于σ^2_{MLE}：

$$
\begin{aligned}
E[\sigma^2_{MLE}]&=E[\frac{1}{N}\sum_{i=1}^{n}(x_i-\mu_{MLE})^2]\\
&=E[\frac{1}{N}\sum_{i=1}^{n}(x_i^2-2x_i\mu_{MLE}+\mu_{MLE}^2)]\\
&=E[\frac{1}{N}\sum_{i=1}^{n}x_i^2-\frac{1}{N}\sum_{i=1}^{n}x_i\cdot 2\mu_{MLE}+\frac{1}{N}\sum_{i=1}^{n}\mu_{MLE}^2]\\
&=E[\frac{1}{N}\sum_{i=1}^{n}x_i^2-\mu_{MLE}^2]\\
&=E[\frac{1}{N}\sum_{i=1}^{n}(x_i^2-\mu^2)-(\mu_{MLE}^2-\mu^2)]\\
&=\frac{1}{N}\sum_{i=1}^{n}(E(x_i^2)-\mu^2)-(E(\mu_{MLE}^2)-\mu^2)\\
&=\frac{1}{N}\sum_{i=1}^{n}(E(x_i^2)-E^2(x_i))-(E(\mu_{MLE}^2)-E^2(\mu_{MLE}))\\
&=\sigma^2-Var(\mu_{MLE})\\
&=\sigma^2-Var(\frac{1}{N}\sum_{i=1}^{n}x_i)\\
&=\sigma^2-\frac{1}{N^2}\sum_{i=1}^{n}\sigma^2\\
&=\frac{N-1}{N}\sigma^2
\end{aligned}
$$

因此，μ_{MLE}是无偏估计，σ_{MLE}是有偏估计

三、多元高斯分布

在推导过程中，我们做以下的规定：

对于多元高斯分布 $X \sim N(\mu, \Sigma)$，pdf为：

$$P(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

其中，$x=\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$，$\mu=\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$，$\Sigma=\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$，$\Sigma$ 一般为正定矩阵或半正定矩阵（这里假设 Σ 为正定矩阵）

3.1 马氏距离

多元高斯分布中 $(x-\mu)^T\Sigma^{-1}(x-\mu)$ 部分其实是一个马氏距离，其值是一个数。当 Σ^{-1} 是一个单位矩阵时，马氏距离即为欧式距离：

对于一个多元高斯分布来说，其 μ、Σ 已固定，因此其pdf是只关于 x 的函数。

因此我们重点关注 $(x-\mu)^T\Sigma^{-1}(x-\mu)$ 这一部分即可。

$\underbrace{(x-\mu)^T}_{1\times p}\underbrace{\Sigma^{-1}}_{p\times p}\underbrace{(x-\mu)}_{p\times 1}$ 的结果是一个数，这个数就是马氏距离。

特别地，当 $\Sigma^{-1}=I$，马氏距离其实就是欧式距离：

$$z_1=\begin{pmatrix} z_{11} \\ z_{12} \end{pmatrix},\ z_2=\begin{pmatrix} z_{21} \\ z_{22} \end{pmatrix},\ (z_1-z_2)^T I^{-1}(z_1-z_2)=(z_{11}-z_{21}, z_{12}-z_{22})\begin{pmatrix} z_{11}-z_{21} \\ z_{12}-z_{22} \end{pmatrix}=(z_{11}-z_{21})^2+(z_{12}-z_{22})^2$$

3.2 $(x-\mu)^T\Sigma^{-1}(x-\mu)$ 的求解

先看看 Σ 部分，对实对称矩阵进行特征值分解，有：

$\Sigma = U\Lambda U^T$（U 是特征向量作列向量组成的正交阵，因此有 $U^T=U^{-1}$，Λ 是特征值组成的对角阵）

$$= \underbrace{(U_1, U_2, \cdots, U_p)}_{p\times p}\underbrace{\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{pmatrix}}_{p\times p}\underbrace{\begin{pmatrix} U_1^T \\ U_2^T \\ \vdots \\ U_p^T \end{pmatrix}}_{p\times p}$$

$$= \underbrace{(\lambda_1 U_1, \lambda_2 U_2, \cdots, \lambda_p U_p)}_{p\times p}\underbrace{\begin{pmatrix} U_1^T \\ U_2^T \\ \vdots \\ U_p^T \end{pmatrix}}_{p\times p}$$

