Table Of Contents:

今日工作

1、阅读《计算机视觉 模型、学习和推理》第七章2-4节。

2、练习生成一维正态分布数据和混合高斯分布可视化。以及二维正态分布数据的生成和可视化。

闭合解

> 在解组件特性相关的方程式时，大多数的时候都要去解偏微分或积分式，才能求得其正确的解。依照求解方法的不同，可以分成以下两类：解析解和数值解。

> 解析解*(analytical solution)*就是一些严格的公式,给出任意的自变量就可以求出其因变量,也就是问题的解, 他人可以利用这些公式计算各自的问题. 所谓的解析解是一种包含分式、三角函数、指数、对数甚至无限级数等基本函数的解的形式。用来求得解析解的方法称为解析法*(analytic techniques*、*analytic methods)*，解析法即是常见的微积分技巧，例如分离变量法等。解析解为一封闭形式*(closed-form)*的函数，因此对任一独立变量，我们皆可将其带入解析函数求得正确的相依变量。因此，解析解也被称为闭合解*(closed-form solution)*。

> 数值解*(numerical solution)*是采用某种计算方法,如有限元的方法, 数值逼近,插值的方法, 得到的解.别人只能利用数值计算的结果, 而不能随意给出自变量并求出计算值. 当无法藉由微积分技巧求得解析解时，这时便只能利用数值分析的方式来求得其数值解了。数值方法变成了求解过程重要的媒介。在数值分析的过程中，首先会将原方程式加以简化，以利后来的数值分析。例如，会先将微分符号改为差分符号等。然后再用传统的代数方法将原方程式改写成另一方便求解的形式。这时的求解步骤就是将一独立变量带入，求得相依变量的近似解。因此利用此方法所求得的相依变量为一个个分离的数值*(discrete values)*，不似解析解为一连续的分布，而且因为经过上述简化的动作，所以可以想见正确性将不如解析法来的好。

> 数值解是在特定条件下通过近似计算得出来的一个数值，而解析解为该函数的解析式解析解就是给出解的具体函数形式，从解的表达式中就可以算出任何对应值；数值解就是用数值方法求出解，给出一系列对应的自变量和解。 *e.g. eq: x^2=5 solution: x=sqrt(5) -- analytical solution*（解析解） *x=2.236 -- numerical solution*（数值解）

http://blog.sina.com.cn/s/blog_65838c3c0101e7tg.html

复杂数据密度建模

隐变量

隐变量可以是离散或连续的。隐变量的个数与x 没有直接的关系。

在之前，数据x 得分布与参数θ 有关，引入隐变量后，这个分布首先是x 和隐变量h 的联合分布关于关于h 的边缘分布，即：

$$Pr(\mathbf{x}|\boldsymbol{\theta}) = \int Pr(\mathbf{x},\mathbf{h}|\boldsymbol{\theta})\mathrm{d}\mathbf{h}$$

那么，现在的参数选择方法是：

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\left[\sum_{i=1}^{I}\log\left[\int Pr(\mathbf{x}_i,\mathbf{h}_i|\boldsymbol{\theta})\mathrm{d}\mathbf{h}_i\right]\right]$$

这个方法称为**期望最大方法。**

期望最大化

这节介绍使用期望最大值方法拟合参数θ 的步骤。

首先，引入 了一个新的函数作为原函数的下界：

$$B[\{q_i(\boldsymbol{h}_i)\},\boldsymbol{\theta}] = \sum_{i=1}^{I}\int q_i(\boldsymbol{h}_i)\log\left[\frac{Pr(\boldsymbol{x}_i,\boldsymbol{h}_i|\boldsymbol{\theta})}{q_i(\boldsymbol{h}_i)}\right]d\boldsymbol{h}_i \leqslant \sum_{i=1}^{I}\log\left[\int Pr(\boldsymbol{x}_i,\boldsymbol{h}_i|\boldsymbol{\theta})d\boldsymbol{h}_i\right]$$

循环交替进行以下两步:

期望步或**E步**:

$$\hat{q}_i(\boldsymbol{h}_i) = Pr(\boldsymbol{h}_i|\boldsymbol{x}_i,\boldsymbol{\theta}^{[t]}) = \frac{Pr(\boldsymbol{x}_i|\boldsymbol{h}_i,\boldsymbol{\theta}^{[t]})Pr(\boldsymbol{h}_i|\boldsymbol{\theta}^{[t]})}{Pr(\boldsymbol{x}_i)}$$

最大化步或**M步**:

$$\hat{\boldsymbol{\theta}}^{[t+1]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\left[\sum_{i=1}^{I}\int \hat{q}_i(\boldsymbol{h}_i)\log[Pr(\boldsymbol{x}_i,\boldsymbol{h}_i|\boldsymbol{\theta})]d\boldsymbol{h}_i\right]$$

这一节并没有证明这样做可以增大下界，只是假设这个命题成立。

混合高斯模型

$$Pr(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{k=1}^{K}\lambda_k \text{Norm}_{\boldsymbol{x}}[\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k]$$

在这个模型中，*h* 作为索引决定哪个分布对应观察数据点*x* 。

混合高斯边缘化

预先定义:

那么:

基于期望最大化的混合模型拟合

E 步:

第*i* 个数据对应第*k* 个正态分布的概率*Pr(h_i=k|x_i, θ_t)* ，记为*r_ik* 。

M 步:

参数更新规则:

k 指第*k* 个分布的参数。

r_ik 越大说明数据*x_i* 属于第*k* 个分布的概率越大，因此对这个分布修正时的影响越高。

拟合时需要考虑的问题:

1、参数*θ* 的初始值。期望值最大算法不能找到全局最优解，通过设置不同的初始值，拟合到不同的局部最优解，最终选择最好的解。

2、必须指定混合分量的个数。

编程练习

生成正态分布的数据*x* 。

使用函数*normal_distribution()* 得到*y* 。

```
def normal_distribution(x, mean, sigma):
    return np.exp(-1 * ((x - mean) ** 2) / (2 * (sigma ** 2))) / (math.sqrt(2 * np.pi) * sigma)
```

生成二维正态分布数据: