深度之眼
deepshare.net

法律声明

本课件包括演示文稿、示例、代码、题库、视频和声音等内容，深度之眼和讲师拥有完全知识产权；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或者机构不得盗版、复制、仿造其中的创意和内容，我们保留一切通过法律手段追究违反者的权利。

课程详情请咨询

- 微信公众号：深度之眼
- 客服微信号：deepshare0920

公众号

微 信

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

deepshare.net
深度之眼

西瓜书公式推导

导师：Sm1les（Datawhale南瓜书项目负责人）

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

deepshare.net
深度之眼

高斯混合模型公式推导

Derivation of Gaussian Mixture Model

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

本节大纲

Outline

先修内容：西瓜书9.1、9.2、9.3、9.4

1.定义

2.参数估计

- EM算法
- 极大似然估计

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

定义：

$$P(\boldsymbol{x})=\sum_{i=1}^{k}\alpha_i\cdot\phi\left(\boldsymbol{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right)$$

该模型共由k个混合成分组成，每个混个成分对应一个高斯分布，其中，$\boldsymbol{x}\in\mathbb{R}^n$ 为α_i混合系数，且 $\alpha_i\geq 0,\sum_{i=1}^{k}\alpha_i=1$ ，$\phi\left(\boldsymbol{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right)$ 为多元高斯分布（当$\boldsymbol{x}$为标量时相应地替换为一元高斯分布）的概率密度函数：

$$\phi\left(\boldsymbol{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right)=\frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right)$$

其生成数据的方式为：首先，依概率 α_i 选择第i个高斯混合成分，接着依据该混合成分的概率分布 $\phi\left(\boldsymbol{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right)$ 生成样本。

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

EM算法：

已知数据集 $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m\}$ 中的样本中的样本均由某个高斯混合模型生成，而每个样本 $\boldsymbol{x}_j$ 是由哪个高斯混合成分生成的是未知的，属于一个隐变量，我们令其为 z_j

$z_j\in\{1,2,...,k\}$ 表示生成样本 $\boldsymbol{x}_j$ 的高斯混合成分，结合高斯混合模型生成数据的方式易知 z_j 的分布律为 $P(z_j=i)=\alpha_i$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

E步： 确定Q函数

$$
\begin{aligned}
Q(\theta|\theta^{(i)}) &= \sum_{Z} P(Z|X,\theta^{(i)}) \ln P(X,Z|\theta) \\
&= \sum_{z_1,z_2,\ldots,z_m} \left\{ \prod_{j=1}^{m} P(z_j|\boldsymbol{x}_j,\theta^{(i)}) \ln \left[\prod_{j=1}^{m} P(\boldsymbol{x}_j,z_j|\theta) \right] \right\} \\
&= \sum_{j=1}^{m} \left[\sum_{z_j} P(z_j|\boldsymbol{x}_j,\theta^{(i)}) \ln P(\boldsymbol{x}_j,z_j|\theta) \right] \\
&= \sum_{j=1}^{m} \left[\sum_{i=1}^{k} P(z_j=i|\boldsymbol{x}_j,\theta^{(i)}) \ln P(\boldsymbol{x}_j,z_j=i|\theta) \right]
\end{aligned}
$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

对于 $P(z_j = i|\boldsymbol{x}_j, \theta^{(i)})$ ：

由贝叶斯定理可知

$$P(z_j = i|\boldsymbol{x}_j) = \frac{P(z_j = i) \cdot P(\boldsymbol{x}_j|z_j = i)}{P(\boldsymbol{x}_j)}$$

$$= \frac{\alpha_i \cdot \phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}{\sum_{l=1}^{k} \alpha_l \cdot \phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l\right)}$$

那么

$$P(z_j = i|\boldsymbol{x}_j, \theta^{(i)}) = \frac{\alpha_i^{(i)} \cdot \phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_i^{(i)}, \boldsymbol{\Sigma}_i^{(i)}\right)}{\sum_{l=1}^{k} \alpha_l^{(i)} \cdot \phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_l^{(i)}, \boldsymbol{\Sigma}_l^{(i)}\right)}$$

