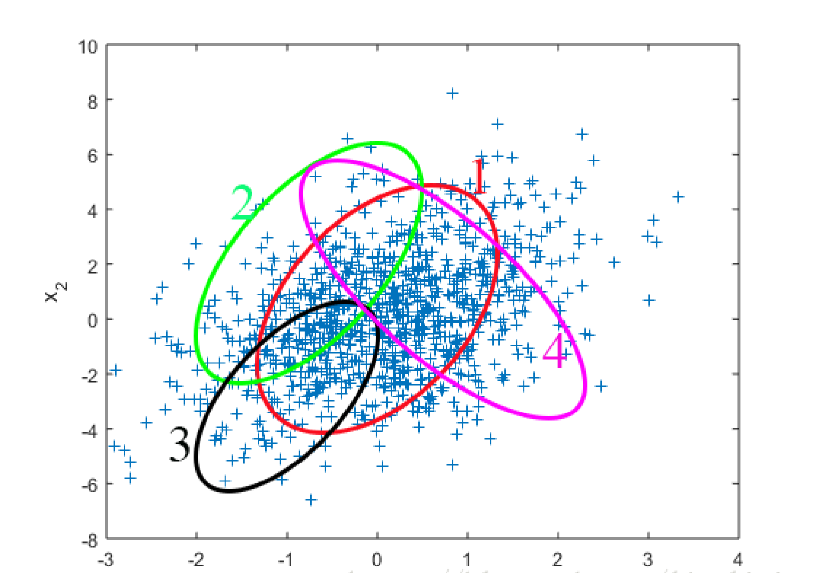
聚类：k-means、GMM

- 极大似然估计 - Maximum Likelihood Estimate
  - 举例：已知数据服从高斯分布，且我们观测到以下数据



根据左侧给出的观测数据，你觉得哪个高斯分布函数比较合适呢？

A. 分布1
B. 分布2
C. 分布3
D. 分布4

聚类：k-means、GMM

- 极大似然估计 - Maximum Likelihood Estimate
- 基本思想
  - 找到一组参数，使得已观测到样本集的联合概率最大化
- 似然函数
  - 假设有样本集$\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_N\}$服从高斯分布$N(\mathbf{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma})$
  - 假设样本抽样是独立的，那么我们同时抽到这N个样本的概率是抽到每个样本概率的乘积，也就是样本集Y的联合概率
  - 此联合概率即为似然函数

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma})=\prod_{i=1}^{N}p(\mathbf{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma}) \xrightarrow[\text{对数形式}]{} \ln L(\boldsymbol{\mu},\boldsymbol{\Sigma})=\sum_{i=1}^{N}\ln p(\mathbf{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma})$$

聚类：k-means、GMM

- 极大似然估计 - Maximum Likelihood Estimate
- 基本思想
  - 找到一组参数，使得已观测到样本集的联合概率最大化
- 最大化似然函数
  - 假设有样本集$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$服从高斯分布$N(\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\max \ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \max \sum_{i=1}^{N} \ln p(\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

  - 请写出1-D高斯分布的最大似然函数

$$\mathrm{f(x)} = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mathrm{l(x)} = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i}(x^i - \mu)^2$$

聚类：k-means、GMM

- 极大似然估计 - Maximum Likelihood Estimate

$$l(x) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_i \left(x^i - \mu\right)^2$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

对目标函数l(x)的参数μ和σ^2求偏导数，令偏导数为0，得到：

l(x)对μ求偏导数$-\frac{1}{2\sigma^2}\sum_i 2\left(x^i-\mu\right)(-1)$

$=\frac{1}{\sigma^2}\sum_i\left(x^i-\mu\right)$，令其等于0，$\boxed{\text{得到}\mu = \frac{1}{n}\sum_i x^i}$

l(x)对σ^2求偏导$-\frac{n}{2}\frac{2\pi}{2\pi\sigma^2} - \frac{1}{2}\sum_i\left(x^i-\mu\right)^2(-1){\sigma^2}^{-2}$

$= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\sum_i\left(x^i-\mu\right)^2{\sigma^2}^{-2}$，*令其等于0，两边乘以*${\sigma^2}^2$

