高斯过程（Gaussian Processes）原理

原创 ting_qifengl 于 2021-11-30 19:12:48 发布 18488 收藏 166 版权
分类专栏： 机器学习 文章标签： 机器学习

机器学习 专栏收录该内容 3 订阅 1 篇文章 订阅专栏

高斯过程（Gaussian Processes, GP）是概率论和数理统计中随机过程的一种，是多元高斯分布 的扩展，被应用于机器学习、信号处理等领域。博主在阅读了数篇文章和博客后才算是基本搞懂了GP的原理，特此记录。本文目前暂对高斯过程的公式推导和高斯过程回归原理及其优缺点进行讲解和阐述，后续根据个人学习进度再更新源码等内容。

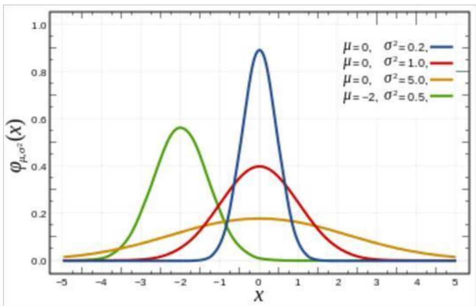
一、一维高斯分布

我们从最简单最常见的一维高斯分布开始。

众所周知，一维高斯分布，又叫一维正态分布 的概率密度函数为：

$$p\left(x\right)=\frac{1}{\sigma\sqrt{2\pi}}exp\left(\frac{-\left(x-\mu\right)^2}{2\sigma^2}\right)\ (1)$$

式中，μ表示均值，σ表示方差，均值和方差唯一的决定了曲线的形状。当μ为0，σ为1时称为标准正态分布。



Normalized Gaussian curves with expected value μ and variance σ^2. The corresponding parameters are $a=\frac{1}{\sigma\sqrt{2\pi}}$, $b=\mu$ and $c=\sigma$.

https://blog.csdn.net/paulfeng20171114

二、多维高斯分布

从一维高斯分布推广到多维高斯分布。

假设**各维度之间相互独立**，则其概率密度函数为：

$$p\left(x_1,x_2,...,x_n\right)=\prod_{i=1}^{n}p\left(x_i\right)=\frac{1}{\left(2\pi\right)^{\frac{n}{2}}\sigma_1\sigma_2...\sigma_n}exp\left(-\frac{1}{2}\left[\frac{\left(x_1-\mu_1\right)^2}{\sigma_1^2}+\frac{\left(x_2-\mu_2\right)^2}{\sigma_2^2}+...+\frac{\left(x_n-\mu_n\right)^2}{\sigma_n^2}\right]\right)\ (2)$$

其中，$\mu_1,\mu_2,\cdots$ 分别是第一维、第二维...的均值，$\sigma_1,\sigma_2,\cdots$ 分别是第一维、第二维...的方差。

使用向量和矩阵表示上式，令

$$\vec{x}-\vec{\mu}=\left[x_1-\mu_1,x_2-\mu_2,\cdots,x_n-\mu_n\right]^T\ (3)$$

$$K=\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_n^2\end{bmatrix}\ (4)$$

则有

$$\sigma_1\sigma_2\cdots\sigma_n=\left|K\right|^{\frac{1}{2}}\ (5)$$

$$\frac{\left(x_1-\mu_1\right)^2}{\sigma_1^2}+\frac{\left(x_2-\mu_2\right)^2}{\sigma_2^2}+\cdots+\frac{\left(x_n-\mu_n\right)^2}{\sigma_n^2}=\left(\vec{x}-\vec{\mu}\right)^TK^{-1}\left(\vec{x}-\vec{\mu}\right)\ (6)$$

将公式（5）和公式（6）代入公式（2）可得

ting_qifengl 关注 42 166 0 专栏

$$p(\vec{x}) = (2\pi)^{-\frac{n}{2}} |K|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T K^{-1} (\vec{x}-\vec{\mu})\right) (7)$$

其中，$\vec{\mu} \in \mathbb{R}^n$是均值向量，$K \in \mathbb{R}^{n \times n}$是协方差矩阵，由于我们假设了各维度之间是相互独立的，所以K是一个对角矩阵。当各维度变量相关时，上式的形式仍然一致，但此时协方差矩阵 不再是对角矩阵，只具备半正定和对称的性质。上式通常也简写为：

