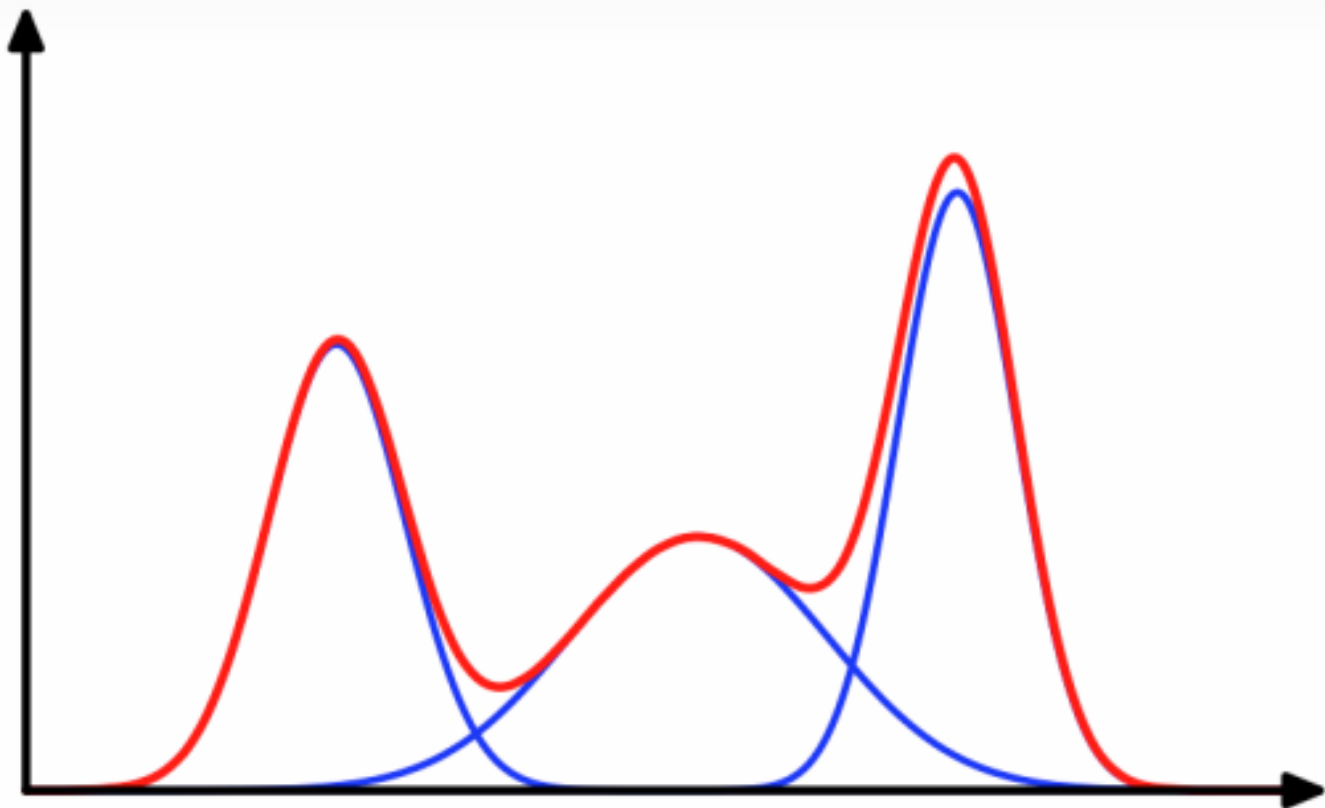


知乎

首发于  
机器学习笔记

## 高斯混合模型 (GMM)



戴文亮

[关注他](#)

1,009 人赞同了该文章

笔者最近在计算机视觉课程里接触到了高斯混合模型 (Gaussian Mixture Model)，遂写一篇笔记来整理记录相关知识点，分享给大家。欢迎讨论、指正！

### 混合模型 (Mixture Model)

混合模型是一个可以用来表示在总体分布 (distribution) 中含有  $K$  个子分布的概率模型，换句话说，混合模型表示了观测数据在总体中的概率分布，它是一个由  $K$  个子分布组成的混合分布。混合模型不要求观测数据提供关于子分布的信息，来计算观测数据在总体分布中的概率。