$$= \sum_{i=1}^{p} U_i \lambda_i U_i^T \quad p\times p$$

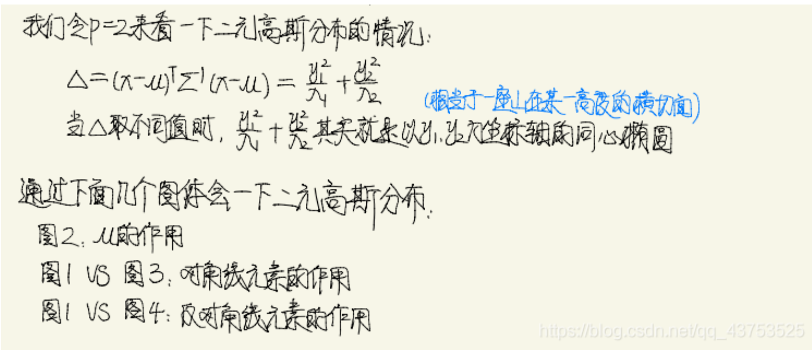
$$\therefore \Sigma^{-1} = \sum_{i=1}^{p}(U_i^T)^{-1}\lambda_i^{-1}U_i^{-1} = \sum_{i=1}^{p} U_i \frac{1}{\lambda_i} U_i^T$$

$$\therefore \overbrace{(x-\mu)^T\Sigma^{-1}(x-\mu)}^{1\times 1} = \sum_{i=1}^{p} \overbrace{(x-\mu)^T U_i}^{1\times p} \overbrace{\frac{1}{\lambda_i}}^{p\times p} \overbrace{U_i^T(x-\mu)}^{p\times 1}$$

令 $y_i=(x-\mu)^T U_i$，则上式 $= \sum_{i=1}^{p} \overset{1\times p}{y_i} \frac{1}{\lambda_i} \overset{p\times 1}{y_i^T}$

特别地，令 p = 2，我们来看一下二元高斯分布的图像情况，并通过几个图来了解一下均值和协方差矩阵对图像分布的影响：

Zicon in广工 关注 0


我们令p=2来看一下二元高斯分布的情况：

$$\Delta=(x-\mu)^T\Sigma^{-1}(x-\mu)=\frac{y_1^2}{\lambda_1}+\frac{y_2^2}{\lambda_2}$$ （相当于一座山在某一高度的横切面）

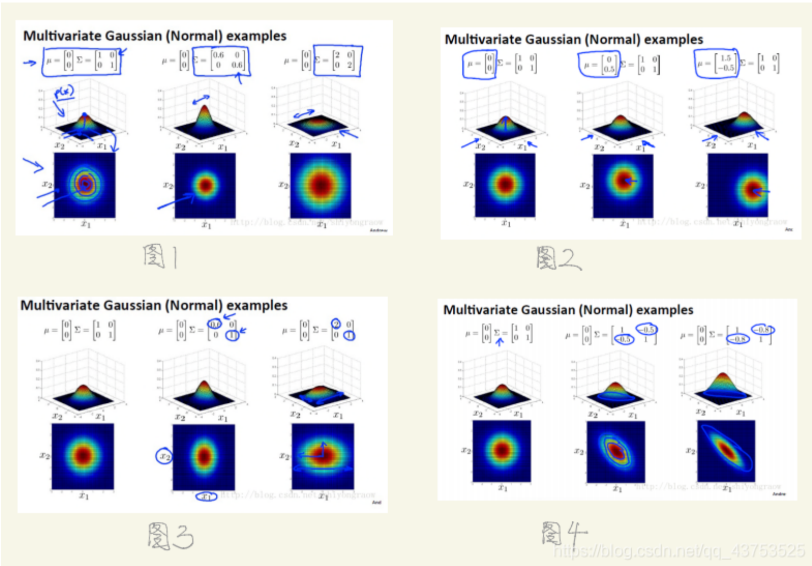
当Δ取不同值时，$\frac{y_1^2}{\lambda_1}+\frac{y_2^2}{\lambda_2}$ 其实就是以 u_1,u_2 为坐标轴的同心椭圆