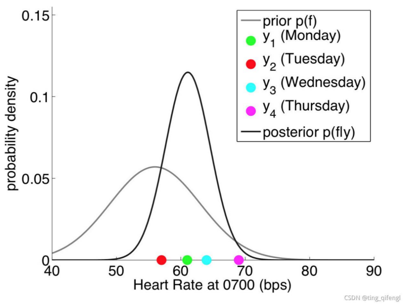
$$x \sim N(\vec{\mu}, K) (8)$$

三、高斯过程

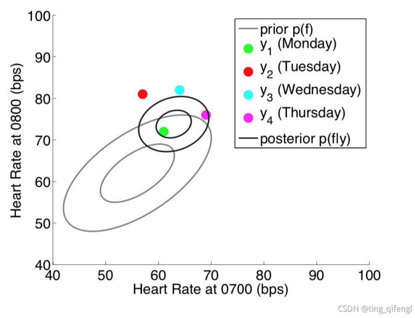
在概率论和统计学中，高斯过程(Gaussian process, GP)是观测值出现在一个连续域(例如时间和空间)的随机过程。**在高斯过程中，连续输入空间中每一点都是与一个正态分布的随机变量相关联。**此外，这些随机变量的每个有限集合都有一个多维正态分布，换句话说他们的任意有限线性组合是一个正态分布。高斯过程的分布是所有(无限多个)随机变量的联合分布，正因如此，它是连续域(例如时间和空间)上的函数分布。

下面通过一个例子来了解一下高斯过程。

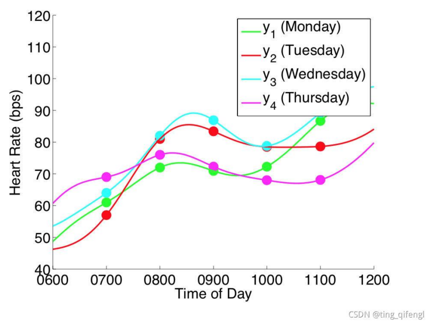
假设我们在周一到周四每天的 7:00 测试了 4 次心率，如下图中 4 个点，可能的高斯分布如图所示（瘦高的那条）。这是一个一维高斯分布，只有每天 7: 00 的心率这个维度。



现在不仅在每天的 7: 00 测心率（横轴），在 8:00 时也进行测量（纵轴），这个时候变成两个维度（二维高斯分布），如下图所示。



更进一步，如果我们在每天的无数个时间点都进行测量，则变成了下图的情况。注意下图中把测量时间作为横轴，则每个颜色的一条线代表一个（无限个时间点的测量）无限维的采样。**当对每次对无限维进行采样得到无限多个点时，其实可以理解为我们采样得到了一个函数。**



当从函数的视角去看待采样，理解了每次采样无限维相当于采样一个函数之后，原本的概率密度函数就不再是点的分布 ，而变成了**函数的分布**。这个连续的无限维高斯分布就是高斯过程。用偏理论的话语表述就是，对于所有$\vec{x} = [x_1, x_2, \cdots, x_n]$，$f(\vec{x}) = [f(x_1), f(x_2), \cdots, f(x_n)]$都服从多元高斯分布，则称$f$是一个高斯过程，表示为：

ting_qifengl 关注 42 166 0 专栏

$$f(\vec{x}) \sim N(\mu(x), K(x, x)) \text{ (9)}$$

一个高斯过程被一个均值函数和协方差函数唯一地定义，并且**一个高斯过程的有限维度的子集都服从一个多元高斯分布**（为了方便理解，可以想象二元高斯分布两个维度各自都服从一个高斯分布）。

同时，**高斯过程的协方差函数就是其核函数。**核函数是高斯过程的核心，它决定了一个高斯过程的性质。核函数在高斯过程中起衡量任意两个点之间的“距离”的作用。不同的核函数有不同的衡量方法，得到的高斯过程的性质也不一样。最常用的一个核函数为高斯核函数，也就是径向基函数 RBF。其基本形式如下，其中σ和l是高斯核的超参数。

$$K(x_i, x_j) = \sigma^2 exp\left(-\frac{\left\|x_i - x_j\right\|_2^2}{2l^2}\right)$$

x_i和x_j表示高斯过程连续域上的两个不同的时间点，$\left\|x_i - x_j\right\|^2$是一个二范式，简单点说就是表示$x_i$和$x_j$之间的距离。

