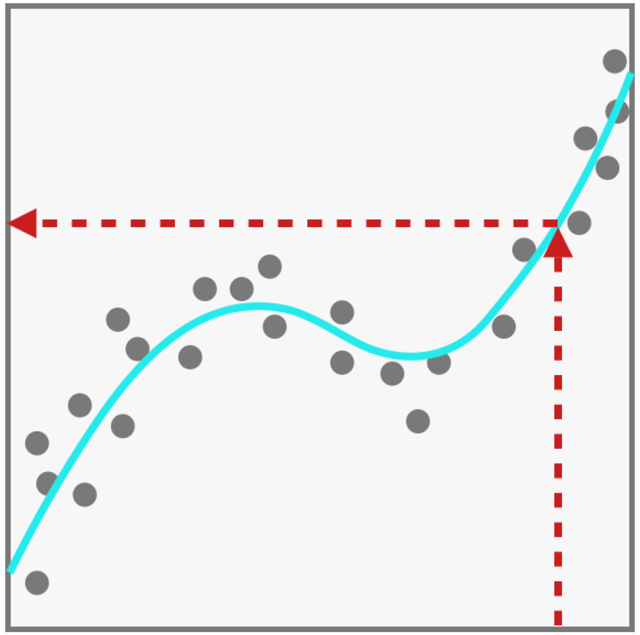
模式识别
Pattern Recognition

李泽榉，复旦大学 生物医学工程与技术创新学院

目录

Recap：线性回归

- 通过已知的输入数据（特征）预测一个连续的数值输出。

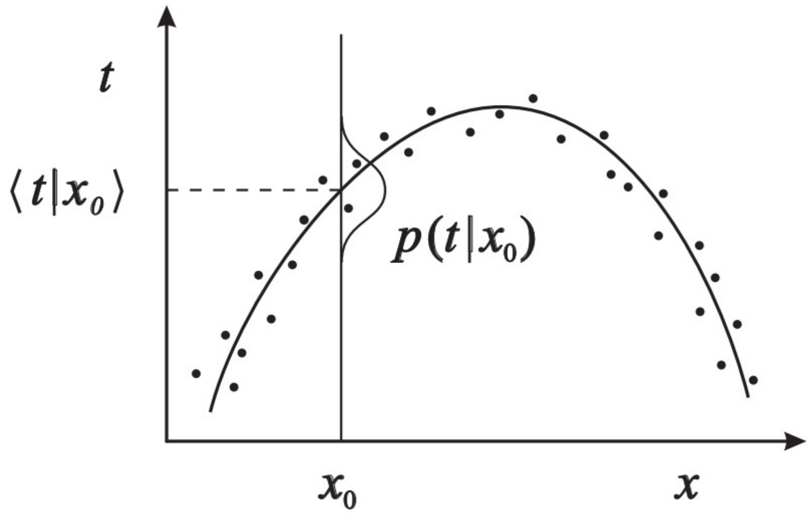


Recap：线性回归

- 对于回归任务的MLE估计本质就是高斯假设下的最小二乘法

Example (contd.)

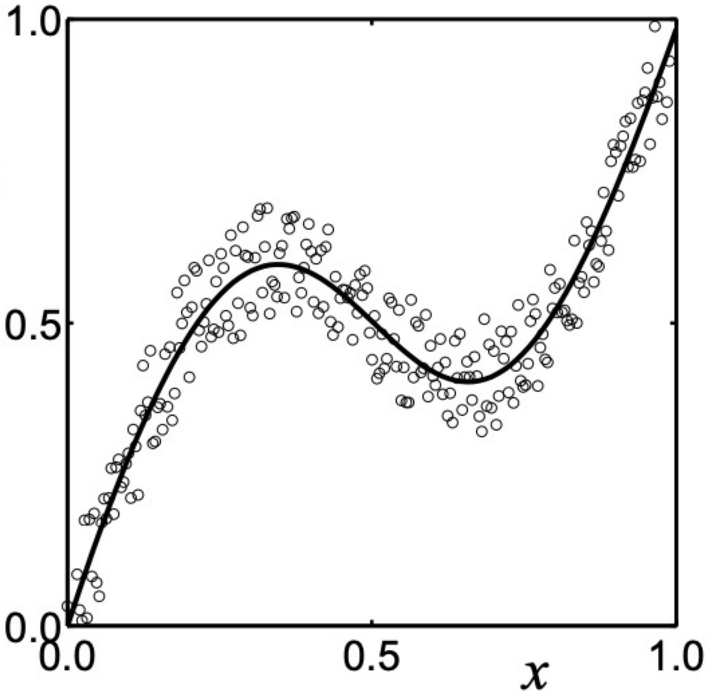
$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= -\sum_{i=1}^{N} \log p(y_i \mid \mathbf{x}_i, \boldsymbol{\theta}) = -\sum_{i=1}^{N} \log \mathcal{N}(y_i \mid \mathbf{x}_i^\top \boldsymbol{\theta}, \sigma^2) \\
&= -\sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\theta})^2}{2\sigma^2}\right) \\
&= -\sum_{i=1}^{N} \log \exp\left(-\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\theta})^2}{2\sigma^2}\right) - \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi\sigma^2}} \\
&= \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \mathbf{x}_i^\top \boldsymbol{\theta})^2 - \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi\sigma^2}}.
\end{aligned}$$



The second term is constant.
$\Longrightarrow$ minimizing $\mathcal{L}(\boldsymbol{\theta}) \Rightarrow$ solving the least-squares problem.

