Machine Learning - 2020-2021
Universitat Pompeu Fabra

Problems 3: Mixture of Gaussian

Goal

The goal of this practice is to understand how Gaussian Mixture Models (GMM) are defined and to learn to estimate a model built not only from one Gaussian but from several (Gaussian Mixture Models). Also, GMM can be used to classify in a similar way as K-means.

Needed Concepts

- Basis of a vector space and change of basis
  If $\mathcal{V} = \{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}\}$ is a basis of $\mathbb{R}^2$ (in cartesian coordinates), then

$$\mathbf{P} = \begin{pmatrix} \vdots & \vdots \\ \mathbf{v}^{(1)} & \mathbf{v}^{(2)} \\ \vdots & \vdots \end{pmatrix}$$

  is the matrix that changes a vector $\mathbf{x}_{\mathcal{V}}$ written in coordinates of the basis $\mathcal{V}$ to $\mathbf{x}_{\mathcal{E}}$ in cartesian coordinates, i.e., $\mathbf{x}_{\mathcal{E}} = \mathbf{P}\ \mathbf{x}_{\mathcal{V}}$.
  Reminder: if $\{v_1, v_2\}$ are orthonormal, then $\mathbf{P}^{-1} = \mathbf{P}^\top$.

- Singular Value Decomposition (SVD): used to diagonalise non-square matrices $A \in \mathbb{R}^{n\times m}$.

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$$

  where $\mathbf{U} \in \mathbb{R}^{n\times n}$, $\mathbf{D} \in \mathbb{R}^{n\times m}$ is a diagonal matrix and $\mathbf{V} \in \mathbb{R}^{m\times m}$.
  Reminder: The columns of $\mathbf{U}$ and $\mathbf{V}$ define new **orthonormal** bases in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

- Gaussian distribution (or multivariate normal distribution)

$$PDF : f(\mathbf{x}) = (2\pi)^{-\frac{k}{2}} \det(\mathbf{\Sigma})^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right]$$

  where k is the dimension of $\mathbf{x}$.

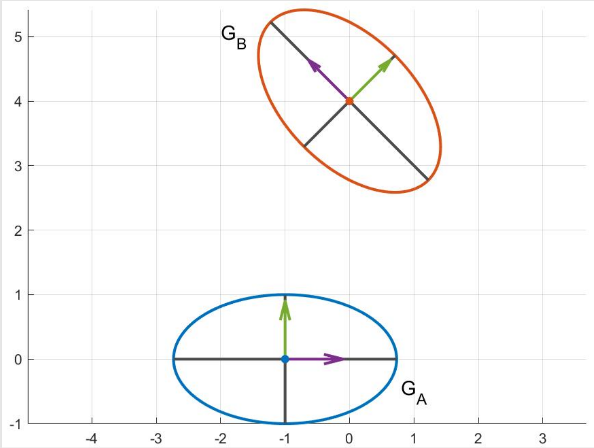
Exercises

1. Consider two different Gaussian models, G_A and G_B with means and covariances:

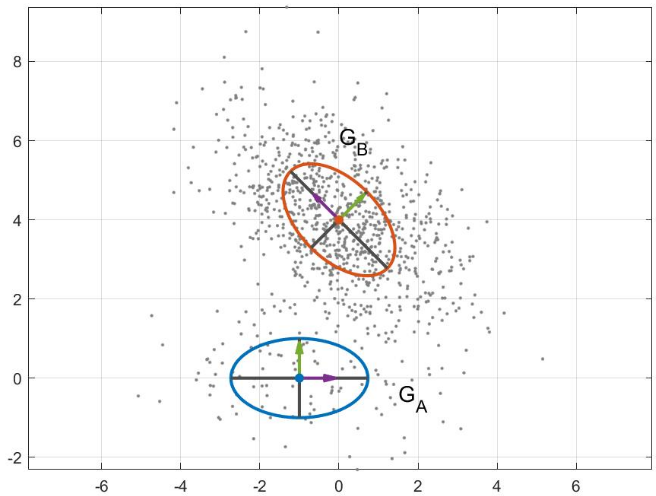
$$\mu_A = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \mathbf{\Sigma}_A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\mu_B = \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \mathbf{\Sigma}_B = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

   (a) Draw the ellipses corresponding to the Mahalanobis distances 1 of both Gaussian models.