一些常用的核函数如下：

Usual covariance functions [edit]

There are a number of common covariance functions:[9]

- Constant : $K_{\mathrm{C}}(x, x') = C$
- Linear: $K_{\mathrm{L}}(x, x') = x^T x'$
- white Gaussian noise: $K_{\mathrm{GN}}(x, x') = \sigma^2 \delta_{x,x'}$
- Squared exponential: $K_{\mathrm{SE}}(x, x') = \exp\left(-\frac{|d|^2}{2\ell^2}\right)$
- Ornstein–Uhlenbeck: $K_{\mathrm{OU}}(x, x') = \exp\left(-\frac{|d|}{\ell}\right)$
- Matérn: $K_{\mathrm{Matern}}(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}|d|}{\ell}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}|d|}{\ell}\right)$
- Periodic: $K_{\mathrm{P}}(x, x') = \exp\left(-\frac{2\sin^2\left(\frac{d}{2}\right)}{\ell^2}\right)$
- Rational quadratic: $K_{\mathrm{RQ}}(x, x') = (1 + |d|^2)^{-\alpha}, \quad \alpha \geq 0$

四、高斯过程回归

高斯过程回归是使用高斯过程先验对数据进行回归分析的非参数模型（指系统的数学模型中非显式地包含可估参数）。简单来讲可以看作是根据先验+观测值推出后验的过程。

根据均值$\mu(x)$和协方差$K(x_i, x_j)$，我们可以定义一个高斯过程，但是此时并没有任何的观测值，是一个先验。在获得了一组观测值之后，可以根据观测值修正这个高斯过程的均值函数和核函数。

设S为训练集，训练数据独立同分布，分布未知，我们定义高斯过程回归模型(Gaussian Process Regression，以下简称GPR)的表达式为：

$$y^{(i)} = f(x^{(i)}) + \varepsilon^{(i)}, \qquad i = 1, \ldots, m$$

其中ε_i是独立同分布的噪声变量，服从$N(0, \sigma^2)$。

在GPR中，我们假设f服从均值为0的GP：

$$f(\cdot) \sim \mathcal{GP}(0, k(\cdot, \cdot))$$

知乎 @Light Sea

设T是测试集，独立同分布且和S的分布相同。

定义：

$$X = \begin{bmatrix} — (x^{(1)})^T — \\ — (x^{(2)})^T — \\ \vdots \\ — (x^{(m)})^T — \end{bmatrix} \in \mathbf{R}^{m \times n} \quad \vec{f} = \begin{bmatrix} f(x^{(1)}) \\ f(x^{(2)}) \\ \vdots \\ f(x^{(m)}) \end{bmatrix}, \quad \vec{\varepsilon} = \begin{bmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \\ \vdots \\ \varepsilon^{(m)} \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} \in \mathbf{R}^m,$$

$$X_* = \begin{bmatrix} — (x_*^{(1)})^T — \\ — (x_*^{(2)})^T — \\ \vdots \\ — (x_*^{(m_*)})^T — \end{bmatrix} \in \mathbf{R}^{m_* \times n} \quad \vec{f}_* = \begin{bmatrix} f(x_*^{(1)}) \\ f(x_*^{(2)}) \\ \vdots \\ f(x \end{bmatrix}, \quad \vec{\varepsilon}_* = \begin{bmatrix} \varepsilon_*^{(1)} \\ \varepsilon_*^{(2)} \\ \vdots \end{bmatrix}, \quad \vec{y}_* = \begin{bmatrix} y_*^{(1)} \\ y_*^{(2)} \\ \vdots \end{bmatrix} \in \mathbf{R}^{m_*}$$

ting_qifengl 关注 42 166 0 专栏

我们的任务就是计算posterior predictive distribution $P\left(y^*|x^*,S\right)$，既然f满足GP，噪声ϵ也满 足高斯分布，根据高斯分布性质可知，这个分布一定是高斯分布。