线性回归的局限

- 建立一个模型，使其能够根据**<u>输入的特征</u>**预测**<u>一个</u>**数值型的目标变量

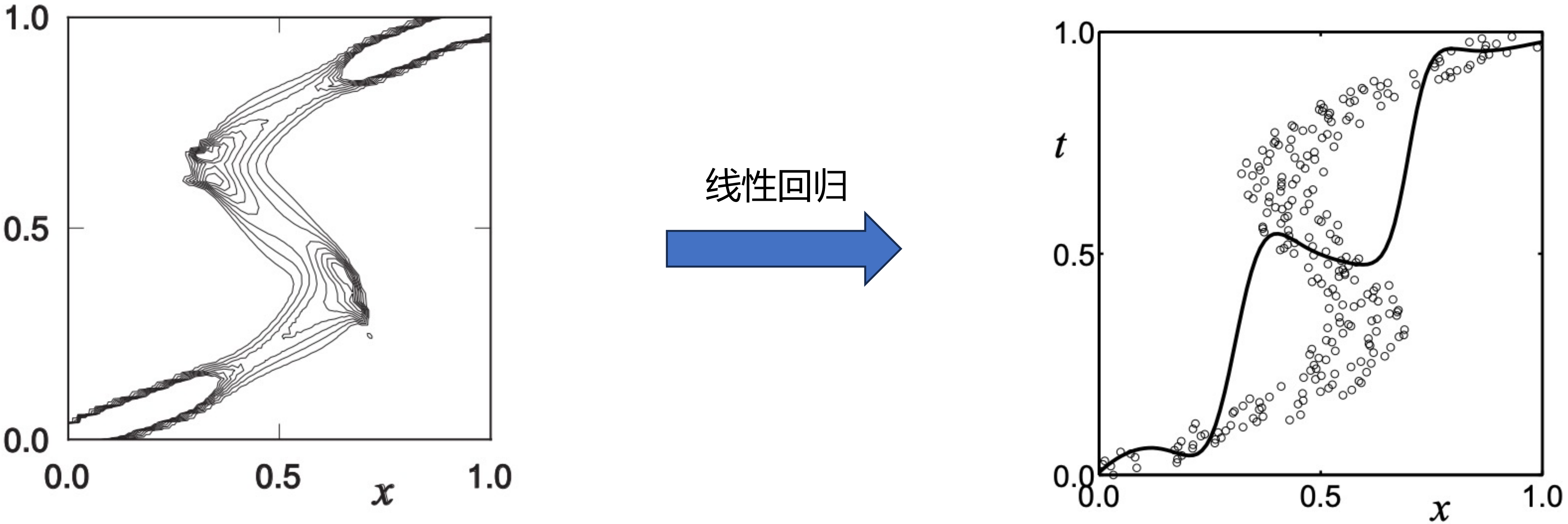


- But wait，类似的输入特征一定是同样的解吗？

线性回归的局限

- How about this?

- 同一个输入特征，有多个可能的输出

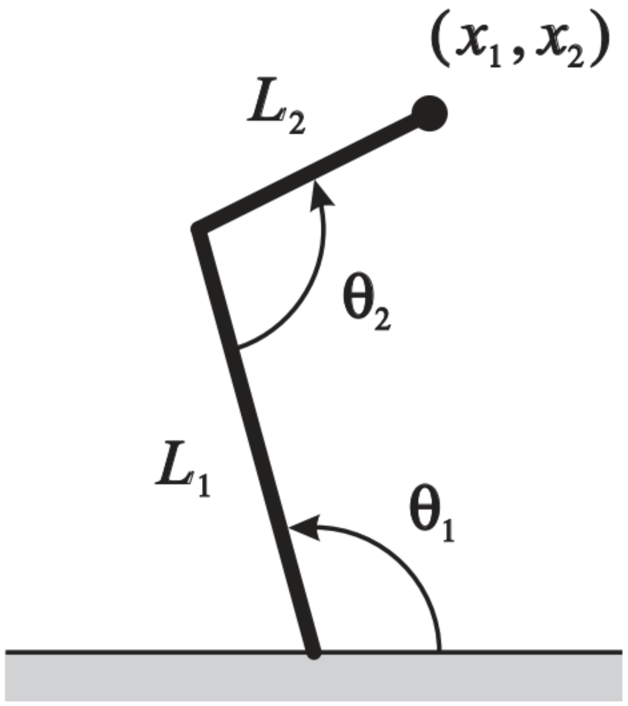


线性回归的局限

- 现实中的例子：机械臂
- 需要从观测得到的(x_1, x_2)推断出(θ_1, θ_2)

$$x_1 = L1\cos(\theta_1) - L2\cos(\theta_1 + \theta_2)$$

$$x_2 = L1\sin(\theta_1) - L2\sin(\theta_1 + \theta_2)$$



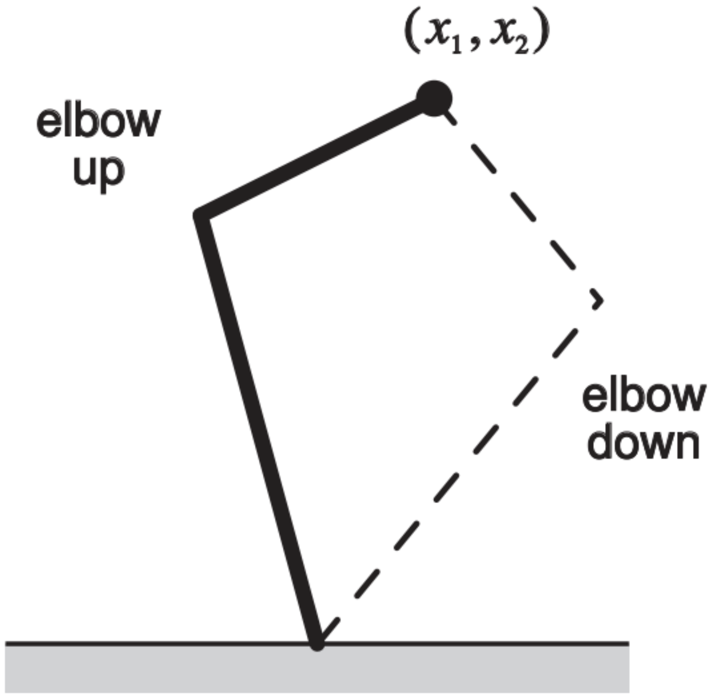
线性回归的局限

- 现实中的例子：机械臂

- 需要从观测得到的(x_1, x_2)推断出(θ_1, θ_2)

