高斯混合模型

Gaussian mixture model(GMM)
learning note For reading translation

我真的不懂忧郁

Delft University of Technology

高斯混合模型

Gaussian mixture model(GMM) learning note For reading translation

by

我真的不懂忧郁

Student Name	Student Number
First Surname	1234567

Instructor:	I. Surname
Teaching Assistant:	I. Surname
Project Duration:	Month, Year - Month, Year
Faculty:	Faculty of Aerospace Engineering, Delft

TU Delft

Preface

A preface...

我真的不懂忧郁
Delft, May 2024

Summary

A summary...

目录

Nomenclature

If a nomenclature is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.

Abbreviations

Abbreviation	Definition
ISA	International Standard Atmosphere
...	

Symbols

Symbol	Definition	Unit
V	Velocity	[m/s]
...		
ρ	Density	[kg/m^3]
...		

Chapter 1

高斯分布

目录大纲以维基百科为准

Chapter 2

Gaussian mixture model

高斯混合模型概述

高斯混合模型 *(Gaussian Mixture Model, GMM)* 是一种流行的概率模型，用于表示具有多个子高斯分布的混合分布。它常用于聚类、分类和密度估计。下面我们从混合模型的角度详细讨论高斯混合模型。

高斯混合模型是一种混合模型，其中每个组件都是一个高斯分布。混合模型的基本思想是将数据视为由若干个子分布生成，每个子分布代表一个不同的群体或类别。

一个 GMM 可以表示为

$$p(x) = \sum_{i=1}^{K} \pi_k \mathcal{N}(x_k | \mu_k, \Sigma_k) \tag{2.1}$$

其中 K 是高斯分布的数量，π_k 是第 k 个高斯分布的混合系数，表示该高斯分布的权重，满足

$$\sum_{i=1}^{K} \pi_k = 1 \tag{2.2}$$

$\mathcal{N}(x_k | \mu_k, \Sigma_k)$ 是第 k 个高斯分布，其均值为 μ_k，协方差矩阵为 Σ_k。

2.1. 从几何的角度

几何角度来看就是多个高斯分布加权平均叠加而成的。

2.2. 从混合模型的角度

由于是多个高斯分布的混叠，所以我们可以用两个随机变量 X, Z，X 称为 *observed variable*，Z 称为 *latent variable*，即对应哪个高斯分布的随机变量。例如一共有 $c_1, \cdots, c_n$ 个高斯分布。

Z	c_1	c_2	$\cdots$	c_k
x	p_1	p_2	$\cdots$	p_k

表 2.1: 概率分布

则观测随机变量总的概率分布为

$$
\begin{aligned}
P(X) &= \sum_{Z} P(X, Z) \\
&= \sum_{k=1}^{K} P(X, Z = C_k) \\
&= \sum_{k=1}^{K} P(Z = C_k) \cdot P(X|Z = C_k) \\
&= \sum_{k=1}^{K} p_k \cdot \mathcal{N}(X|\mu_k, \Sigma_k)
\end{aligned}
\tag{2.3}
$$

Chapter 3

高斯混合模型的参数估计:
极大似然估计

3.1. 极大似然估计 (maximum likelihood estimation)

给定概率分布 D，已知其概率密度函数（连续分布）或者概率质量函数（离散分布）为 f_D，以及一个分布参数 θ，我们可以从这个分布中抽一个具有 n 个值的采样 $X_1, X_2, \cdots, X_n$，利用 f_D 计算出其似然函数

$$L(\theta|x_1, \cdots, x_n) = f_\theta(x_1, \cdots, x_n) \tag{3.1}$$

若 D 是离散分布，f_θ 即是在参数为 θ 时观测到这一采样的概率；若其是连续分布，f_θ 则为 $X_1, X_2, \cdots, X_n$ 联合分布的概率密度函数在观测值处的取值。

从数学上来说，我们可以在 θ 的所有可能取值中寻找一个值使得似然函数取到最大值。这个使可能性最大的 $\widehat{\theta}$ 值即称为 θ 的最大似然估计。由定义，最大似然估计是样本的函数。