   Solution:
   SVD to $\mathbf{\Sigma}_A \longrightarrow$ (not needed) $\lambda_1^A = 3, \mathbf{u}_A^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \lambda_2^A = 1, \mathbf{u}_A^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
   SVD to $\mathbf{\Sigma}_B \longrightarrow \quad \lambda_1^B = 3, \mathbf{u}_B^{(1)} = \frac{\sqrt{2}}{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \lambda_2^B = 1, \mathbf{u}_B^{(2)} = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$



(b) (*In Jupyter Notebook*) Draw 1000 random samples from a mixture of Gaussians G_A and G_B for different values of the mixing coefficients (modeling point of view).
For example, this figure shows 1000 random samples using $\pi_A = 0.1$ and $\pi_B = 0.9$.

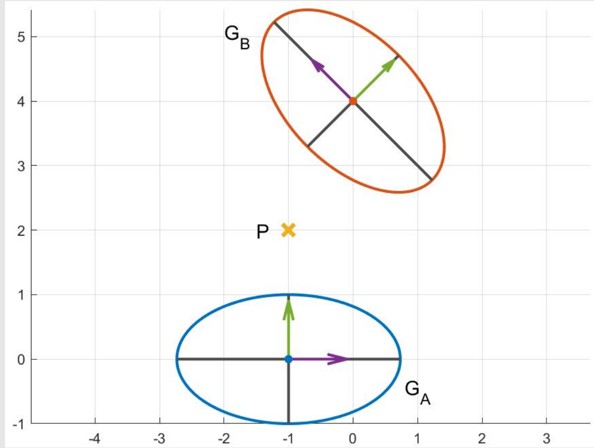


(c) Estimate the probability a posteriori of a point $P = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ of being generated by G_A and G_B (clustering point of view) given mixing coefficient $\pi_A = 0.1$ and $\pi_B = 0.9$.

Solution:

$$\left.\begin{aligned} P(P|G_A) = \mathcal{N}(P|\mu_1, \mathbf{\Sigma}_1) = (2\pi)^{-1}\det(\mathbf{\Sigma}_1)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(P-\mu_1)^\top \mathbf{\Sigma}_1^{-1}(P-\mu_1)\right] = 0.0124 \\ P(P|G_B) = \mathcal{N}(P|\mu_2, \mathbf{\Sigma}_2) = (2\pi)^{-1}\det(\mathbf{\Sigma}_2)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(P-\mu_2)^\top \mathbf{\Sigma}_2^{-1}(P-\mu_2)\right] = 0.0089 \end{aligned}\right\} \Longrightarrow$$

$$\begin{aligned} \Longrightarrow P(A|P) &= \frac{\pi_1 \mathcal{N}(P|\mu_1, \mathbf{\Sigma}_1)}{\pi_1 \mathcal{N}(P|\mu_1, \mathbf{\Sigma}_1) + \pi_2 \mathcal{N}(P|\mu_2, \mathbf{\Sigma}_2)} = \frac{0.1 \cdot 0.0124}{0.1 \cdot 0.0124 + 0.9 \cdot 0.0089} = 0.1343 \\ P(B|P) &= \frac{\pi_2 \mathcal{N}(P|\mu_2, \mathbf{\Sigma}_2)}{\pi_1 \mathcal{N}(P|\mu_1, \mathbf{\Sigma}_1) + \pi_2 \mathcal{N}(P|\mu_2, \mathbf{\Sigma}_2)} = 1 - P(A|P) = 0.8657 \end{aligned}$$

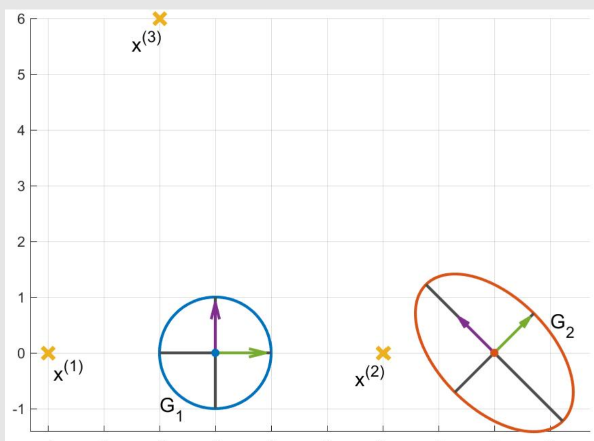


2. Given the points $\left\{ \mathbf{x}^{(1)} = \begin{pmatrix} -4 \\ 0 \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ 6 \end{pmatrix} \right\} \subset \mathbb{R}^2$, derive an Expectation-Maximisation update for a Mixture of Gaussian with initial condition

$$\pi_1 = 0.8, \mu_1 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \mathbf{\Sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\pi_2 = 0.2, \mu_2 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \mathbf{\Sigma}_2 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

(a) Draw the given points and the initial Gaussian models (ellipses). What do you expect the responsibilities to be? $r_1^{(1)} > r_2^{(1)}$ or otherwise? And what about $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(3)}$?