$$x_1 = L1\cos(\theta_1) - L2\cos(\theta_1 + \theta_2)$$

$$x_2 = L1\sin(\theta_1) - L2\sin(\theta_1 + \theta_2)$$



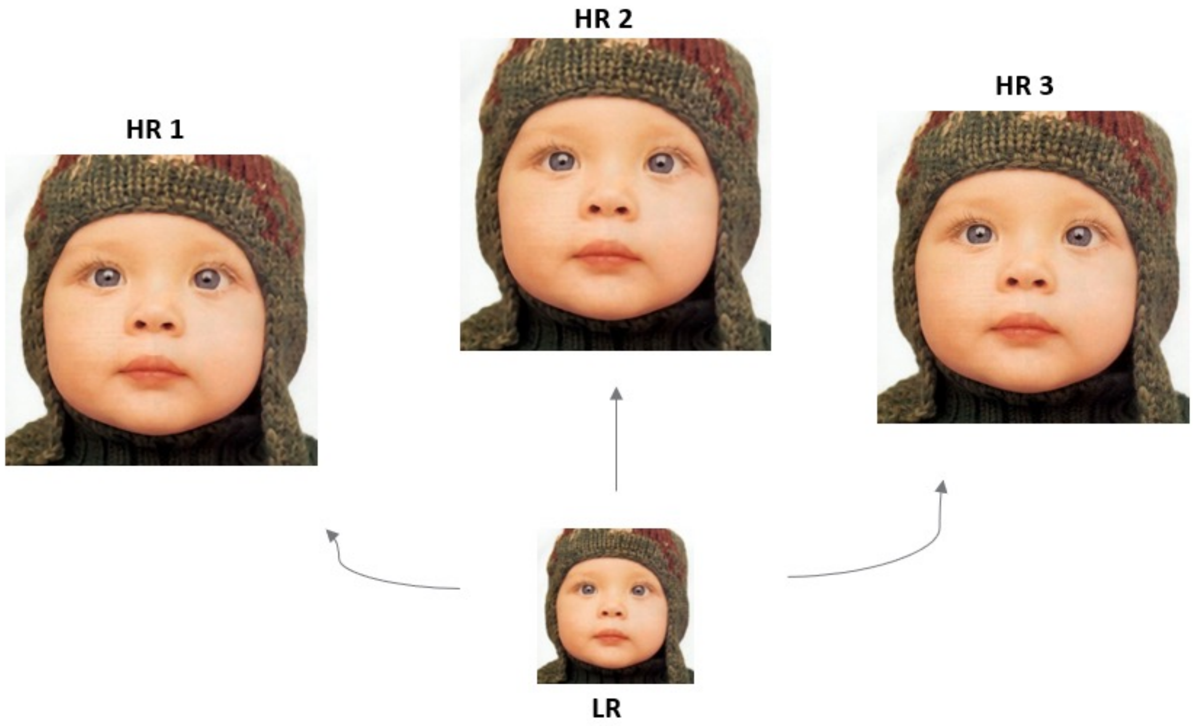
- 但是问题也是有两个解的

线性回归的局限

- 现实中的例子：图像重建

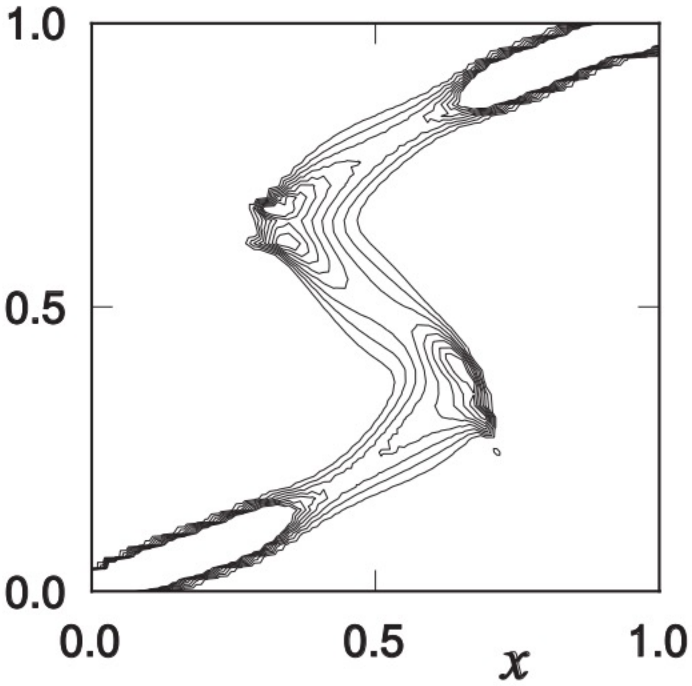
- 一个多分辨率，可以对应多个高分辨率

- “ill-posed”problem

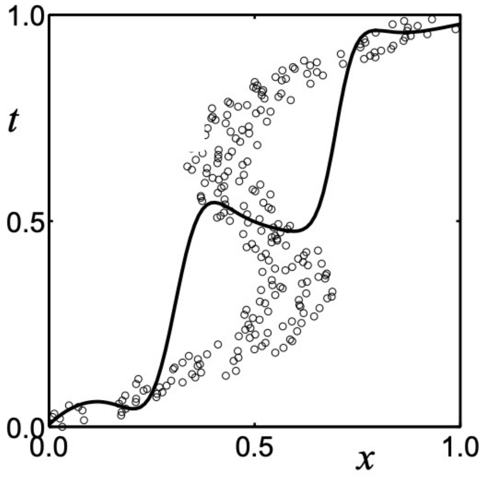


混合模型

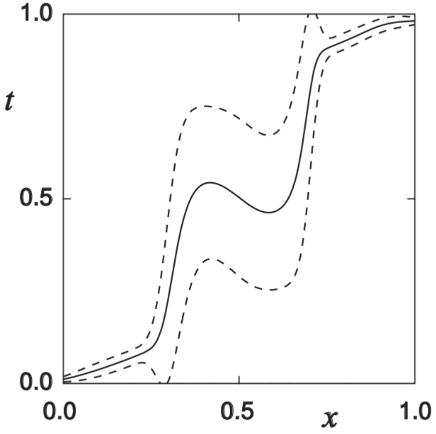
- 混合模型可以对多输出进行建模
- 对于同一个输入特征，取最大可能输出

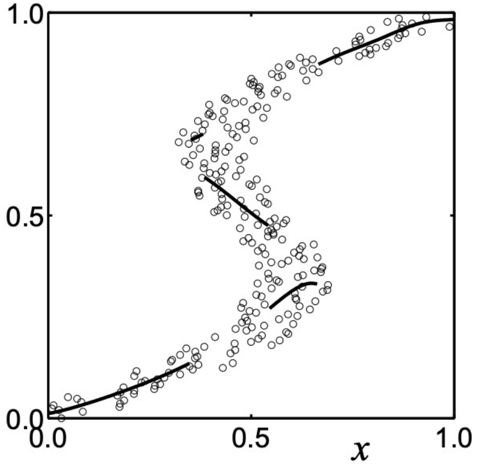


线性回归



混合模型





混合模型

- 利用K的简单分布的凸集合来表示一个复杂的分布

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k p_k(\mathbf{x}),$$

$$0 \le \pi_k \le 1, \sum_{k=1}^{K} \pi_k = 1.$$

π_k: *mixture weights.*

高斯混合模型

- 为了简单，假设每一个子分布都服从高斯分布

Gaussian Mixture Model

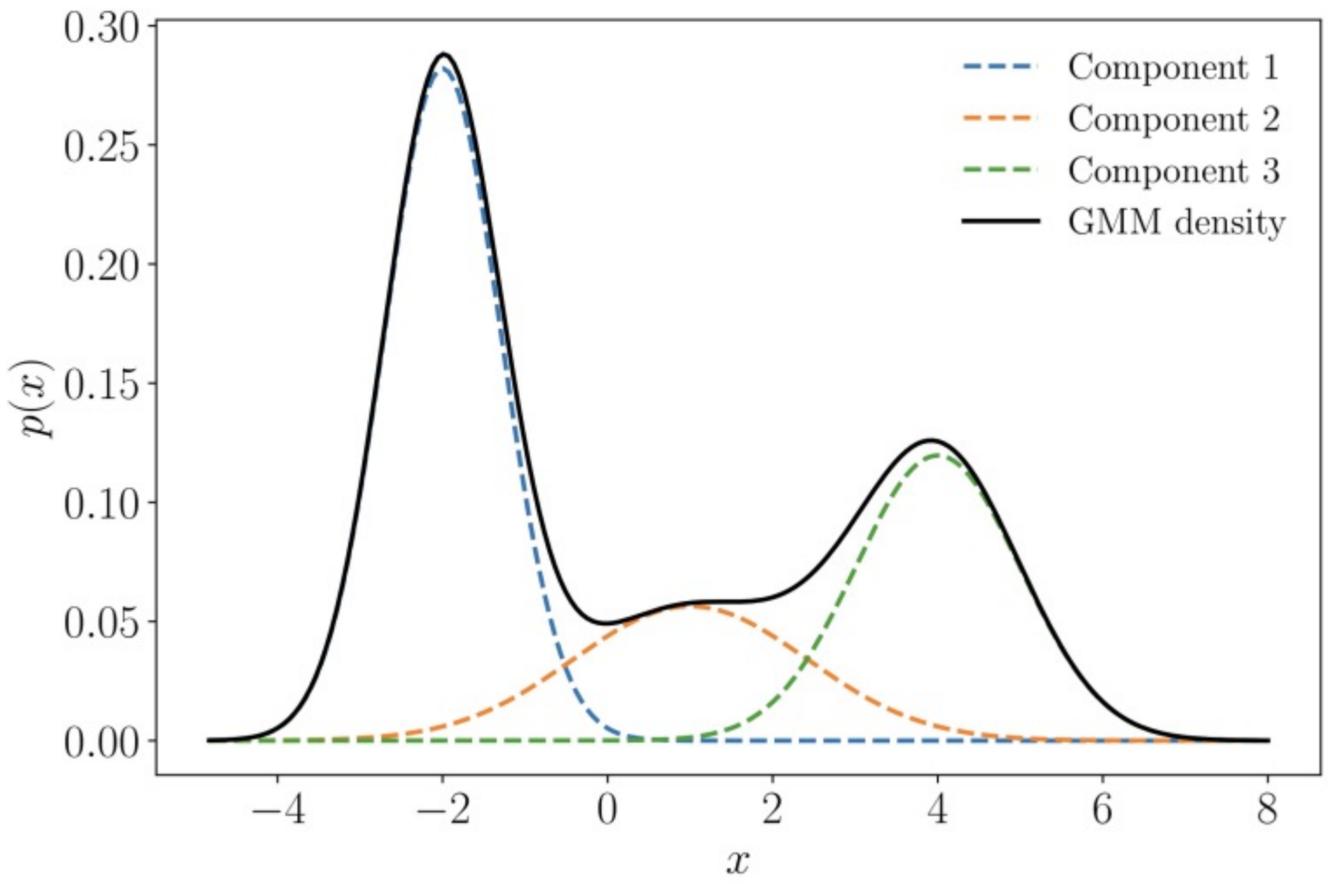
A Gaussian mixture model is a density model where we combine a finite number of K Gaussian distributions $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ such that

$$p(\mathbf{x} \mid \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$0 \leq \pi_k \leq 1, \sum_{k=1}^{K} \pi_k = 1,$$

where $\boldsymbol{\theta} := \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k \mid k = 1, \ldots, K\}$.

高斯混合模型



$$p(x \mid \boldsymbol{\theta}) = 0.5\mathcal{N}(x \mid -2, 0.5) + 0.2\mathcal{N}(x \mid 1, 2) + 0.3\mathcal{N}(x \mid 4, 1).$$

目录

模型优化

- 问题假设

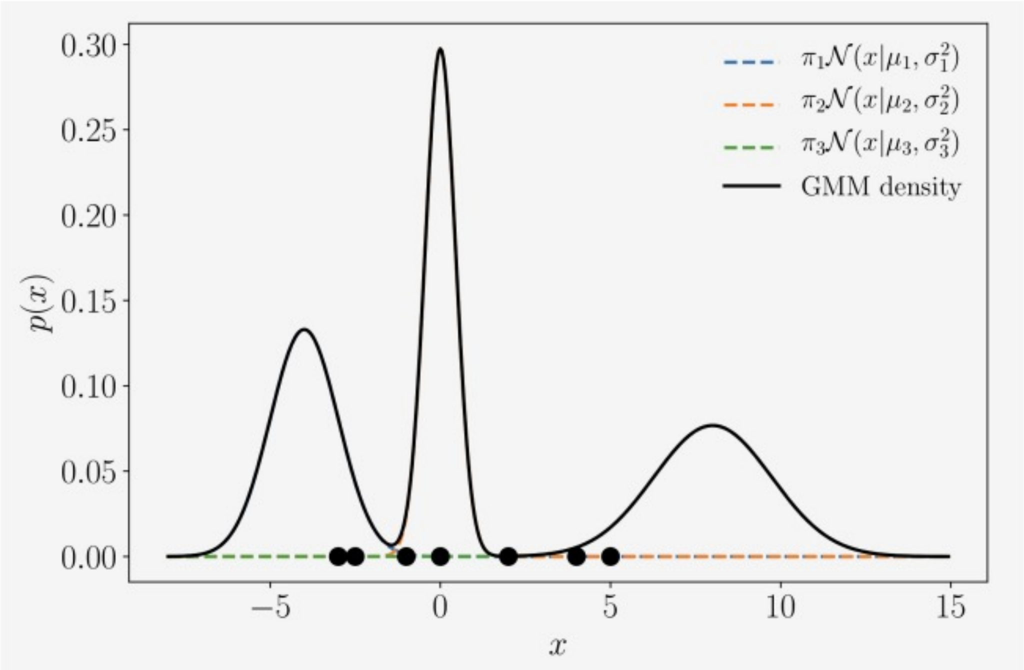
A dataset $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, where each $\mathbf{x}_i$ is drawn i.i.d. from an unknown distribution $p(\mathbf{x})$.