通过下面几个图体会一下二元高斯分布：

图2：μ的作用

图1 VS 图3：对角线元素的作用

图1 VS 图4：反对角线元素的作用

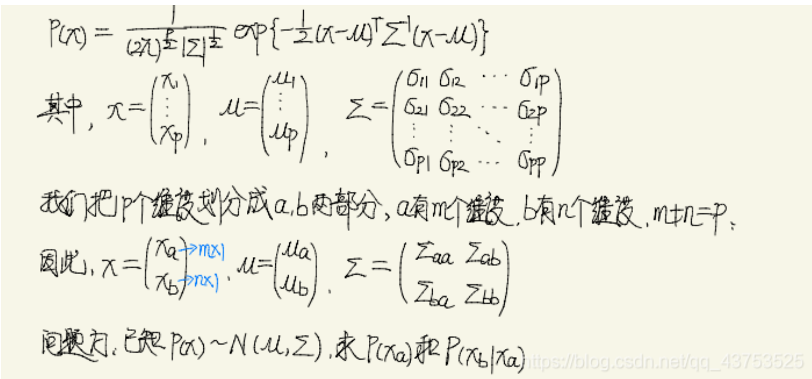


3.3 实际应用过程的问题

1. 实际应用中，协方差矩阵 Σ 有 $\frac{p(p+1)}{2}$ 个自由参数，复杂度为 $O(p^2)$，因此在高维时常常假设 Σ 为对角阵
2. 单个高斯分布的拟合能力是有限的，因此后续引入了高斯混合模型等模型

四、联合分布 -> 边缘分布 + 条件分布

4.1 数据说明


$$P(x)=\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

其中，$x=\begin{pmatrix}x_1\\ \vdots\\ x_p\end{pmatrix}$，$\mu=\begin{pmatrix}\mu_1\\ \vdots\\ \mu_p\end{pmatrix}$，$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp}\end{pmatrix}$

我们把p个维度划分成a,b两部分，a有m个维度，b有n个维度，m+n=p.

因此，$x=\begin{pmatrix}x_a \to m(x)\\ x_b \to n(x)\end{pmatrix}$，$\mu=\begin{pmatrix}\mu_a\\ \mu_b\end{pmatrix}$，$\Sigma=\begin{pmatrix}\Sigma_{aa} & \Sigma_{ab}\\ \Sigma_{ba} & \Sigma_{bb}\end{pmatrix}$

问题为：已知 $P(x)\sim N(\mu,\Sigma)$，求 $P(x_a)$ 和 $P(x_b|x_a)$

4.2 关键推论

在后面的推导中，我们会经常用到下面这个推理，此处证明略：

$$\text{若已知}\begin{cases} x\sim N(\mu,\Sigma) \\ y=Ax+B \end{cases} \Longrightarrow y\sim N(A\mu+B, A\Sigma A^T)$$

4.3 求解边缘分布 $P(x_a)$

首先求解边缘概率 $P(x_a)$:

$$\underbrace{x_a}_{y}=\underbrace{(I_m,O_n)}_{A}\underbrace{\begin{pmatrix}x_a\\x_b\end{pmatrix}}_{x}$$

因此根据4.2的推论，有：

$$E(x_a)=\overbrace{(I_m,O_n)}^{A}\overbrace{\begin{pmatrix}\mu_a\\\mu_b\end{pmatrix}}^{\mu}=\mu_a,$$

$$Var(x_a)=(I_m,O_n)\begin{pmatrix}\Sigma_{aa}&\Sigma_{ab}\\\Sigma_{ba}&\Sigma_{bb}\end{pmatrix}\begin{pmatrix}I_m\\O_n\end{pmatrix}=(\Sigma_{aa}\ \Sigma_{ab})\begin{pmatrix}I_m\\O_n\end{pmatrix}=\Sigma_{aa}$$

$$\boxed{\text{综上},\ x_a\sim N(\mu_a,\Sigma_{aa})}$$

4.4 求解条件分布 $P(x_b|x_a)$

使用配方法也可以推导出条件分布$P(x_b|x_a)$，但在这里我们使用巧妙且比较简单的构造法进行推导：

我们构造如下函数：

$$\begin{cases} x_{b\cdot a}=x_b-\Sigma_{ba}\Sigma_{aa}^{-1}x_a \\ \mu_{b\cdot a}=\mu_b-\Sigma_{ba}\Sigma_{aa}^{-1}\mu_a \\ \Sigma_{bb\cdot a}=\Sigma_{bb}-\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \end{cases}$$