Solution:
$r_1^{(1)} > r_2^{(1)}$, $r_1^{(2)} < r_2^{(2)}$, $r_1^{(3)} < r_2^{(3)}$.



(b) (Expectation step) Compute the responsibilities. Are they as you expected?

Solution:

$\boxed{\mathbf{x}^{(1)}}$

$$\left.\begin{array}{l} P(\mathbf{x}^{(1)}|G_1) = 0.0018 \\ \\ P(\mathbf{x}^{(1)}|G_2) = 5 \cdot 10^{-11} \end{array}\right\} \Longrightarrow$$

$$\Longrightarrow \begin{cases} r_1^{(1)} = P(1|\mathbf{x}^{(1)}) = \dfrac{\pi_1 P(\mathbf{x}^{(1)}|G_1)}{\pi_1 P(\mathbf{x}^{(1)}|G_1) + \pi_2 P(\mathbf{x}^{(1)}|G_2)} = \dfrac{0.8 \cdot 0.0018}{0.8 \cdot 0.0018 + 0.2 \cdot 5 \cdot 10^{-11}} = 1 \\ \\ r_2^{(1)} = P(2|\mathbf{x}^{(1)}) = 1 - P(2|\mathbf{x}^{(2)}) = 0 \end{cases}$$

$\boxed{\mathbf{x}^{(2)}}$

$$\left.\begin{array}{l} P(\mathbf{x}^{(2)}|G_1) = 0.0018 \\ P(\mathbf{x}^{(2)}|G_2) = 0.0242 \end{array}\right\} \Longrightarrow \begin{array}{l} r_1^{(2)} = P(1|\mathbf{x}^{(2)}) = 0.226 \\ r_2^{(2)} = P(2|\mathbf{x}^{(2)}) = 1 - P(2|\mathbf{x}^{(2)}) = 0.774 \end{array}$$

$\boxed{\mathbf{x}^{(3)}}$

$$\left.\begin{array}{l} P(\mathbf{x}^{(3)}|G_1) = 1.4702 \cdot 10^{-9} \\ P(\mathbf{x}^{(3)}|G_2) = 5.6458 \cdot 10^{-7} \end{array}\right\} \Longrightarrow \begin{array}{l} r_1^{(3)} = P(1|\mathbf{x}^{(3)}) = 0.0103 \\ r_2^{(3)} = P(2|\mathbf{x}^{(3)}) = 1 - P(2|\mathbf{x}^{(2)}) = 0.9897 \end{array}$$

(c) (Maximisation step) Update the means, covariance matrices and mixing coefficients.

Solution:

$N_1 = r_1^{(1)} + r_1^{(2)} + r_1^{(3)} = 1.2363$

$N_2 = r_2^{(1)} + r_2^{(2)} + r_2^{(3)} = 1.7637$

Means

$$\mu_1 = \frac{1}{N_1} \sum_{n=1}^{3} r_1^{(n)} \mathbf{x}^{(n)} = \begin{pmatrix} -2.8865 \\ 0.05 \end{pmatrix}$$

$$\mu_2 = \frac{1}{N_2} \sum_{n=1}^{3} r_2^{(n)} \mathbf{x}^{(n)} = \begin{pmatrix} -0.2446 \\ 3.3669 \end{pmatrix}$$

Covariance matrices

$$\mathbf{\Sigma}_1 = \frac{1}{N_1} \sum_{n=1}^{3} r_1^{(n)} (\mathbf{x}^{(n)} - \mu_1)(\mathbf{x}^{(n)} - \mu_1)^\top = \begin{pmatrix} 5.3743 & 0.0444 \\ 0.0444 & 0.2977 \end{pmatrix}$$

$$\mathbf{\Sigma}_2 = \frac{1}{N_2} \sum_{n=1}^{3} r_2^{(n)} (\mathbf{x}^{(n)} - \mu_2)(\mathbf{x}^{(n)} - \mu_2)^\top = \begin{pmatrix} 3.9402 & -5.9103 \\ -5.9103 & 8.8654 \end{pmatrix}$$

Mixing coefficients

$$\pi_1 = \frac{N_1}{N} = \frac{1.2363}{3} = 0.4121$$

$$\pi_2 = \frac{N_2}{N} = \frac{1.7637}{3} = 0.5879$$

(d) (Convergence) Compute the log-likelihood at the current iteration.