Parameters: $\boldsymbol{\theta} := \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k \mid k = 1, \ldots, K\}$.

模型优化

- 初始状态

$$\mathcal{X} = \{-3, -2.5, -1, 0, 2, 4, 5\}.$$
$$K = 3.$$
$$p_1(x) = \mathcal{N}(x \mid -4, 1),\ p_2(x) = \mathcal{N}(x \mid 0, 0.2),\ p_3(x) = \mathcal{N}(x \mid 8, 3).$$
$$\pi_1 = \pi_2 = \pi_3 = 1/3.$$



模型优化

- 利用最大似然估计（MLE）来计算参数

By the i.i.d. assumption, we have the factorized likelihood

$$p(\mathcal{X} \mid \boldsymbol{\theta}) = \prod_{i=1}^{N} p(\mathbf{x}_i \mid \boldsymbol{\theta}), \quad p(\mathbf{x}_i \mid \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).$$

Then the log-likelihood is

$$\mathcal{L} := \log p(\mathcal{X} \mid \boldsymbol{\theta}) = \sum_{i=1}^{N} \log p(\mathbf{x}_i \mid \boldsymbol{\theta}) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).$$

模型优化

- 问题是，需要优化的参数太多
- 所以用EM算法来优化参数
- 本质上就是一个时间只优化一个参数，固定其他参数

模型优化

- 对三个变量分别分析

Necessary conditions for a local optimum of $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}_k} = \mathbf{0}^\top \iff \sum_{i=1}^{N} \frac{\partial \log p(\mathbf{x}_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k} = \mathbf{0}^\top$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\Sigma}_k} = \mathbf{0}^\top \iff \sum_{i=1}^{N} \frac{\partial \log p(\mathbf{x}_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_k} = \mathbf{0}^\top$$