求 $P(x_b|x_a)$ 时，x_a 作为已知常量，只需知道 $x_{b\cdot a}\sim N(\mu_{b\cdot a},\Sigma_{bb\cdot a})$ 就能基于推论求 $P(x_b|x_a)$

$$\underbrace{\overset{n\times1}{x_{b\cdot a}}}_{y}=\underbrace{(-\overset{n\times m}{\Sigma_{ba}}\overset{m\times m}{\Sigma_{aa}^{-1}},I_{n\times n})}_{A}\underbrace{\begin{pmatrix}\overset{m\times1}{x_a}\\\overset{n\times1}{x_b}\end{pmatrix}}_{x}$$

$$\therefore E(x_{b\cdot a})=(-\Sigma_{ba}\Sigma_{aa}^{-1}\ \ I_{n\times n})\begin{pmatrix}\mu_a\\\mu_b\end{pmatrix}=-\Sigma_{ba}\Sigma_{aa}^{-1}\mu_a+\mu_b=\mu_{b\cdot a}$$

$$Var(x_{b\cdot a})=(-\Sigma_{ba}\Sigma_{aa}^{-1}\ \ I_{n\times n})\begin{pmatrix}\Sigma_{aa}&\Sigma_{ab}\\\Sigma_{ba}&\Sigma_{bb}\end{pmatrix}\begin{pmatrix}-\Sigma_{aa}^{-1}\Sigma_{ba}^T\\I_{n\times n}\end{pmatrix}$$

$$=(-\Sigma_{ba}+\Sigma_{ba},\ -\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}+\Sigma_{bb})\begin{pmatrix}-\Sigma_{aa}^{-1}\Sigma_{ba}^T\\I_{n\times n}\end{pmatrix}$$

$$=\Sigma_{bb\cdot a}$$

$$\therefore x_{b\cdot a}\sim N(\mu_{b\cdot a},\Sigma_{bb\cdot a})$$

由 $x_b=\overset{x}{x_{b\cdot a}}+\overset{B}{\Sigma_{ba}\Sigma_{aa}^{-1}x_a}$ 得：

$$E(x_b|x_a)=\mu_{b\cdot a}+\Sigma_{ba}\Sigma_{aa}^{-1}x_a,$$

$$Var(x_b|x_a)=\Sigma_{bb\cdot a}$$

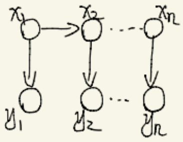
综上，

$$\boxed{\begin{cases} P(x_a)=N(\mu_a,\Sigma_{aa}) \\ P(x_b|x_a)=N(\mu_{b\cdot a}+\Sigma_{ba}\Sigma_{aa}^{-1}x_a,\Sigma_{bb\cdot a}) \end{cases} \xrightarrow{\text{对称性}} \begin{cases} P(x_b)=N(\mu_b,\Sigma_{bb}) \\ P(x_a|x_b)=N(\mu_{a\cdot b}+\Sigma_{ab}\Sigma_{bb}^{-1}x_b,\Sigma_{aa\cdot b}) \end{cases}}$$

五、线性高斯系统

5.1 问题介绍

假设有一个线性高斯动态系统:



$x_{t+1}=Fx_t+h+\varepsilon,\ \varepsilon\sim N(0,\Lambda^{-1})$

$y_{t+1}=Ax_{t+1}+b+\epsilon,\ \epsilon\sim N(0,L^{-1})$

$y|x$ 表示 x 已确定，y 的波动仅看 ϵ

即已知 $P(x)=N(\mu,\Lambda^{-1})$，$P(y|x)=N(Ax+b,L^{-1})$，求 $P(y)$ 和 $P(x|y)$

5.2 求解 $P(y)$

先求解 $P(y)$:

$$y=Ax+b+\epsilon,\ \epsilon\sim N(0,L^{-1})$$

$$E(y)=E(Ax+b+\epsilon)\overset{4.2推论}{=}E(Ax+b)+E(\epsilon)=A\mu+b,$$

$$Var(y)=Var(Ax+b+\epsilon)\overset{4.2推论}{=}Var(Ax+b)+Var(\epsilon)=A\Lambda^{-1}A^T+L^{-1}$$

$$\boxed{综上,\ P(y)=N(A\mu+b,\ A\Lambda^{-1}A^T+L^{-1})}$$

5.3 求解 $P(x|y)$

接下来求解 $P(x|y)$:

①我们令 $z=\begin{pmatrix}x\\y\end{pmatrix}$，求 $P(z)$

②由4.3的结论即可求得 $P(x|y)$

①: $z=\begin{pmatrix}x\\y\end{pmatrix}\sim N\left(\begin{pmatrix}\mu\\A\mu+b\end{pmatrix},\begin{pmatrix}\Lambda^{-1}&\Delta\\\Delta&A\Lambda^{-1}A^T+L^{-1}\end{pmatrix}\right)$

