Mixture of Gaussians

1. Mixture of Gaussians for Classification.

In an unsupervised classification problem, we are given data $\{\mathrm{x}_i\}_{i=1}^n$ and a number of classes m, and we try to learn:

$$p(y|\mathrm{x}) = \frac{p(\mathrm{x}|y)p(y)}{p(x)} = \frac{p(\mathrm{x}|y)p(y)}{\sum_{j=1}^m p(\mathrm{x}|y=j)p(y=j)}, \qquad (1)$$

where $y \in 1 \dots m$ is the class label, with the hypothesis:

$$y \sim \mathrm{Multinomial}(\phi) \qquad (2)$$

$$(\mathrm{x}|y=j) \sim \mathcal{N}(\boldsymbol{\mu}_j, \Sigma_j) \qquad \forall j \in 1 \dots m. \qquad (3)$$

Hence, the parameters of this model are ϕ, $\{\boldsymbol{\mu}_j\}_{j=1}^m$, and $\{\Sigma_j\}_{j=1}^m$. Since the labels y_i of each data point are unknown, we need the EM-algorithm to find locally-maximal log-likelihood estimates for ϕ, $\{\boldsymbol{\mu}_j\}_{j=1}^m$, and $\{\Sigma_j\}_{j=1}^m$.

The attached file `data5a.txt` has two numbers a and n on the first line; a is the dimension of the data, and n is the number of data points. Then n lines follow with a numbers: $x_1, \dots, x_a$ that constitute the data $\{\mathrm{x}_i\}_{i=1}^n$, where $\mathrm{x}_i \in \mathbb{R}^a$.

Plot the data points. Based on the plot, how many classes m do you think the data set should be classified into?

Implement the EM-algorithm to learn the parameters ϕ, $\{\boldsymbol{\mu}_j\}_{j=1}^m$, and $\{\Sigma_j\}_{j=1}^m$ for a given dataset $\{\mathrm{x}_i\}_{i=0}^n$ and a given number of classes m. Apply the EM-algorithm to the dataset using the number of classes m chosen in the previous question. How did you initialize the EM-algorithm? How does the initialization affect the result? Describe what you see. Shows plots of the means and the 1-sigma contours of the Gaussian distributions $\mathrm{x}|y=j$ for $j \in 1 \dots m$ on top of the data upon convergence. show plots of the evolution of the expected log-likelihood during the iteration of the EM-algorithm.

Mixture of Gaussians for Non-linear Regression

for a given number m of constituent Gaussians, where $\mathcal{N}(\mathrm{x}; \boldsymbol{\mu}, \Sigma)$ is the probability density at x of a Gaussian distribution with mean $\boldsymbol{\mu}$ and variance Σ:

$$\mathcal{N}(\mathrm{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^a |\Sigma|}} \exp(\frac{1}{2}(\mathrm{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathrm{x} - \boldsymbol{\mu})), \qquad (5)$$

where a is the dimension of x.

Also here, ϕ, $\{\boldsymbol{\mu}_j\}_{j=1}^m$, and $\{\Sigma_j\}_{j=1}^m$ are the parameters of the model, which can be learned using the same EM-algorithm as above given a set of data.

a)

Thus, the expected value and variance of $\mathbf{x}$ is given by,

$$\mu = E(\mathbf{x}) = \sum_{j=1}^m \phi_j \mu_j$$

$$\Sigma = Var(\mathbf{x}) = \sum_{i=1}^m \phi_j \Sigma_j + \sum_{j=1}^m \phi_j (\mu_j - \mu)(\mu_j - \mu)^T$$

b)

Given $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ where $\mathbf{x}$ and $\mathbf{y}$ are draw from different mixture of Gaussians respectively. The marginal distribution of $\mathbf{x}$ is nothing but the same mixture of Gaussians from which it is derived. Thus,

$$p(\mathbf{x}) = \sum_{j=1}^{m} \phi_j \mathcal{N}(\mu_j^x, \Sigma_j^x)$$

c)

The conditional distribution of $\mathbf{x}|\mathbf{y}$ is also mixture of Gaussians given by,

$$p(\mathbf{x}|\mathbf{y}) = \sum_{j=1}^{m} \phi_j^{\mathbf{x}|\mathbf{y}} \mathcal{N}(\mathbf{x}; \mu_j^{\mathbf{x}|\mathbf{y}}, \Sigma_j^{\mathbf{x}|\mathbf{y}})$$

The parameters of the this distribution are given as,

$$\begin{aligned} \phi_j^{\mathbf{x}|\mathbf{y}} &= \frac{\phi_j \mathcal{N}(\mathbf{y}; \mu_j^y, \Sigma_j^y)}{\sum_{k=1}^{m} \phi_k \mathcal{N}(\mathbf{y}; \mu_k^y, \Sigma_k^y)} \\ \mu_j^{\mathbf{x}|\mathbf{y}} &= \mu_j^x + \Sigma_j^{xy}(\Sigma_j^y)^{-1}(\mathbf{y} - \mu_j^y) \\ \Sigma_j^{\mathbf{x}|\mathbf{y}} &= \Sigma_j^y - \Sigma_j^{yx}(\Sigma_j^x)^{-1}\Sigma_j^{xy} \end{aligned}$$

The attached file `data5b.txt` has one number n on the first line, the number of data points. Then n lines follow with 2 numbers: x and y, that constitute the data $\{x_i, y_i\}_{i=1}^{n}$, where $x_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$.

(d) Plot the data and compute the parameters of the joint mixture of Gaussians distribution of $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ for varying number m of constituent Gaussians using the EM-algorithm. Plot the means and 1-sigma contours of the constituent Gaussians on top of the data.

(e) Plot the graphs of $E(y|x)$ and the 1-sigma countours $E(y|x) \pm \sqrt{(\mathrm{Var}(y|x))}$ as functions of x on top of the data. What happens if you change the number m of constituent Gaussians.