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = 0 \iff \sum_{i=1}^{N} \frac{\partial \log p(\mathbf{x}_i \mid \boldsymbol{\theta})}{\partial \pi_k} = 0.$$

Applying the chain rule:

$$\frac{\partial \log p(\mathbf{x}_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{p(\mathbf{x}_i \mid \boldsymbol{\theta})} \frac{\partial p(\mathbf{x}_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

and

$$\frac{1}{p(\mathbf{x}_i \mid \boldsymbol{\theta})} = \frac{1}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

模型优化

- 对三个变量分别分析
- 分别进行推导可以得到

Theorem [Update of the Means]

The update of the mean parameters $\boldsymbol{\mu}_k$, $k = 1, \dots, K$, of the GMM is given by

$$\boldsymbol{\mu}_k^{new} = \frac{\sum_{i=1}^N r_{ik} \mathbf{x}_i}{\sum_{i=1}^N r_{ik}}.$$

Theorem [Update of the Covariances]

The update of the covariance parameters $\boldsymbol{\Sigma}_k$, $k = 1, \dots, K$, of the GMM is given by

$$\boldsymbol{\Sigma}_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top,$$

where

$$r_{ik} := \frac{\pi_k \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

and $N_k := \sum_{i=1}^N r_{ik}$.

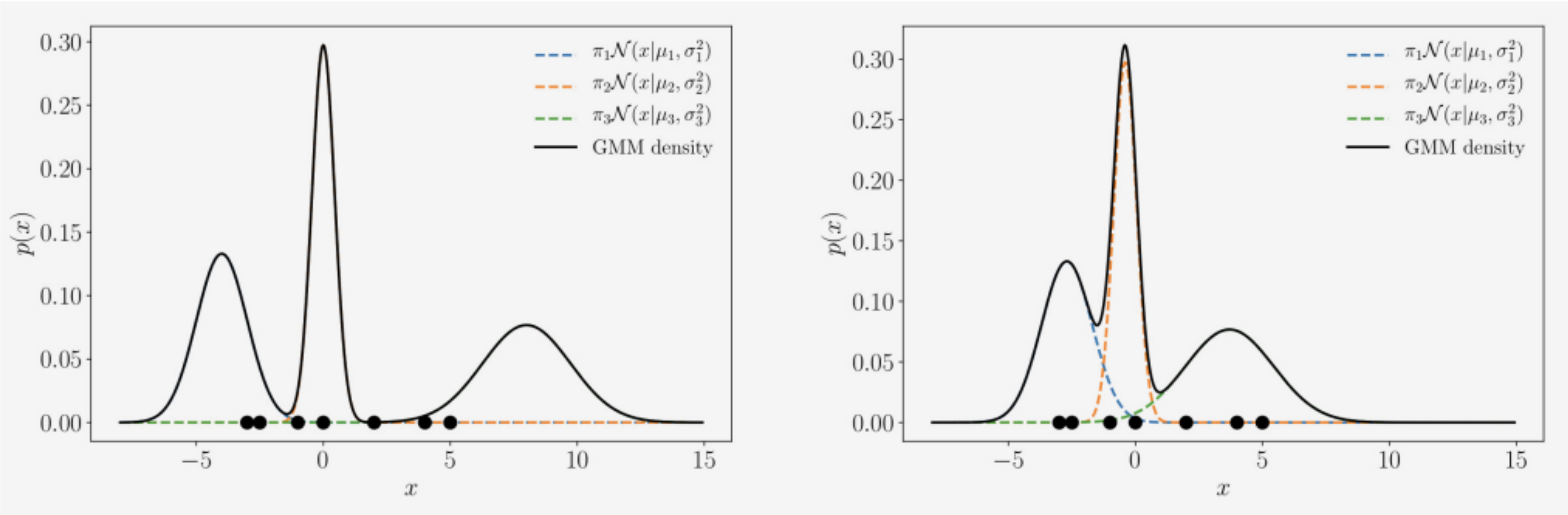
Theorem [Update of the Mixture Weights]

The update of the mixture weights of the GMM is given by

$$\pi_k^{new} = \frac{N_k}{N}, \quad k = 1, \dots, K.$$

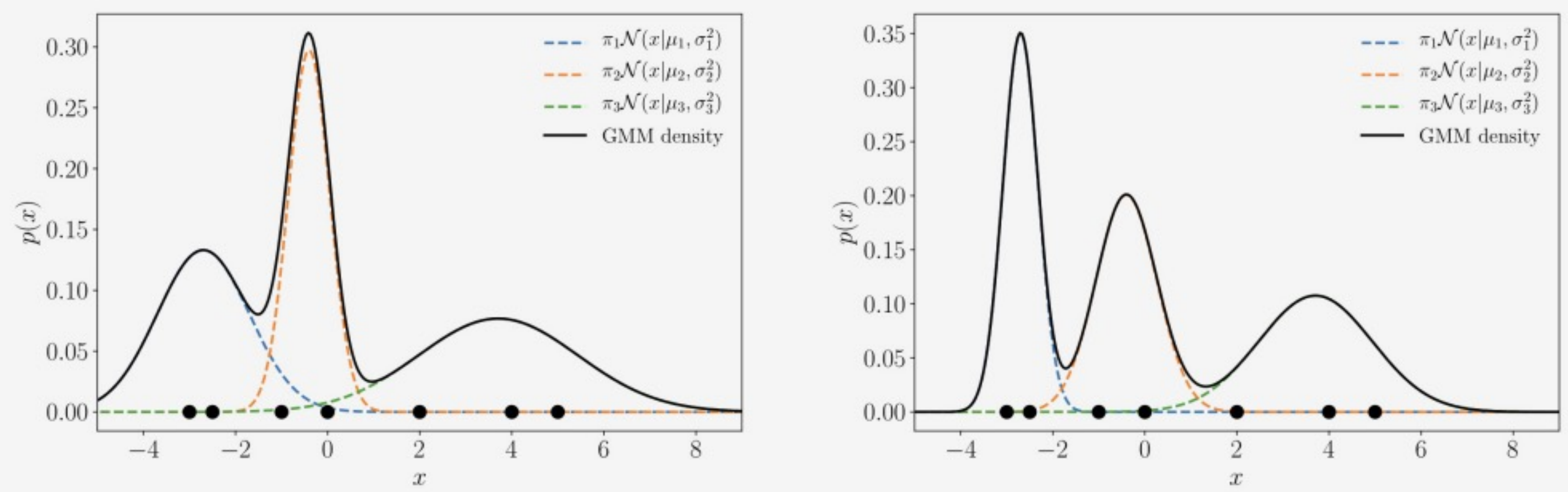
- N: the number of data points.
- $N_k := \sum_{i=1}^N r_{ik}$.