将其简记为γ_{ji}

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

对于 $P(\boldsymbol{x}_j, z_j = i|\theta)$ ：

$$P(\boldsymbol{x}_j, z_j = i|\theta) = P(\boldsymbol{x}_j|z_j = i, \theta) \cdot P(z_j = i|\theta)$$

$$= \phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right) \cdot \alpha_i$$

将上式代入Q函数可得

$$Q(\theta|\theta^{(i)}) = \sum_{j=1}^{m}\left[\sum_{i=1}^{k} P(z_j = i|\boldsymbol{x}_j, \theta^{(i)}) \ln P(\boldsymbol{x}_j, z_j = i|\theta)\right]$$

$$= \sum_{j=1}^{m}\sum_{i=1}^{k} \gamma_{ji} \ln\left[\alpha_i \cdot \phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)\right]$$

deepshare.net
深度之眼

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

$$
\begin{aligned}
Q(\theta|\theta^{(i)}) &= \sum_{j=1}^{m}\sum_{i=1}^{k}\gamma_{ji}\left[\ln\alpha_i+\ln\phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right)\right] \\
&= \sum_{j=1}^{m}\sum_{i=1}^{k}\left[\gamma_{ji}\ln\alpha_i+\gamma_{ji}\ln\phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right)\right] \\
&= \sum_{j=1}^{m}\sum_{i=1}^{k}\left\{\gamma_{ji}\ln\alpha_i+\gamma_{ji}\ln\left[\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right)\right]\right\} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{k}\left\{\gamma_{ji}\ln\alpha_i+\gamma_{ji}\left[\ln\frac{1}{(2\pi)^{\frac{n}{2}}}-\frac{1}{2}\ln|\boldsymbol{\Sigma}_i|-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^{T}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right]\right\} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{k}\left\{\gamma_{ji}\ln\alpha_i+\gamma_{ji}\ln\frac{1}{(2\pi)^{\frac{n}{2}}}-\frac{1}{2}\gamma_{ji}\ln|\boldsymbol{\Sigma}_i|-\frac{1}{2}\gamma_{ji}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^{T}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right\}
\end{aligned}
$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

M步：求使得Q函数达到极大的$\theta^{(i+1)}$

求 $\boldsymbol{\mu}_i^{(i+1)}$：

对于Q函数关于$\boldsymbol{\mu}_i$求偏导

$$\frac{\partial Q(\theta,\theta^{(i)})}{\partial \boldsymbol{\mu}_i}=\sum_{j=1}^{m}\left\{0+0-0-\frac{1}{2}\gamma_{ji}\frac{\partial\left((\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right)}{\partial \boldsymbol{\mu}_i}\right\}$$

$$=-\sum_{j=1}^{m}\frac{1}{2}\gamma_{ji}\frac{\partial\left(\boldsymbol{x}_j^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x}_j-\boldsymbol{x}_j^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i-\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x}_j+\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i\right)}{\partial \boldsymbol{\mu}_i}$$

$$=-\sum_{j=1}^{m}\frac{1}{2}\gamma_{ji}\frac{\partial\left(-\boldsymbol{x}_j^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i-\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x}_j+\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i\right)}{\partial \boldsymbol{\mu}_i}$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

由于 $\boldsymbol{x}_j^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i$ 和 $\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j$ 均为标量且 $\boldsymbol{\Sigma}_i$ 为对称矩阵，所以

$$\left(\boldsymbol{x}_j^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i\right)^T = \boldsymbol{\mu}_i^T \left(\boldsymbol{\Sigma}_i^{-1}\right)^T \boldsymbol{x}_j = \boldsymbol{\mu}_i^T \left(\boldsymbol{\Sigma}_i^T\right)^{-1} \boldsymbol{x}_j = \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j = \boldsymbol{x}_j^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i$$

代入 $\dfrac{\partial Q(\theta, \theta^{(i)})}{\partial \boldsymbol{\mu}_i} = -\sum_{j=1}^{m} \dfrac{1}{2}\gamma_{ji} \dfrac{\partial\left(-\boldsymbol{x}_j^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i - \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j + \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i\right)}{\partial \boldsymbol{\mu}_i}$ 可得

$$\frac{\partial Q(\theta, \theta^{(i)})}{\partial \boldsymbol{\mu}_i} = -\sum_{j=1}^{m} \frac{1}{2}\gamma_{ji} \frac{\partial\left(-2\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j + \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i\right)}{\partial \boldsymbol{\mu}_i}$$