根据高斯过程的性质，有:

$$\left[\begin{matrix}\vec{f}\\ \vec{f}_*\end{matrix}\right]\Bigg| X, X_* \sim \mathcal{N}\left(\vec{0}, \left[\begin{matrix}K(X,X) & K(X,X_*)\\ K(X_*,X) & K(X_*,X_*)\end{matrix}\right]\right),$$

知乎 @Light Sea

其中:

$$\begin{aligned}
\vec{f} \in \mathbf{R}^m \text{ such that } \vec{f} &= \left[f(x^{(1)}) \quad \cdots \quad f(x^{(m)})\right]^T\\
\vec{f}_* \in \mathbf{R}^{m_*} \text{ such that } \vec{f}_* &= \left[f(x_*^{(1)}) \quad \cdots \quad f(x_*^{(m)})\right]^T\\
K(X,X) \in \mathbf{R}^{m\times m} \text{ such that } (K(X,X))_{ij} &= k(x^{(i)},x^{(j)})\\
K(X,X_*) \in \mathbf{R}^{m\times m_*} \text{ such that } (K(X,X_*))_{ij} &= k(x^{(i)},x_*^{(j)})\\
K(X_*,X) \in \mathbf{R}^{m_*\times m} \text{ such that } (K(X_*,X))_{ij} &= k(x_*^{(i)},x^{(j)})\\
K(X_*,X_*) \in \mathbf{R}^{m_*\times m_*} \text{ such that } (K(X_*,X_*))_{ij} &= k(x_*^{(i)},x_*^{(j)}).
\end{aligned}$$

知乎 @Light Sea

同样的，对噪声也有类似的结论:

$$\left[\begin{matrix}\vec{\varepsilon}\\ \vec{\varepsilon}_*\end{matrix}\right] \sim \mathcal{N}\left(\vec{0}, \left[\begin{matrix}\sigma^2 I & \vec{0}\\ \vec{0}^T & \sigma^2 I\end{matrix}\right]\right).$$

知乎 @Light Sea

因为两个独立多元高斯变量的和还是多元高斯变量，因此有:

$$\left[\begin{matrix}\vec{y}\\ \vec{y}_*\end{matrix}\right]\Bigg| X, X_* = \left[\begin{matrix}\vec{f}\\ \vec{f}_*\end{matrix}\right] + \left[\begin{matrix}\vec{\varepsilon}\\ \vec{\varepsilon}_*\end{matrix}\right] \sim \mathcal{N}\left(\vec{0}, \left[\begin{matrix}K(X,X)+\sigma^2 I & K(X,X_*)\\ K(X_*,X) & K(X_*,X_*)+\sigma^2 I\end{matrix}\right]\right)$$

根据多元高斯分布的性质(性质3)，可以看出，y^*满足多元高斯分布:

$$\vec{y_*} \mid \vec{y}, X, X_* \sim \mathcal{N}(\mu^*, \Sigma^*)$$

利用这个性质有:

$$\begin{aligned}
\mu^* &= K(X_*,X)\left(K(X,X)+\sigma^2 I\right)^{-1}\vec{y}\\
\Sigma^* &= K(X_*,X_*)+\sigma^2 I - K(X_*,X)\left(K(X,X)+\sigma^2 I\right)^{-1}K(X,X_*).
\end{aligned}$$

上式的具体推导可以参考此链接。

这样就完成了预测。

五、高斯过程回归算法的优缺点

- 优点
  - （采用 RBF 作为协方差函数）具有平滑性质，能够拟合非线性数据;
  - 高斯过程回归天然支持得到模型关于预测的不确定性（置信区间），直接输出关于预测点值的概率分布;
  - 通过最大化边缘似然这一简洁的方式，高斯过程回归可以在不需要交叉验证的情况下给出比较好的正则化效果。
- 缺点
  - 高斯过程是一个非参数模型，每次的 inference 都需要对所有的数据点进行（矩阵求逆）。对于没有经过任何优化的高斯过程回归，n个样本点时间复杂度大概是$O\left(n^3\right)$，空间复杂度是$O\left(n^2\right)$，在数据量大的时候高斯过程变得 intractable;
  - 高斯过程回归中，先验是一个高斯过程，likelihood 也是高斯的，因此得到的后验仍是高斯过程。在 likelihood 不服从高斯分布的i approximate 使其仍为高斯过程;