模型优化



- $\mu_1 : -4 \to -2.7.$
- $\mu_2 : 0 \to -0.4.$
- $\mu_3 : 8 \to 3.7.$

模型优化

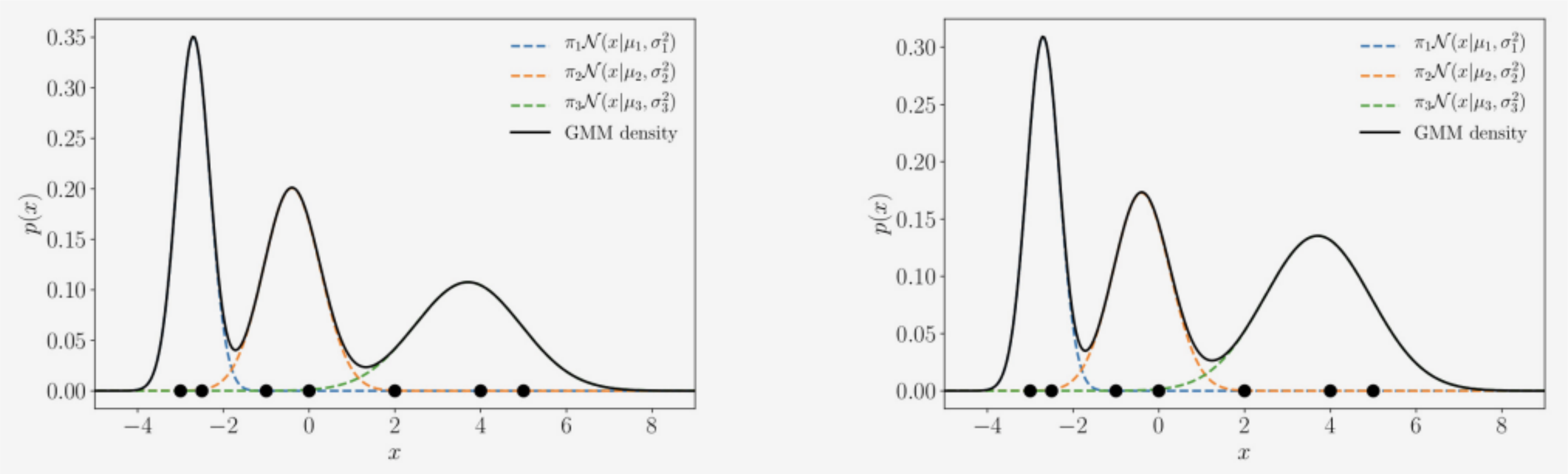


(a) GMM density and individual components prior to updating the variances.

(b) GMM density and individual components after updating the variances.

- $\sigma_1^2 : 1 \to 0.14.$
- $\sigma_2^2 : 0.2 \to 0.44.$
- $\sigma_3^2 : 3 \to 1.53.$

模型优化

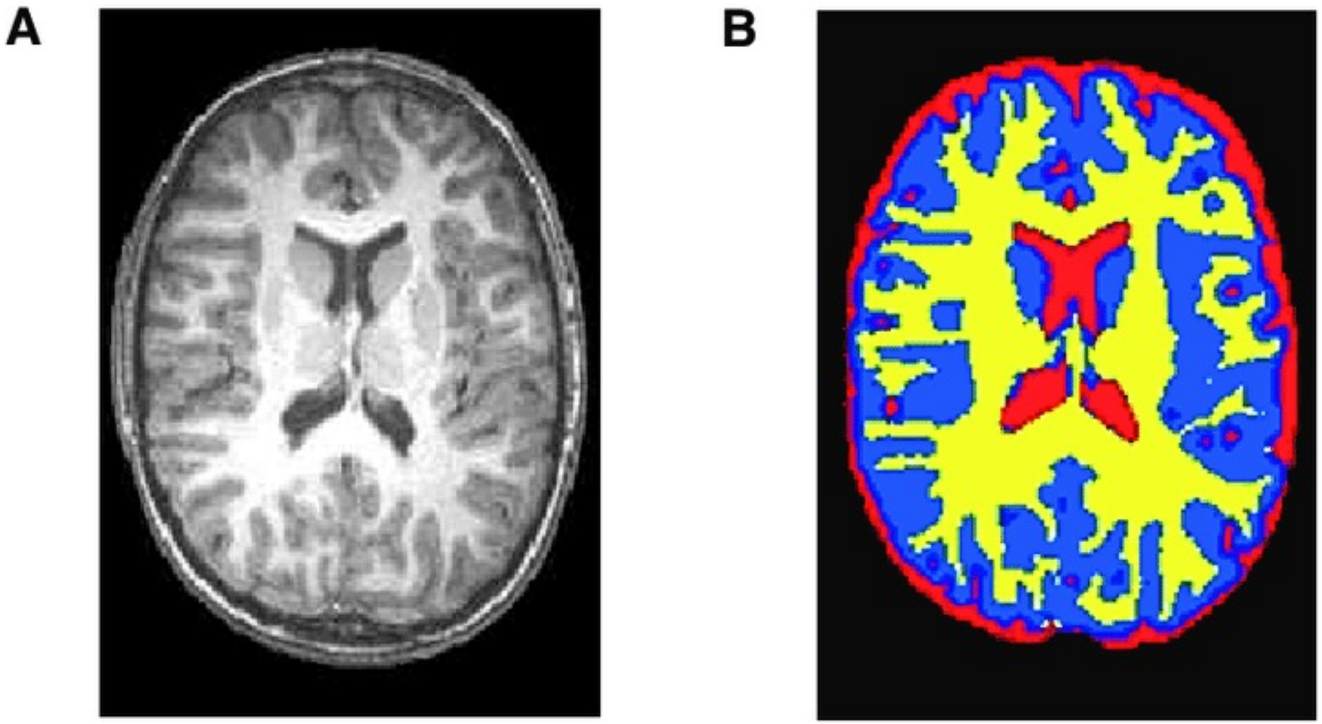


- $\pi_1 : \frac{1}{3} \to 0.29.$
- $\pi_2 : \frac{1}{3} \to 0.29.$
- $\pi_3 : \frac{1}{3} \to 0.42.$