### 高斯模型

#### 单高斯模型

当观测数据  $x$  是  $n$  维数据 (即  $x \in \mathbb{R}^n$ ) 时，高斯分布满足以下概率密度函数 (Probability Density Function, PDF)：

已赞同 1009



63 条评论

[分享](#)[喜欢](#)[收藏](#)[申请转载](#)

其中  $\mu$  为数据均值（期望）， $\sigma$  为数据标准差（Standard deviation）。

当样本数据  $X$  是多维数据（Multivariate）时，高斯分布遵从下方概率密度函数：

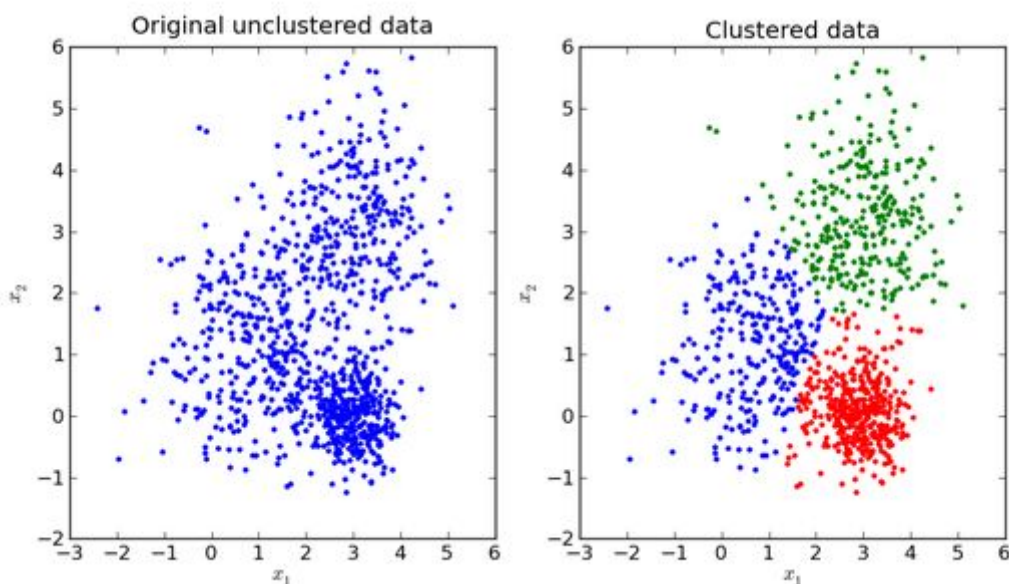
$$P(x|\theta) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2}\right)$$

其中， $\mu$  为数据均值（期望）， $\Sigma$  为协方差（Covariance）， $D$  为数据维度。

## 高斯混合模型

高斯混合模型可以看作是由  $K$  个单高斯模型组合而成的模型，这  $K$  个子模型是混合模型的隐变量（Hidden variable）。一般来说，一个混合模型可以使用任何概率分布，这里使用高斯混合模型是因为高斯分布具备很好的数学性质以及良好的计算性能。

举个不是特别稳妥的例子，比如我们现在有一组狗的样本数据，不同种类的狗，体型、颜色、长相各不相同，但都属于狗这个种类，此时单高斯模型可能不能很好的来描述这个分布，因为样本数据分布并不是一个单一的椭圆，所以用混合高斯分布可以更好的描述这个问题，如下图所示：



图中每个点都由  $K$  个子模型中的某一个生成

首先定义如下信息：

- $x_j$  表示第  $j$  个观测数据， $j = 1, 2, \dots, N$

- $\phi(x|\theta_k)$  是第  $k$  个子模型的高斯分布密度函数,  $\theta_k = (\mu_k, \sigma_k^2)$ 。其展开形式与上面介绍的单高斯模型相同
- $\gamma_{jk}$  表示第  $j$  个观测数据属于第  $k$  个子模型的概率

高斯混合模型的概率分布为:

$$P(x|\theta) = \sum_{k=1}^K \alpha_k \phi(x|\theta_k)$$

对于这个模型而言, 参数  $\theta = (\tilde{\mu}_k, \tilde{\sigma}_k, \tilde{\alpha}_k)$ , 也就是每个子模型的期望、方差 (或协方差)、在混合模型中发生的概率。

## 模型参数学习

对于单高斯模型, 我们可以用最大似然法 (Maximum likelihood) 估算参数  $\theta$  的值,

$$\theta = \operatorname{argmax}_{\theta} L(\theta)$$

这里我们假设了每个数据点都是独立的 (Independent), 似然函数由概率密度函数 (PDF) 给出。

$$L(\theta) = \prod_{j=1}^N P(x_j|\theta)$$

由于每个点发生的概率都很小, 乘积会变得极其小, 不利于计算和观察, 因此通常我们用 Maximum Log-Likelihood 来计算 (因为 Log 函数具备单调性, 不会改变极值的位置, 同时在 0-1 之间输入值很小的变化可以引起输出值相对较大的变动):

$$\log L(\theta) = \sum_{j=1}^N \log P(x_j|\theta)$$

对于高斯混合模型, Log-Likelihood 函数是:

$$\log L(\theta) = \sum_{j=1}^N \log P(x_j|\theta) = \sum_{j=1}^N \log \left( \sum_{k=1}^K \alpha_k \phi(x|\theta_k) \right)$$

(hidden variable)，因此 log 里面还有求和，对于每个子模型都有未知的  $\alpha_k, \mu_k, \sigma_k$ ，直接求导无法计算。需要通过迭代的方法求解。

## EM 算法

EM 算法是一种迭代算法，1977 年由 Dempster 等人总结提出，用于含有隐变量 (Hidden variable) 的概率模型参数的最大似然估计。

每次迭代包含两个步骤：

1. E-step: 求期望  $E(\gamma_{jk}|X, \theta)$  for all  $j = 1, 2, \dots, N$
2. M-step: 求极大，计算新一轮迭代的模型参数

这里不具体介绍一般性的 EM 算法 (通过 Jensen 不等式得出似然函数的下界 Lower bound，通过极大化下界做到极大化似然函数)，只介绍怎么在高斯混合模型里应用从而推算出模型参数。

通过 EM 迭代更新高斯混合模型参数的方法 (我们有样本数据  $x_1, x_2, \dots, x_N$  和一个有  $K$  个子模型的高斯混合模型，想要推算出这个高斯混合模型的最佳参数)：

- 首先初始化参数
- E-step: 依据当前参数，计算每个数据  $j$  来自子模型  $k$  的可能性

$$\gamma_{jk} = \frac{\alpha_k \phi(x_j | \theta_k)}{\sum_{k=1}^K \alpha_k \phi(x_j | \theta_k)}, j = 1, 2, \dots, N; k = 1, 2, \dots, K$$

- M-step: 计算新一轮迭代的模型参数

$$\mu_k = \frac{\sum_j^N (\gamma_{jk} x_j)}{\sum_j^N \gamma_{jk}}, k = 1, 2, \dots, K$$

$$\Sigma_k = \frac{\sum_j^N \gamma_{jk} (x_j - \mu_k)(x_j - \mu_k)^T}{\sum_j^N \gamma_{jk}}, k = 1, 2, \dots, K \quad (\text{用这一轮更新后的 } \mu_k)$$

$$\alpha_k = \frac{\sum_{j=1}^N \gamma_{jk}}{N}, k = 1, 2, \dots, K$$

- 重复计算 E-step 和 M-step 直至收敛 (  $\|\theta_{i+1} - \theta_i\| < \epsilon$  ,  $\epsilon$  是一个很小的正数，表示经过一次