相对熵

最大似然估计可以从相对熵推导而来。相对熵衡量了使用一个给定分布 Q 来近似另一个分布 P 时的信息损失，对于离散随机变量

$$D_{KL}(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)} \tag{3.2}$$

其中 P 是真实分布，Q 是近似分布。在最大似然估计的情景下，假设分布拥有一系列参数 θ，我们希望通过样本得到参数的估计值 $\hat{\theta}$，我们可以利用相对熵来评估估计的好坏

$$D_{KL}(p_\theta(x)||p_{\hat{\theta}}(x)) = \sum_{x \in E} p_\theta(x) log \frac{p_\theta(x)}{p_{\hat{\theta}}(x)} \tag{3.3}$$

根据数学期望的定义，上式可以改写

$$D_{KL}(p_\theta(x)||p_{\hat{\theta}}(x)) = \mathbb{E}_\theta[log(\frac{p_\theta(x)}{p_{\hat{\theta}}(x)})] = \mathbb{E}_\theta[log\, p_\theta(x)] - \mathbb{E}_\theta[log\, p_{\hat{\theta}}(x)] \tag{3.4}$$

KL 值越大，参数估计越坏，因此，需要通过改变估计参数 $\hat{\theta}$ 的值来获得最小的值，所对应的参数极为最佳估计参数

$$\hat{\theta}_{best} = arg \min_{\hat{\theta}} D_{KL}(p_\theta(x)||p_{\hat{\theta}}(x)) \tag{3.5}$$

假设有 n 个样本，根据大数定律，

$$\mathbb{E}_\theta[log\ p_{\hat{\theta}(x)}] \rightsquigarrow \frac{1}{n}\sum_{i=1}^{n} log\ p_{\hat{\theta}}(x) \tag{3.6}$$

因此我们可以通过下式去估计

$$D_{KL}(p_\theta(x)||p_{\hat{\theta}}(x)) = \mathbb{E}_\theta[log\ p_{\hat{\theta}(x)}] - \frac{1}{n}\sum_{i=1}^{n} log\ p_{\hat{\theta}}(x_i) \tag{3.7}$$

对于一个已知的分布，其参数 θ 是确定的。因此，$\mathbb{E}_\theta[log\ p_{\hat{\theta}(x)}]$ 为常数。因此，我们可以通过最小化 KL 值获得最佳估计参数：

$$\hat{\theta} = arg \min_{\theta} \mathbb{E}_\theta[log\ p_{\hat{\theta}(x)}] - \frac{1}{n}\sum_{i=1}^{n} log\ p_{\hat{\theta}}(x_i) \tag{3.8}$$

只要求和项最大，那么整体就最小，这个优化问题等价于

$$\begin{aligned} & arg \max_{\theta} \frac{1}{n}\sum_{i=1}^{n} log\ p_{\hat{\theta}}(x_i) \\ & \Rightarrow arg \max_{\theta} log[\prod_{i=1}^{n} p_{\hat{\theta}}(x_i)] \\ & \Rightarrow arg \max_{\theta} \prod_{i=1}^{n} p_{\hat{\theta}}(x_i) \end{aligned} \tag{3.9}$$

因此，要得到最佳参数估计值，只需要最大化 $\prod_{i=1}^{n} p_{\hat{\theta}}(x_i)$，这就是最大似然函数。

3.2. 极大似然估计看高斯混合模型

极大似然估计来解单一高斯问题是存在解析解的，但是实际上极大似然估计来看高斯混合模型的解析解是求不出来的。这一节主要是为了看一下为什么。

设

$$X\ :\ Observed\ Data\ \rightarrow\ X = (x_1, x_2, \cdots, x_n) \tag{3.10}$$

$$(X, Z)\ :\ Complete\ Data \tag{3.11}$$

$$\theta\ :\ Parameter\ \rightarrow\ \theta = \{p_1, \cdots, p_k.\mu_1, \cdots, \mu_k, \sum_1, \cdots, \sum_k\} \tag{3.12}$$

极大似然估计去估计参数

$$\begin{aligned} \hat{\theta_{MLE}} &= arg \max_{\theta} log\ P(X) \\ &= arg \max_{\theta} log \prod_{i=1}^{N} P(x_i) \\ &= arg \max_{\theta} \sum_{i=1}^{N} log\ P(x_i) \end{aligned} \tag{3.13}$$