Solution:

Notice that the updated $\mathbf{\Sigma}_2$ has an eigenvalue $= 0$, $\lambda_2 = 0$. As seen the exercise 4 from "Problems 2: Gaussian Models", when a covariance matrix has an eigenvalue $= 0$, we cannot compute the PDF, given that the covariance matrix is not invertible. To solve this issue, we set $\lambda_2 := 0.001$.

$$\ln P(X|\pi, \mu, \mathbf{\Sigma}) = \sum_{n=1}^{3} \ln \left[\sum_{k=1}^{2} \pi_k \mathcal{N}(\mathbf{x}^{(n)}|\mu_k, \mathbf{\Sigma}_k) \right] = \cdots = -4.47$$

(e) Draw the updated Gaussian models. Do the updated Gaussian models look more accurate than the initial ones? What do you think the tendency of the EM algorithm is in this case?

Solution:

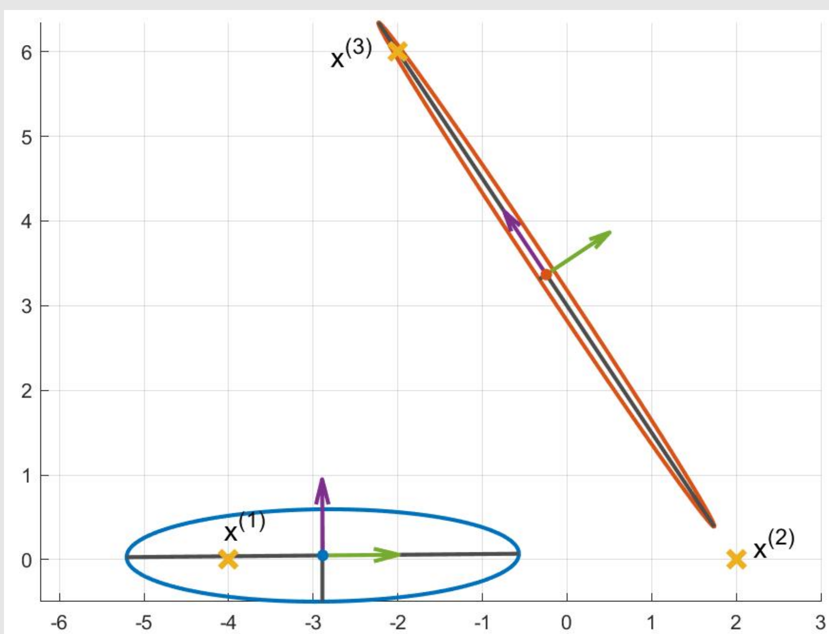
$\boxed{G_1}$

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 5.3743 & 0.0444 \\ 0.0444 & 0.2977 \end{pmatrix} \implies \begin{cases} \lambda_1 = 5.38, \mathbf{u}^{(1)} = \begin{pmatrix} 1 \\ 0.0087 \end{pmatrix} \\ \\ \lambda_2 = 0.3, \mathbf{u}^{(2)} = \begin{pmatrix} -0.0087 \\ 1 \end{pmatrix} \end{cases}$$

$\boxed{G_2}$

$$\mathbf{\Sigma}_2 = \begin{pmatrix} 3.9402 & -5.9103 \\ -5.9103 & 8.8654 \end{pmatrix} \implies \begin{cases} \lambda_1 = 12.8, \mathbf{u}^{(1)} = \begin{pmatrix} -0.56 \\ 0.83 \end{pmatrix} \\ \\ \lambda_2 = 0, \mathbf{u}^{(2)} = \begin{pmatrix} 0.83 \\ 0.56 \end{pmatrix} \end{cases}$$

Since λ_2 of $\mathbf{\Sigma}_2$ is equal to 0, we set it to a small number, p.ex. 0.01.



3. (In Jupyter Notebook) In this exercise we are going to use sklearn with a real dataset, to see how can we use Gaussian Mixture Models to extract information in a real world example. Open the ipython notebook "*P3_MixtureOfGaussian.ipynb*" and follow the instructions for this exercise.

Solution: Done in the ipython Notebook.