由矩阵微分公式 $\dfrac{\partial \boldsymbol{x}^T \boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a}, \dfrac{\partial \boldsymbol{x}^T \mathbf{B} \boldsymbol{x}}{\partial \boldsymbol{x}} = \left(\mathbf{B} + \mathbf{B}^T\right)\boldsymbol{x}$ 可得

$$= \sum_{j=1}^{m} \frac{1}{2}\gamma_{ji}\left(2\boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j - 2\boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i\right)$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

$$\frac{\partial Q(\theta, \theta^{(i)})}{\partial \boldsymbol{\mu}_i} = \sum_{j=1}^{m} \gamma_{ji} \left(\boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j - \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i \right)$$

令上式等于0可得

$$\sum_{j=1}^{m} \gamma_{ji} \left(\boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j - \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i \right) = 0$$

$$\boldsymbol{\Sigma}_i^{-1} \cdot \sum_{j=1}^{m} \gamma_{ji} \left(\boldsymbol{x}_j - \boldsymbol{\mu}_i \right) = 0$$

$$\sum_{j=1}^{m} \gamma_{ji} \left(\boldsymbol{x}_j - \boldsymbol{\mu}_i \right) = 0$$

$$\boldsymbol{\mu}_i = \frac{\sum_{j=1}^{m} \gamma_{ji} \boldsymbol{x}_j}{\sum_{j=1}^{m} \gamma_{ji}} \Rightarrow \boldsymbol{\mu}_i^{(i+1)} = \frac{\sum_{j=1}^{m} \gamma_{ji} \boldsymbol{x}_j}{\sum_{j=1}^{m} \gamma_{ji}}$$

此即为式9.34

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

求 $\boldsymbol{\Sigma}_i^{(i+1)}$ ：

对于Q函数关于 $\boldsymbol{\Sigma}_i$ 求偏导

$$\frac{\partial Q(\theta,\theta^{(i)})}{\partial \boldsymbol{\Sigma}_i}=\sum_{j=1}^{m}\left\{0+0-\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left(\frac{1}{2}\gamma_{ji}\ln|\boldsymbol{\Sigma}_i|\right)-\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left[\frac{1}{2}\gamma_{ji}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right]\right\}$$

$$=\sum_{j=1}^{m}\left\{-\frac{1}{2}\gamma_{ji}\frac{\partial\left(\ln|\boldsymbol{\Sigma}_i|\right)}{\partial \boldsymbol{\Sigma}_i}-\frac{1}{2}\gamma_{ji}\frac{\partial\left[(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right]}{\partial \boldsymbol{\Sigma}_i}\right\}$$

由矩阵微分公式 $\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}}=|\mathbf{X}|\cdot(\mathbf{X}^{-1})^T,\frac{\partial \boldsymbol{a}^T\mathbf{X}^{-1}\boldsymbol{b}}{\partial \mathbf{X}}=-\mathbf{X}^{-T}\boldsymbol{a}\boldsymbol{b}^T\mathbf{X}^{-T}$ 可得

$$=\sum_{j=1}^{m}\left\{-\frac{1}{2}\gamma_{ji}\cdot\frac{1}{|\boldsymbol{\Sigma}_i|}\cdot|\boldsymbol{\Sigma}_i|\cdot(\boldsymbol{\Sigma}_i^{-1})^T-\frac{1}{2}\gamma_{ji}\cdot(-\boldsymbol{\Sigma}_i)^{-T}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-T}\right\}$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

$$\frac{\partial Q(\theta,\theta^{(i)})}{\partial \boldsymbol{\Sigma}_i}=\sum_{j=1}^{m}\left\{-\frac{1}{2}\gamma_{ji}\boldsymbol{\Sigma}_i^{-1}+\frac{1}{2}\gamma_{ji}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}\right\}$$