那次。

Reference

- 1. 《统计学习方法》第九章 - EM算法及其推广——李航
- 2. [Mixture model - Wikipedia](#)
- 3. [高斯混合模型 \(GMM\) 介绍以及学习笔记](#)

编辑于 2020-03-28 22:47

「如果你喜欢，不妨支持我继续写作，谢谢」

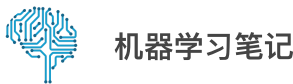
赞赏

3 人已赞赏



概率论   机器学习   统计学

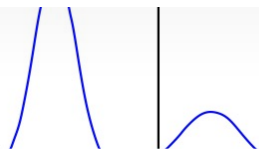
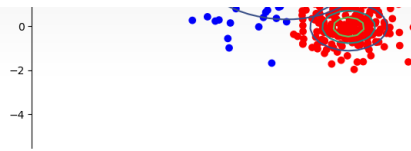
文章被以下专栏收录



机器学习笔记

推荐阅读

知乎

首发于  
机器学习笔记w/haker  
高斯混合模型(GMM)雷俊毅  
简单的混合高斯模型  
发表于广东人的机...

63 条评论

⇌ 切换为时间排序

写下你的评论...



Altriaex

2017-10-29

"如何计算高斯混合模型的参数呢？这里我们无法使用最大似然法来求导解析，因为  $K$  个高斯模型的和不是一个高斯模型，直接求导无法计算。只能通过迭代的方法求解。"

从这句话感觉你理解上似乎有点问题。

你这里还是在学习使得likelihood最大的模型参数，然后跟可导性没啥关系。

下面是我的理解。

使用像EM这样的方法，而不直接求likelihood函数的最值的原因是，likelihood函数里除了每个component的参数，还有alpha。

所以无法直接求出最值。就像你不能用一个方程解出两个未知数一样。

👍 8



戴文亮 (作者) 回复 Altriaex

2017-10-31

嗯其实我本来的理解有一点偏差，我的意思是后面是 $K$ 个高斯的和，不能像单高斯模型那样直接求导出使得likelihood最大的mu和sigma，你说得是有道理的，也就是说不知道 $x$ 属于哪一个子分布，求导的话log里还有求和，且未知参数太多，每个alpha, mu, sigma都不一样。

👍 4



WMY 回复 Altriaex

01-21

因为log里面还有求和，所以这个函数求不出导数的解析解形式。就是说对于高斯混合模型，求出参数的解析解只能通过数值法来迭代，通过最优化方法求出个近似

已赞同 1009



63 条评论

🔗 分享

❤️ 喜欢

★ 收藏

📄 申请转载



您好，我是车辆专业的，想用机器学习的方法做结构设计变量和响应之间的拟合！我的问题是：我的设计变量有离散的和连续的，而离散的变量取不同值会导致响应有较大的波动，这样用一般的拟合方法，如多项式，RBF等，精度较差！我看了您介绍的这个混合模型预测响应，很感兴趣，不知道能不能解决我这样的类似的问题，请教