ting_qifengl 关注 42 166 0 专栏

- RBF 是最常用的协方差函数，但在实际中通常需要根据问题和数据的性质选择恰当的协方差函数。

六、参考链接

高斯过程 Gaussian Processes 原理、可视化及代码实现 -

高斯过程回归 - 简书

如何通俗易懂地介绍 Gaussian Process？ - 知乎

[公式推导]用最简洁的方法证明多元正态分布的条件分布 - 司徒鲜生 - 博客园

高斯过程和高斯过程回归 - 知乎

文章知识点与官方知识档案匹配，可进一步学习相关知识

OpenCV技能树 首页 概览 14808 人正在系统学习中

透彻理解高斯过程Gaussian Process (GP) 冯喆–AI工匠 6万+
透彻理解高斯过程Gaussian Process (GP) 一、整体说说 为了理解高斯过程，我们就首先需要准备一下预备知识...

万字总结复杂而奇妙的高斯过程! lsxxx2011的专栏 2223
高斯过程的理论知识非参数方法的基本思想高斯过程的基本概念高斯过程的Python实现使用Numpy手动实现使用`...

高斯过程_Phoenix Studio的博客 2-20
高斯过程定义了先验函数。观察到某些函数值后,可通过代数运算以将其转换为后验函数。 在这种情况下,连续函数...

机器学习中的随机过程（高斯过程） 最新发布 scott198512的博客 279
高斯过程的直观解释与推导、实现过程

GPR(高斯过程回归) 尔过留香的博客 791
高斯过程简单来说，就是一系列关于连续域（时间或空间）的随机变量的联合，而且其中每一个时间或空间点上...

开发一个app多少钱
app开发费用一般多少钱

高斯过程是什么? migue_math 1375
本文译自： Duane Rich 的文章Which is your favorite Machine Learning algorithm? 几乎所有的有监督机器学习算...

机器学习笔记：高斯过程 qq_40206371的博客 1277
1 高斯分布回顾 一维高斯分布 当维度上升到多维的时候（比如有限的p维），我们称之为高维高斯分布，记作： ...

高斯过程 Gaussian Process sinat_37836116的博客 3986
一、什么是高斯过程 高斯过程是一种随机过程，即按时间或者空间索引的随机变量的集合。这个集合中的有...

高斯过程是什么？从视觉上理解机器学习中的高斯过程——Gaussian Pr... 一个苦逼研究僧的博客 913
高斯过程Gaussian Process是什么?

机器学习之高斯过程 weixin_47166032的博客 3451
高斯过程(Gaussian Process) 在机器学习领域里，高斯过程是一种假设训练数据来自无限空间、并且各特征都符...

高斯拟合原理_高斯过程回归(GPR) weixin_39714275的博客 8975
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.高斯过程回归有两个视角： ...

高斯过程回归GPR 温故而知新 4万+
关键字：核函数，RBF超参数调优 对这个很熟悉了，简单写一下 本人用matlab实现了一下: https://download.csdn...

浅析高斯过程回归（Gaussian process regression） 热门推荐 qq_20195745的博客 6万+
前言 高斯过程回归的和其他回归算法的区别是：一般回归算法给定输入X，希望得到的是对应的Y值，拟合函数可...

1.7. 高斯过程(Gaussian Processes) u010016927的博客 2万+
针对机器学习的高斯过程(Gaussian Processes for Machine Learning,即 GPML) 是一个通用的监督学习方法，主...

高斯过程回归(Gaussian Processes Regression, GPR)简介 HelloWorldTM的博客 5284
一、高斯过程简介二、高斯分布1. 一元高斯分布2. 多元高斯分布三、高斯过程回归1. 高斯过程2.高斯过程回归四...

高斯过程（原理和代码实现） 渊茶 2203
1.高斯过程原理 高斯过程：每个点的观测值都是高斯分布，这里面的观测值就是输出Y，观测点的组合也符合高斯...

高斯过程的matlab程序实现及其参数优化
高斯过程matlab编程实战详解 本次结论公式均来自《Gaus

ting_qifengl 关注 42 166 0 专栏