令上式等于0可得

$$\sum_{j=1}^{m}\left\{-\frac{1}{2}\gamma_{ji}\boldsymbol{\Sigma}_i^{-1}+\frac{1}{2}\gamma_{ji}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}\right\}=0$$

$$\sum_{j=1}^{m}\left\{-\frac{1}{2}\gamma_{ji}+\frac{1}{2}\gamma_{ji}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\right\}=0$$

$$\frac{1}{2}\sum_{j=1}^{m}\gamma_{ji}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T=\frac{1}{2}\sum_{j=1}^{m}\gamma_{ji}$$

$$\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{m}\gamma_{ji}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T=\sum_{j=1}^{m}\gamma_{ji}$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

$$\boldsymbol{\Sigma}_i^{-1} = \frac{\sum_{j=1}^{m} \gamma_{ji}}{\sum_{j=1}^{m} \gamma_{ji}(\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T}$$

$$\boldsymbol{\Sigma}_i = \frac{\sum_{j=1}^{m} \gamma_{ji}(\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T}{\sum_{j=1}^{m} \gamma_{ji}} \Rightarrow \boldsymbol{\Sigma}_i^{(i+1)} = \frac{\sum_{j=1}^{m} \gamma_{ji}(\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T}{\sum_{j=1}^{m} \gamma_{ji}}$$

此即为式9.35

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

求 $\alpha_i^{(i+1)}$ ：

由于 α_i 存在约束 $\sum_{i=1}^{k} \alpha_i = 1$ ，所以考虑使用拉格朗日乘子法，由拉格朗日乘子法可得拉格朗日函数为

$$
\begin{aligned}
L(\boldsymbol{\alpha}, \lambda) &= Q(\theta, \theta^{(i)}) + \lambda \left(\sum_{i=1}^{k} \alpha_i - 1 \right) \\
&= \sum_{j=1}^{m} \sum_{i=1}^{k} \left\{ \gamma_{ji} \ln \alpha_i + \gamma_{ji} \ln \frac{1}{(2\pi)^{\frac{n}{2}}} - \frac{1}{2} \gamma_{ji} \ln |\boldsymbol{\Sigma}_i| - \frac{1}{2} \gamma_{ji} (\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu}_i) \right\} + \lambda \left(\sum_{i=1}^{k} \alpha_i - 1 \right)
\end{aligned}
$$

对拉格朗日函数关于 α_i 求偏导

$$
\frac{\partial L(\boldsymbol{\alpha}, \lambda)}{\partial \alpha_i} = \sum_{j=1}^{m} \left\{ \frac{\partial \left(\gamma_{ji} \ln \alpha_i \right)}{\partial \alpha_i} + 0 - 0 - 0 \right\} + \lambda \frac{\partial \left(\sum_{i=1}^{k} \alpha_i - 1 \right)}{\partial \alpha_i}
$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

$$\frac{\partial L(\boldsymbol{\alpha}, \lambda)}{\partial \alpha_i} = \sum_{j=1}^{m} \frac{\gamma_{ji}}{\alpha_i} + \lambda$$

令上式等于0可得

$$\sum_{j=1}^{m} \frac{\gamma_{ji}}{\alpha_i} + \lambda = 0$$

$$\frac{1}{\alpha_i} \sum_{j=1}^{m} \gamma_{ji} = -\lambda$$

$$\alpha_i = -\frac{1}{\lambda} \sum_{j=1}^{m} \gamma_{ji}$$

由于 $\sum_{i=1}^{k} \alpha_i = 1$，则对上式两边关于i求和可得

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

$$\sum_{i=1}^{k}\alpha_i = -\frac{1}{\lambda}\sum_{i=1}^{k}\sum_{j=1}^{m}\gamma_{ji}$$

$$1 = -\frac{1}{\lambda}\sum_{i=1}^{k}\sum_{j=1}^{m}\gamma_{ji}$$

$$\lambda = -\sum_{i=1}^{k}\sum_{j=1}^{m}\gamma_{ji}$$

又因为

$$\sum_{i=1}^{k}\sum_{j=1}^{m}\gamma_{ji} = \sum_{j=1}^{m}\sum_{i=1}^{k}\gamma_{ji} = \sum_{j=1}^{m}\sum_{i=1}^{k}P(z_j = i|\boldsymbol{x}_j, \theta^{(i)}) = \sum_{j=1}^{m}1 = m$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