Chapter 4

高斯混合模型的参数估计: EM 求解

EM 其实是期望最大算法。

4.1. E-step

$$
\begin{aligned}
Q(\theta,\theta^{(t)}) &= \int_x log\ P(x,z|\theta)\cdot P(z|x,\theta^{(t)})dz \\
&= \sum_{i=1}^{N}\sum_{z_i} log\left[P_{z_1}\cdot\mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\right]\cdot\frac{p_{z_i}\cdot\mathcal{N}(x_i|\mu_{z_i}^{(t)},\Sigma_{z_i}^{(t)})}{\sum_{k=1}^{K}p_k^{(t)}\cdot\mathcal{N}(x_i|\mu_{z_i}^{(t)},\Sigma_{z_i}^{(t)})} \\
&= \sum_{i=1}^{N}\sum_{z_i} log\left[P_{z_1}\cdot\mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\right]\cdot P(z_i|x_i,\theta^{(t)}) \\
&= \sum_{i=1}^{N}\sum_{z_i} log\left[P_{z_1}\cdot\mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\right]\cdot P(z_i=c_k|x_i,\theta^{(t)}) \\
&= \sum_{i=1}^{N}\sum_{z_i} \left[log\ P_{z_1}+log\ \mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\right]\cdot P(z_i=c_k|x_i,\theta^{(t)})
\end{aligned} \tag{4.1}
$$

转化为最优化问题

$$\theta^{(t+1)} = arg\ \max_{\theta}\ Q(\theta,\theta^{(t)}) \tag{4.2}$$

求 $p_k^{(t+1)}$

$$
\begin{cases}
p_k^{(t+1)} = arg\ \max\limits_{p_k} \sum\limits_{k=1}^{K}\sum\limits_{i=1}^{N} log\ p_k\cdot P(z_i=c_k|x_i.\theta^{(t)}) \\
s.t.\quad \sum\limits_{k=1}^{K} p_k = 1
\end{cases} \tag{4.3}
$$

4.2. M-step

求 $p^{(t+1)} = (p_1^{(t+1)}, p_2^{(t+1)}, \cdots, p_k^{(t+1)})$，

$$\begin{cases} \max\limits_{p} \sum\limits_{k=1}^{K} \sum\limits_{i=1}^{N} log\, p_k \cdot P(z_i = c_k | x_i, \theta^{(t)}) \\ s.t. \quad \sum\limits_{k=1}^{K} p_k = 1 \end{cases} \tag{4.4}$$

由拉格朗日乘子法，结论是

$$p_k^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} P(z_i = c_k | x_i, \theta^{(t)}) \tag{4.5}$$

References

[1] I. Surname, I. Surname, and I. Surname. “The Title of the Article”. In: *The Title of the Journal* 1.2 (2000), pp. 123–456.

Chapter A

Source Code Example

Adding source code to your report/thesis is supported with the package listings. *An example can be found below. Files can be added using* \lstinputlisting[language=<language>]{<filename>}.

```python
"""
ISA Calculator: import the function, specify the height and it will return a
list in the following format: [Temperature,Density,Pressure,Speed of Sound].
Note that there is no check to see if the maximum altitude is reached.
"""

import math
g0 = 9.80665
R = 287.0
layer1 = [0, 288.15, 101325.0]
alt = [0,11000,20000,32000,47000,51000,71000,86000]
a = [-.0065,0,.0010,.0028,0,-.0028,-.0020]

def atmosphere(h):
    for i in range(0,len(alt)-1):
        if h >= alt[i]:
            layer0 = layer1[:]
            layer1[0] = min(h,alt[i+1])
            if a[i] != 0:
                layer1[1] = layer0[1] + a[i]*(layer1[0]-layer0[0])
                layer1[2] = layer0[2] * (layer1[1]/layer0[1])**(-g0/(a[i]*R))
            else:
                layer1[2] = layer0[2]*math.exp((-g0/(R*layer1[1]))*(layer1[0]-layer0[0]))
    return [layer1[1],layer1[2]/(R*layer1[1]),layer1[2],math.sqrt(1.4*R*layer1[1])]
```

Chapter B

Task Division Example

If a task division is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.

表 **B.1:** Distribution of the workload

	Task	Student Name(s)
	Summary	
Chapter 1	Introduction	
Chapter 2		
Chapter 3		
Chapter *		
Chapter *	Conclusion	
	Editors	
	CAD and Figures	
	Document Design and Layout	