其中，

$$\begin{aligned}\Delta=cov(x,y)&=E[(x-E(x))(y-E(y))^T]\\&=E[(x-\mu)(Ax+b+\varepsilon-A\mu-b)^T]\\&=E[(x-\mu)(Ax-A\mu+\varepsilon)^T]\\&=E[(x-\mu)(Ax-A\mu)^T]+E[(x-\mu)\varepsilon^T]\\&=E[(x-\mu)(x-\mu)^TA^T]\\&=Var(x)\cdot A^T\\&=\Lambda^{-1}A^T\end{aligned}$$

$x\perp\varepsilon$，因此 $x-\mu\perp\varepsilon$

$\therefore E[(x-\mu)\varepsilon^T]=E[x-\mu]E[\varepsilon^T]$，$E[(x-\mu)\varepsilon^T]=0$，$E[\varepsilon^T]=0$

② 由4.3可知: $P(x_a|x_b)=N(\mu_{a\cdot b}+\Sigma_{ab}\Sigma_{bb}^{-1}x_b,\ \Sigma_{aa\cdot b})$

$$\begin{aligned}E[x|y]&=\mu_a-\Sigma_{ab}\Sigma_{bb}^{-1}\mu_b+\Sigma_{ab}\Sigma_{bb}^{-1}x_b\\&=\mu-\Lambda^{-1}A^T(L^{-1}+A\Lambda^{-1}A^T)^{-1}(A\mu+b)+\Lambda^{-1}A^T(L^{-1}+A\Lambda^{-1}A^T)^{-1}y\\&=\mu+\Lambda^{-1}A^T(L^{-1}+A\Lambda^{-1}A^T)^{-1}(y-A\mu-b)\end{aligned}$$

$$Var[x|y]=\Sigma_{aa\cdot b}=\Sigma_{aa}-\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}=\Lambda^{-1}-\Lambda^{-1}A^T(L^{-1}+A\Lambda^{-1}A^T)^{-1}A\Lambda^{-1}$$

$$\boxed{综上,\ P(x|y)=N\left(\mu+\Lambda^{-1}A^T(L^{-1}+A\Lambda^{-1}A^T)^{-1}(y-A\mu-b),\ \Lambda^{-1}-\Lambda^{-1}A^T(L^{-1}+A\Lambda^{-1}A^T)^{-1}A\Lambda^{-1}\right)}$$

六、参考资料

1. 哔哩哔哩白板推导系列视频
2. 机器学习第九周（三）–多元高斯分布

目录 0. 前言 1. 高斯分布（Gaussian distribution） 2. 参数估计 3. 异常检测算法（原始模型） 4. ...

【机器学习】高斯分布 zxiaohui666 358
1.单变量高斯分布单变量高斯分布概率密度函数定义为：p(x)=12πσ−−−√exp{−12(x−μσ)2}(1.1)...

【机器学习 基本概念】高斯分布_魏晓蕾的博客_高斯分布... 4-6
今天,我要介绍我们早就知道的一种分布,它叫做高斯分布。高斯分布在概率论中算是比较核心的一...

【机器学习课程】高斯分布_积流成河 8-13
【机器学习课程】高斯分布 1 概率密度函数 对于一个连续的概率密度方程,p (x) p(x)p(x),满足下...

机器学习 | 二：高斯分布 机器学习之家 908
今日话题 讲解了独立同分布的概念，高斯分布，一维高斯分布。01-独立同分布 指随机过程...

机器学习--正态分布 weixin_33973609的博客 534
2019独角兽企业重金招聘Python工程师标准>>> ...

机器学习概率知识补充——高斯分布(1)_guozhi Tang的博客 3-21
在机器学习中,会接触到大量的概率分布,掌握这些概率分布的数学原理是十分重要的,下面介绍其...

【机器学习系列】高斯分布-最大似然估计求解_【AI机器... 3-3
公众号:AI机器学习与知识图谱 研究方向:自然语言处理与知识图谱 前言:机器学习系列文章常含有...

正态分布为何如此重要? AI科技大本营 5378
作者 | Farhad Malik译者 | Monanfei责编 | 夕颜出品 | AI科技大本营（ID: rgznai100）为什么正态...

python gmm em算法 2维数据_高斯混合模型(GMM) weixin_39768247的博客 220
一个例子高斯混合模型（Gaussian Mixed Model）指的是多个高斯分布函数的线性组合，理论上...