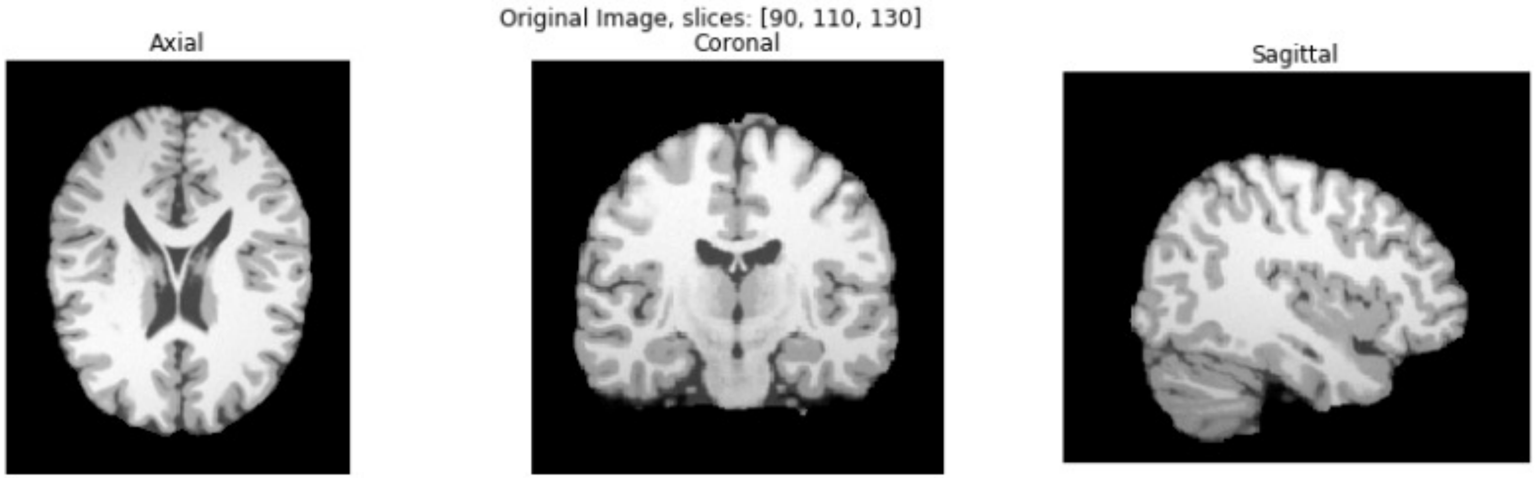
目录

医学图像分割

- 将MRI图像分割成为GM，WM和CSF



医学图像分割


Original Image, slices: [90, 110, 130]
Axial
Coronal
Sagittal

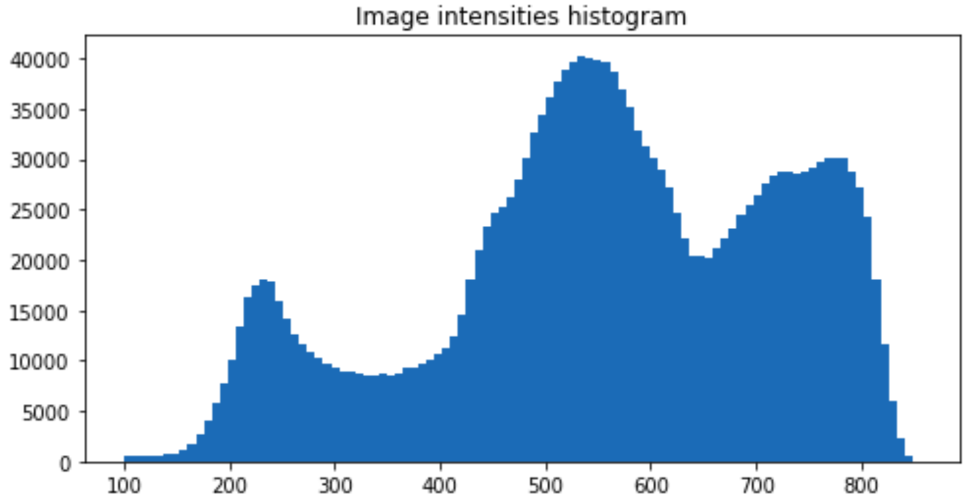


Image intensities histogram
40000
35000
30000
25000
20000
15000
10000
5000
0
100
200
300
400
500
600
700
800

医学图像分割

- 假设D是图像，I是期望得到的标注
- 我们希望最大化p$(I|D)$
- 直接优化就是如果有K-means，只能得到hard label，那么这个决定是武断的。我们丢失了“这个点很有可能是白质”的重要信息。

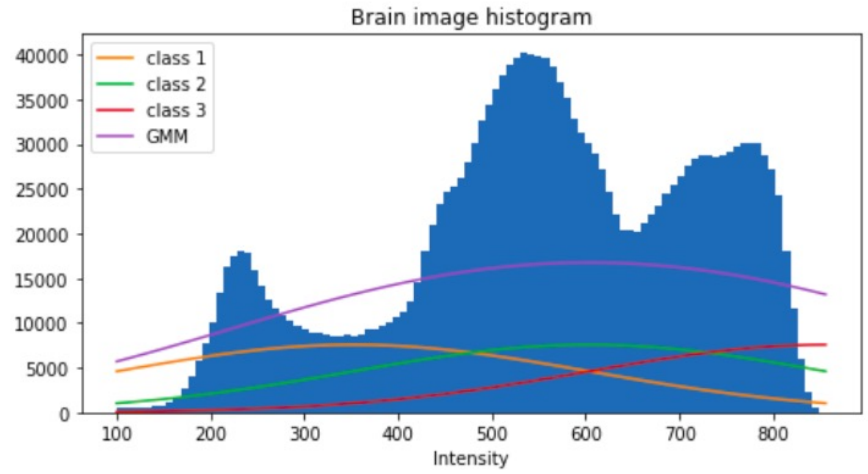
医学图像分割

- 根据贝叶斯法则，我们有$\mathrm{p}(I|D) \propto \mathrm{p}(D|I)p(I)$
- 分割->不同组织下的后验概率加权之和最大