所以

$$\lambda = -\sum_{i=1}^{k}\sum_{j=1}^{m}\gamma_{ji} = -m$$

$$\alpha_i = -\frac{1}{\lambda}\sum_{j=1}^{m}\gamma_{ji} = \frac{1}{m}\sum_{j=1}^{m}\gamma_{ji}$$

由于$0 \leq \gamma_{ji} = P(z_j = i|\boldsymbol{x}_j, \theta^{(i)}) \leq 1$，所以

$$0 \leq \sum_{j=1}^{m}\gamma_{ji} \leq m \Rightarrow 0 \leq \frac{1}{m}\sum_{j=1}^{m}\gamma_{ji} \leq 1$$

那么此时解得的 α_i 是有效解，可以作为下一次迭代的初始参数，也即

$$\alpha_i = \frac{1}{m}\sum_{j=1}^{m}\gamma_{ji} \Rightarrow \alpha_i^{(i+1)} = \frac{1}{m}\sum_{j=1}^{m}\gamma_{ji}$$

此即为式9.38

高斯混合模型

Gaussian Mixture Model

极大似然估计法:

根据极大似然估计法可得对数似然函数为

$$
\begin{aligned}
LL(D) &= \ln\left(\prod_{j=1}^{m} P\left(\boldsymbol{x}_j\right)\right) \\
&= \sum_{j=1}^{m} \ln P\left(\boldsymbol{x}_j\right) \\
&= \sum_{j=1}^{m} \ln\left(\sum_{i=1}^{k} \alpha_i \cdot \phi\left(\boldsymbol{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)\right)
\end{aligned}
$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

对对数似然函数关于$\boldsymbol{\mu}_i$求偏导

$$
\begin{aligned}
\frac{\partial LL(D)}{\partial \boldsymbol{\mu}_i} &= \frac{\partial}{\partial \boldsymbol{\mu}_i}\left[\sum_{j=1}^{m} \ln\left(\sum_{i=1}^{k} \alpha_i \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right)\right] \\
&= \sum_{j=1}^{m} \frac{\partial}{\partial \boldsymbol{\mu}_i}\left[\ln\left(\sum_{i=1}^{k} \alpha_i \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right)\right] \\
&= \sum_{j=1}^{m} \frac{\alpha_i \cdot \frac{\partial}{\partial \boldsymbol{\mu}_i}(\phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i))}{\sum_{l=1}^{k} \alpha_l \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} \\
&= \sum_{j=1}^{m} \frac{\alpha_i \cdot \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right)}{\sum_{l=1}^{k} \alpha_l \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} \frac{\partial}{\partial \boldsymbol{\mu}_i}\left(-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right)
\end{aligned}
$$

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼
关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

$$\frac{\partial LL(D)}{\partial \boldsymbol{\mu}_i} = \sum_{j=1}^{m} \frac{\alpha_i \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k} \alpha_l \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} \cdot (-\frac{1}{2}) \cdot \frac{\partial}{\partial \boldsymbol{\mu}_i} (\boldsymbol{x}_j^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j - \boldsymbol{x}_j^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i - \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j + \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i)$$

令 $\frac{\alpha_i \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k} \alpha_l \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} = \gamma_{ji}$ 则

$$\begin{aligned}\frac{\partial LL(D)}{\partial \boldsymbol{\mu}_i} &= \sum_{j=1}^{m} \gamma_{ji} \cdot (-\frac{1}{2}) \cdot \frac{\partial}{\partial \boldsymbol{\mu}_i} (\boldsymbol{x}_j^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j - \boldsymbol{x}_j^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i - \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j + \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i) \\ &= \sum_{j=1}^{m} \gamma_{ji} \left(\boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_j - \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i\right)\end{aligned}$$