CS229《机器学习》笔记 | 多元高斯分布的边缘分布和条件分布 3-22
吴恩达的《机器学习(CS229)》Lecture note 9(Part X Factor analysis)中提及了多元高斯分布的边...

机器学习中用来防止过拟合的方法有哪些? T_Janey 354
机器学习中用来防止过拟合的方法有哪些

高斯拟合原理_关于拟合 weixin_39613824的博客 340
首先，关于拟合的文章，刘博：深度学习（Deep Learning）基础概念8：L2正则化（L2 Regulari...

高斯(Gaussian)拟合的实现 Du_Shuang的博客 1万+
转载于：https://blog.csdn.net/c914620529/article/details/50393238 高斯拟合(Gaussian Fitting)...

【Python】机器学习笔记10-高斯混合模型（Gaussian ... weixin_41429999的博客 2277
本文的参考资料：《Python数据科学手册》； 本文的源代上传到了Gitee上； 本文用到的包： %...

【机器学习基础】概率分布之高斯分布 天堂的鸽子 920
本系列为《模式识别与机器学习》的读书笔记。 一，多元高斯分布 考虑高斯分布的几何形式，高...

机器学习基础专题：高斯分布 八门金锁的技术博客 59
记号和术语 $X \in R^{N*p} = (x_1, x_2, ..., x_N)^T $ xi=(xi1,...,xip)Tx_i = (x_{i1}, ..., x_{ip})^Txi=(...

机器学习——高斯过程 qq_27436347的博客 700
高斯过程 所谓高斯，即高斯分布 所谓过程，即随机过程 高斯分布 一维高斯 p(x)=N(μ,σ2)p(x)=N...

认识高斯分布 ACdreamer 1万+
今天，我要介绍我们早就知道的一种分布，它叫做高斯分布。高斯分布在概率论中算是比较核心...

从零实现机器学习算法（十九） 高斯混合模型 **最新发布** 学如不及,犹恐失之 158
01. 高斯混合模型简介 高斯混合模型（Gaussian Mixed Model，GMM）和隐马尔可夫模型(Hidde...

高斯拟合原理_高斯过程回归(GPR) weixin_39714275的博客 1932
1.高斯过程是定义在连续域上的无限多个服从高斯分布的随机变量所组成的随机过程2.高斯过程...

机器学习中的高斯过程 **热门推荐** lotus 3万+
转自：http://www.datalearner.com/blog/1051459170229238 关于高斯过程，其实网上已经...

机器学习知识点(二十二)高斯分布（正态分布）基础知识 医疗影像检索 7115
1、概念 正态分布（Normal distribution）又名高斯分布（Gaussian distribution），是一个在数学...

高斯分布机器学习实践项目 人工智能讲师团 624
多元高斯分布与一元高斯分布的关系 首先一维标准正态分布： 二维标准正态分布，就是两个独立...

【机器学习】高斯分布为什么普遍和常用? artzers的专栏 1万+
似然函数到高斯分布为了得到精确值，我们需要

热门文章

简洁易懂:利用栈实现二叉树的前序、中序、后序遍历 2815

多项式朴素贝叶斯算法 1420

手把手教你如何把jupyter notebook切换到其他配置好的conda虚拟环境 1149

linux服务器上遇到UnicodeEncodeError: 'ascii' codec can't encode characters in ordinal not in range(128)的问题 741

EM算法总结:从 ELBO + KL散度出发 652

分类专栏

- 笔记 3篇
- 深度学习 6篇
- 因果推断 2篇
- 数据结构 2篇
- 推荐系统 3篇
- 数据分析

最新评论

逻辑回归公式推导
涛哥的AI之路: 对数似然应该时ln吧,log话求导就成了1/h(x)ln10了,而不是1/h(x)

线性回归公式推导总结
涛哥的AI之路: 您好,就是梯度下降的推导中最后乘以Xi是怎么推导导的呀

目录

文章目录

一、高斯分布的概率密度函数

二、一元高斯分布的极大似然估计

2.1 μ M L E , σ M L E \mu_{MLE},\si…

2.2 验证 μ M L E , σ M L E \mu_{ML…

三、多元高斯分布

3.1 马氏距离

3.2 (x − μ) T Σ − 1 (x − μ) (x-\mu)…

3.3 实际应用过程的问题

四、联合分布 -> 边缘分布 + 条件分布

4.1 数据说明

4.2 关键推论

4.3 求解边缘分布 P (x a) P(x_a) P(…

4.4 求解条件分布 P (x b ∣ x a) P(x…