$0 = -n\sigma^2 + \sum_i\left(x^i-\mu\right)^2$

$\boxed{\text{得到}\sigma^2 = \frac{1}{n}\sum_i\left(x^i-\mu\right)^2}$

请推导极大似然估计的参数μ_{ML}和σ^2_{ML}

聚类：k-means、GMM

聚类：k-means、GMM

- 混合高斯分布的极大似然估计
- 混合模型
  - $\sum_k \pi_k N(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
- 似然函数
  - $p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})=\prod_{i=1}^{N}[\sum_k \pi_k N(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)]$
- 对数似然函数
  - $L = \ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{N} \ln[\sum_k \pi_k N(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)]$

聚类：k-means、GMM

- 极大似然估计 - Maximum Likelihood Estimate
- 混合高斯分布的极大似然估计
- 对数似然函数
  - $L = \ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{N} \ln \sum_k \pi_k N(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
- 针对$\boldsymbol{\mu}_k$求导
  - 先处理第i项

第i项

$$\ln\left(\pi_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{\left(x^i-\mu_1\right)^2}{2\sigma_1^2}} + \pi_2 \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{\left(x^i-\mu_2\right)^2}{2\sigma_2^2}} + \ldots\right)$$

对$\boldsymbol{\mu}_k$求导

Log(x)导数为x分之1

$$\frac{1}{\sum_k \pi_k N(\mathbf{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)} \pi_k \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{\left(x^i-\mu_k\right)^2}{2\sigma_k^2}} \left(\frac{-2\left(x^i-\mu_k\right)(-1)}{2\sigma_k^2}\right) = \frac{\pi_k N(\mathbf{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_k \pi_k N(\mathbf{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)} \frac{\left(x^i-\mu_k\right)}{\sigma_k^2}$$

所有项

$$\frac{\pi_k N(\mathbf{x}_1|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_k \pi_k N(\mathbf{x}_1|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)} \frac{\left(x^1-\mu_k\right)}{\sigma_k^2} + \frac{\pi_k N(\mathbf{x}_2|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_k \pi_k N(\mathbf{x}_2|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)} \frac{\left(x^2-\mu_k\right)}{\sigma_k^2} + \frac{\pi_k N(\mathbf{x}_3|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_k \pi_k N(\mathbf{x}_3|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)} \frac{\left(x^3-\mu_k\right)}{\sigma_k^2} + \ldots = 0$$

聚类：k-means、GMM

- 极大似然估计 - Maximum Likelihood Estimate

$$\underbrace{\boxed{\frac{\pi_k N(\mathbf{x_1}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_k \pi_k N(\mathbf{x_1}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}}}_{\gamma_{1k}}\frac{(x^1-\boldsymbol{\mu}_k)}{\sigma_k^2}+\underbrace{\boxed{\frac{\pi_k N(\mathbf{x_2}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_k \pi_k N(\mathbf{x_2}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}}}_{\gamma_{2k}}\frac{(x^2-\boldsymbol{\mu}_k)}{\sigma_k^2}+\underbrace{\boxed{\frac{\pi_k N(\mathbf{x_3}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_k \pi_k N(\mathbf{x_3}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}}}_{\gamma_{3k}}\frac{(x^3-\boldsymbol{\mu}_k)}{\sigma_k^2}+\ldots=0$$

$$\boldsymbol{\mu}_k=\frac{1}{N_k}\sum_{i=1}^{N}\gamma_{ik}\,\mathbf{x}_i$$

Where $N_k=\sum_{i=1}^{N}\gamma_{ik}$

聚类：k-means、GMM

- 极大似然估计 - Maximum Likelihood Estimate
- 类似的方式求解$\boldsymbol{\Sigma}_k$与 π_k

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma_{ik}\,\mathbf{x}_i$$

Where $N_k = \sum_{i=1}^{N}\gamma_{ik}$

$$\pi_k = \frac{N_k}{N}$$

推导过程冗长，请参考word文档

名称	类型	大小	修改日期
miscGMM的极大似然估计(AddOn).docx	Microsoft Word ...	24 KB	2021/11/18 19:15

聚类：k-means、GMM