令上式等于0可得 $\boldsymbol{\mu}_i = \frac{\sum_{j=1}^{m} \gamma_{ji} \boldsymbol{x}_j}{\sum_{j=1}^{m} \gamma_{ji}}$ 此即为式9.34

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

对对数似然函数关于 $\boldsymbol{\Sigma}_i$ 求偏导

$$\frac{\partial LL(D)}{\partial \boldsymbol{\Sigma}_i} = \frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left[\sum_{j=1}^{m} \ln\left(\sum_{i=1}^{k} \alpha_i \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right)\right]$$

$$= \sum_{j=1}^{m} \frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left[\ln\left(\sum_{i=1}^{k} \alpha_i \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right)\right]$$

$$= \sum_{j=1}^{m} \frac{\alpha_i \cdot \frac{\partial}{\partial \boldsymbol{\Sigma}_i}(\phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i))}{\sum_{l=1}^{k} \alpha_l \cdot \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼
关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

其中

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left(\phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)\right) &= \frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left[\frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right)\right] \\
&= \frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left\{\exp\left[\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right)\right)\right]\right\} \\
&= \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)\cdot\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left[\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right)\right)\right] \\
&= \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)\cdot\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left[\ln\frac{1}{(2\pi)^{\frac{n}{2}}}-\frac{1}{2}\ln\left|\boldsymbol{\Sigma}_i\right|-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right] \\
&= \phi(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)\cdot\left[-\frac{1}{2}\frac{\partial\left(\ln\left|\boldsymbol{\Sigma}_i\right|\right)}{\partial \boldsymbol{\Sigma}_i}-\frac{1}{2}\frac{\partial\left[(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\right]}{\partial \boldsymbol{\Sigma}_i}\right]
\end{aligned}
$$

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

所以

$$
\begin{aligned}
\frac{\partial LL(D)}{\partial \boldsymbol{\Sigma}_i} &= \sum_{j=1}^{m} \frac{\alpha_i \cdot \phi(\boldsymbol{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k} \alpha_l \cdot \phi(\boldsymbol{x}_j | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} \left[-\frac{1}{2} \frac{\partial \left(\ln |\boldsymbol{\Sigma}_i| \right)}{\partial \boldsymbol{\Sigma}_i} - \frac{1}{2} \frac{\partial \left[(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu}_i) \right]}{\partial \boldsymbol{\Sigma}_i} \right] \\
&= \sum_{j=1}^{m} \gamma_{ji} \left[-\frac{1}{2} \frac{\partial \left(\ln |\boldsymbol{\Sigma}_i| \right)}{\partial \boldsymbol{\Sigma}_i} - \frac{1}{2} \frac{\partial \left[(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu}_i) \right]}{\partial \boldsymbol{\Sigma}_i} \right] \\
&= \sum_{j=1}^{m} \left\{ -\frac{1}{2} \gamma_{ji} \boldsymbol{\Sigma}_i^{-1} + \frac{1}{2} \gamma_{ji} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} \right\}
\end{aligned}
$$

令上式等于0可得 $\boldsymbol{\Sigma}_i = \dfrac{\sum_{j=1}^{m} \gamma_{ji} (\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T}{\sum_{j=1}^{m} \gamma_{ji}}$ 此即为式9.35

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

由于 α_i 存在约束 $\sum_{i=1}^{k}\alpha_i=1$，所以考虑使用拉格朗日乘子法，由拉格朗日乘子法可得拉格朗日函数为

$$\begin{aligned}L(\boldsymbol{\alpha},\lambda)&=LL(D)+\lambda\left(\sum_{i=1}^{k}\alpha_i-1\right)\\&=\sum_{j=1}^{m}\ln\left(\sum_{i=1}^{k}\alpha_i\cdot\phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right)\right)+\lambda\left(\sum_{i=1}^{k}\alpha_i-1\right)\end{aligned}$$