👍 3



凝聚态小土豆 回复 宁静致远

2020-07-13

混合模型跟RBF有啥区别呢

👍 赞



angus zou 回复 宁静致远

01-10

离散数据类别要够多才行，不然分布不开

👍 赞



冯迁

2017-10-28

得配上代码

👍 2



Konan 回复 冯迁

2019-05-14

为什么好多人都说要配代码呢

👍 赞



冯迁 回复 Konan

2019-05-15

可能是好多人都觉得理论容易吧，数学专业转code

👍 赞

展开其他 3 条回复



堃昕或懿翹

01-10

我觉得 $\alpha$ 和 $\gamma$ 表示的意义作者需要重新好好解释一下，初学者可能都不好理解

👍 2



罗布淖尔牧羊少年

11-13

多元高斯和混合高斯分布不是一回事儿，多元高斯分布是指数据在每个维度上均成高斯分布，而混合高斯分布是多个高斯模型的混合，其分模型也可以是多元的

👍 1



西部骚大叔

2017-10-28

还有双T混合模型和高斯T分布混合模型，算法中不仅有EM算法还有模拟退火算法等

👍 1



冬炫

2020-11-26

高斯混合模型就是一种 某数据 $x$ 可能由 $K$ 个分布产生。但是每个分布产生该 $x$ 的概率又随着它所隶属的 $k$ 分布的参数 $\theta$ 所决定。各个分布目前感觉是独立的。不知道对不对？哈哈。多元高斯分布，直接可以使用一个数学公式即可求出该点产生的概率。并没有它隶属的其他信息，应该就是一个分布。

👍 1



苏鲁支 回复 冬炫

2020-12-18

您好，推导过程中用的并不是概率值，而是概率密度，这样是不是有问题呢？

👍 赞



罗布淖尔牧羊少年

11-13

$\alpha$ 表示 $K$ 个高斯混合子模型的选取概率，总概率为1； $\gamma$ 表示隐变量，即我们观测看到观测值 $x_i$ ，但是不知道是高斯混合模型中的哪个子模型产生产生的 $x_i$ ，因此定义了一个隐变量 $\gamma$ ， $\gamma$ 为0-1分布，每个观测数值只对应一个 $\gamma$ 为1，其他分模型对应的 $\gamma$ 均为0；因此建立了从输出到模型的联系。可以通过迭代的方法求解每个分模型的参数

👍 赞



Hi18

2020-07-11

求问答者，是不是所有连续分布都可以用GM模型去描述。或者这有一个问题，多个相关同分布高斯随机变量最大值的分布可以用GM去模拟吗？谢谢



👍 赞



jinzhao

2020-06-29

$\alpha$  是隐变量，也就是分类变量的数量比例，例如学生的身高，男生和女生的数量分别为50与50，那么，50%和50%两个比例。下面的 $r_{jk}$  是根据值所在位置算出来的单个点的概率。

👍 赞





戴文亮 (作者) 回复 jinzhao

2020-07-01

$\alpha_k$  是描述整体而言 数据来自第k个子分布的概率, 或者说这个分布的占比。 $r_{jk}$  是某一个数据来自第 k 个子分布的可能性。你看一下迭代公式里,  $\alpha_k$ 的更新方法, 应该就可以理解。

👍 1



陈三章 回复 jinzhao

07-21

在李航老师的书中,  $r_{jk}$ 定义为数据是否属于第k个子模型, 是0-1随机变量

👍 赞



王汪汪

2020-06-16

可以问一下 $\alpha$ 最初的值是如何得到的嘛? (因为看迭代的话, 感觉还是在已知 $\alpha$ 的情况下求下一个, 若是追溯到第一个 $\alpha_0$ , 感觉我也可以从零开始迭代, 这样就不make sense了。)感谢!

👍 赞



陈伟 回复 王汪汪

2020-06-30

$\alpha$ 可以设置 $1/K$ , 默认所有观测变量来自不同的子高斯分布的概率是一样。举个例子, 如果只有两个子高斯, 那么 $\alpha$ 为0.5, 0.5

👍 2



炖一年

2020-05-28

M-step: 计算新一轮迭代的模型参数里,  $\mu$ 和 $\sigma$ 计算公式的图片没有标注j从1开始

👍 赞



薛定谔

2020-04-03

不错, 不错。

👍 赞



KShao

2020-03-06

K 个高斯模型的和不是一个高斯模型吗?

👍 赞



Laughing

2019-08-04

相当清晰, 由衷感谢!

还是我哪里理解有错误，希望大家指点一下

👍 赞



郝翌骏 回复 阿呆i

2019-10-28

这个是 初始化每个component对应的高斯分布的参数时的技巧,先把 标准差初始化比较大. 随着E-M迭代,标准差会慢慢缩小.

👍 赞