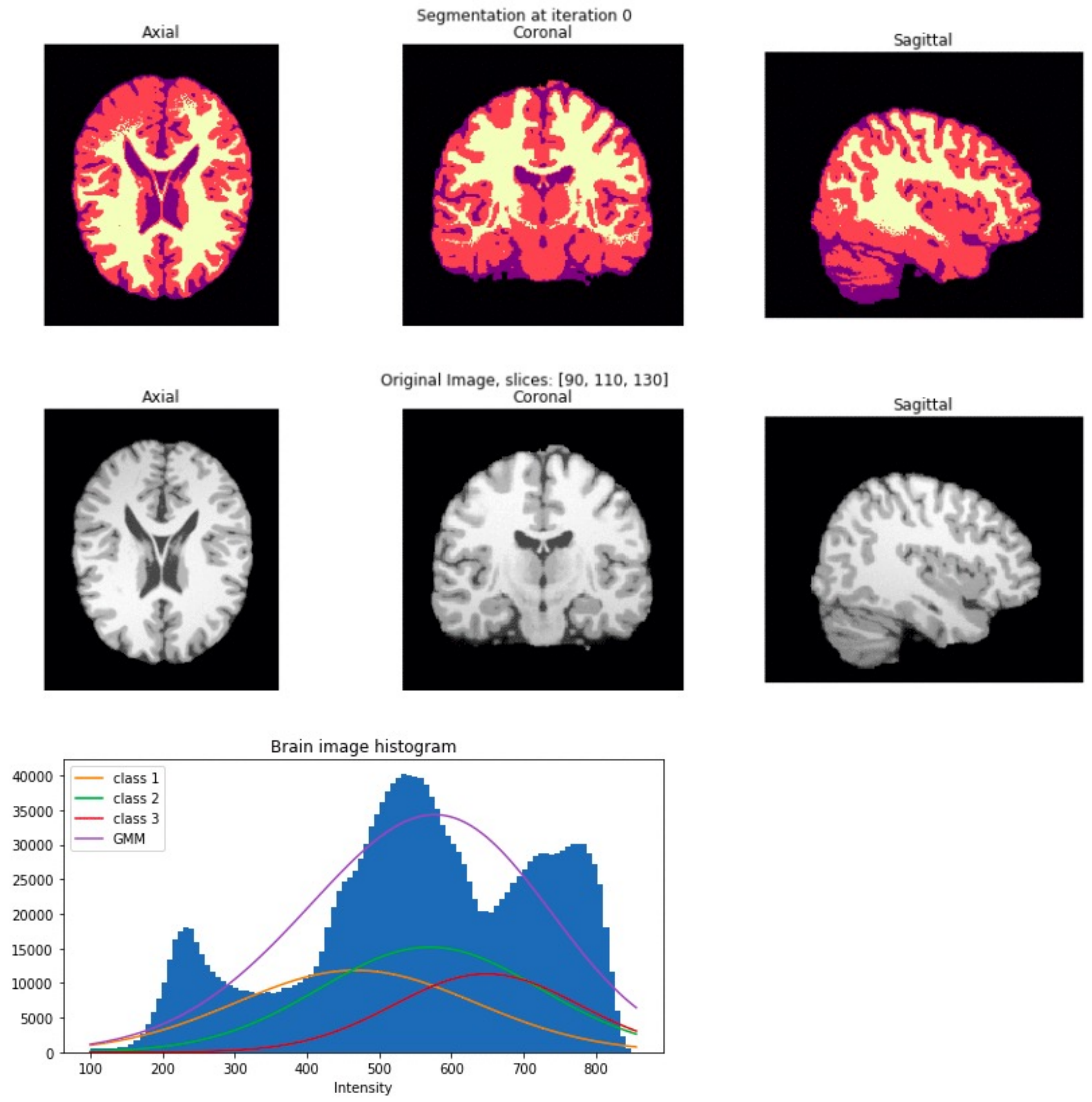
医学图像分割

- 根据贝叶斯法则，我们有p$(I|D) \propto \boxed{\mathrm{p}(D|I)}\ \boxed{p(I)}$
- 分割->不同组织下的后验概率加权之和最大

$$\ln p(\mathbf{D}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi) = \sum_{i}^{I} \ln\{\sum_{k}^{K} \boxed{\pi_k} \boxed{\mathcal{N}(\mathbf{d}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}\}.$$

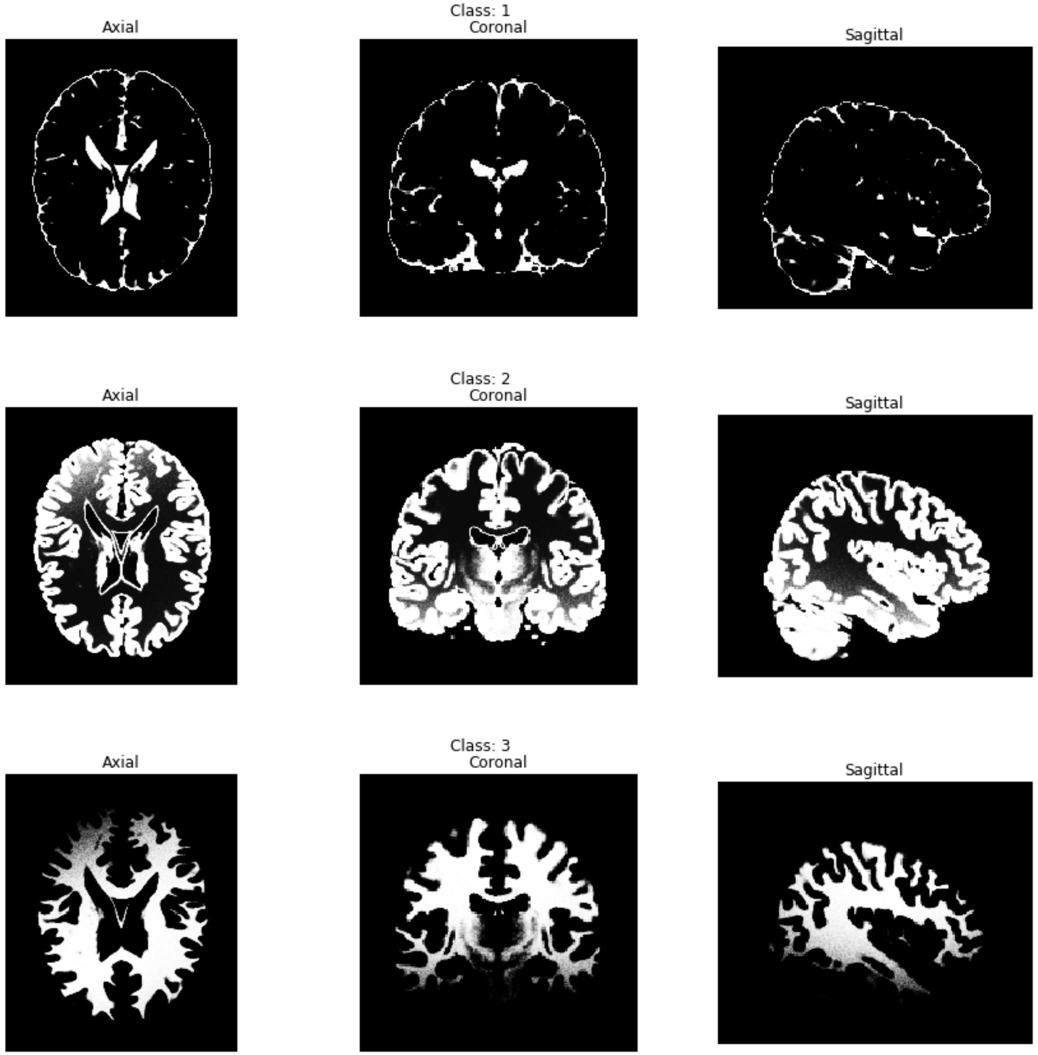


医学图像分割



Segmentation at iteration 0
Axial
Coronal
Sagittal
Original Image, slices: [90, 110, 130]
Axial
Coronal
Sagittal
Brain image histogram
class 1
class 2
class 3
GMM
40000
35000
30000
25000
20000
15000
10000
5000
0
100
200
300
400
500
600
700
800
Intensity

医学图像分割


Class: 1
Axial
Coronal
Sagittal
Class: 2
Axial
Coronal
Sagittal
Class: 3
Axial
Coronal
Sagittal

小结

- 回归模型不能对多输出情况进行建模
- 混合模型可以通过生成模型的思想建模多输出问题
- 高斯混合模型本质是用EM算法来进行优化的
- 高斯混合模型是k-means聚类的一个概率化形式
- EM算法不保证全局最优，但往往总是有效