对拉格朗日函数关于 α_i 求偏导

$$\frac{\partial L(\boldsymbol{\alpha},\lambda)}{\partial\alpha_i}=\frac{\partial}{\partial\alpha_i}\left[\sum_{j=1}^{m}\ln\left(\sum_{i=1}^{k}\alpha_i\cdot\phi\left(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right)\right)\right]+\frac{\partial}{\partial\alpha_i}\left[\lambda\left(\sum_{i=1}^{k}\alpha_i-1\right)\right]$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

deepshare.net
深度之眼

$$\frac{\partial L(\boldsymbol{\alpha}, \lambda)}{\partial \alpha_i} = \sum_{j=1}^{m} \frac{\partial}{\partial \alpha_i} \left[\ln \left(\sum_{i=1}^{k} \alpha_i \cdot \phi(\boldsymbol{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right) \right] + \lambda$$

$$= \sum_{j=1}^{m} \frac{\phi(\boldsymbol{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k} \alpha_l \cdot \phi(\boldsymbol{x}_j | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} + \lambda$$

令上式等于0可得

$$\sum_{j=1}^{m} \frac{\phi(\boldsymbol{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k} \alpha_l \cdot \phi(\boldsymbol{x}_j | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} + \lambda = 0$$

两边同时乘以 α_i 可得

$$\sum_{j=1}^{m} \frac{\alpha_i \cdot p(\boldsymbol{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k} \alpha_l \cdot p(\boldsymbol{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)} + \lambda \alpha_i = 0$$

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

高斯混合模型

Gaussian Mixture Model

深度之眼 deepshare.net

关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

$$\sum_{j=1}^{m} \gamma_{ji} = -\lambda \alpha_i$$

两边关于i求和可得

$$\sum_{i=1}^{k} \sum_{j=1}^{m} \gamma_{ji} = -\lambda \sum_{i=1}^{k} \alpha_i$$

$$m = -\lambda$$

所以

$$\alpha_i = \frac{1}{m} \sum_{j=1}^{m} \gamma_{ji}$$

此即为式9.35

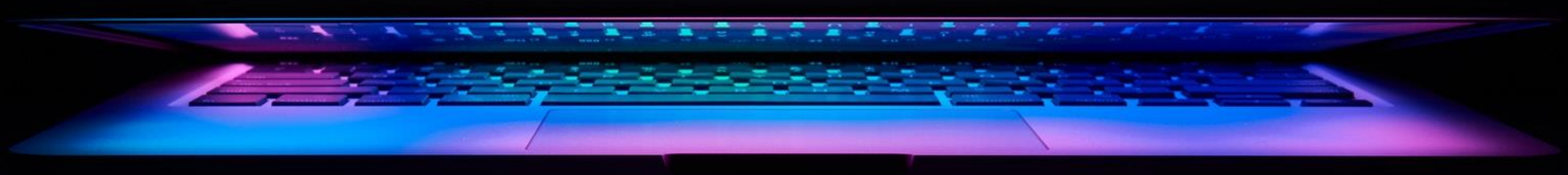
结 语

在这次课程中，我们学习了西瓜书

高斯混合模型的公式推导

那么在下次课程中，我们将会学习西瓜书

隐马尔可夫模型的公式推导



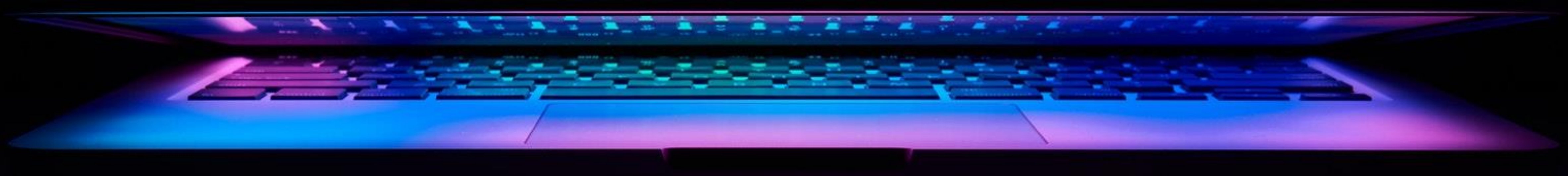
关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料

deepshare.net
深度之眼

联系我们:
电话：18001992849
邮箱：service@deepshare.net
Q Q： 2677693114

公众号

客服微信



关注公众号深度之眼，后台回复资料，获取AI必学书籍及完整实战学习资料