

**Séance 3**  
**Calcul des probabilités d'un Mélange Gaussien**

---

Tout au long de ce TP nous allons étudier les propriétés des modèles de lois Gaussiennes. Un mélange sera représenté par une structure de données contenant les informations suivantes :

|                        |   |
|------------------------|---|
| mix.type: 'gmm'        | le nom de la structure manipulée                                |
| mix.nin: 2             | la dimension  |
| mix.ncentres: 3        | le nombre de composantes du mélange                             |
| mix.covar_type: 'diag' | le type de matrice de covariance « sphérique », diagonale, full |
| mix.priors: []         | proba a priori des composantes                                  |
| mix.centres: []        | position des centres  |
| mix.covars: []         | les matrices de covariances                                     |

Nous allons utiliser certaines des fonctions de la boîte à outils NetLab téléchargeable à partir du lien ci-dessous pour réaliser le travail demandé.

<http://www.ncrg.aston.ac.uk/netlab/index.php>

**1. Génération**

A l'aide de la fonction `gmm()`, créer un mélange de 3 composantes de covariance diagonale puis modifier les centres pour les placer en (3,1) (0.5,0.5) (1,3) et modifier les covariances pour avoir sur la diagonale de chaque composante: (0.8, 0.2) (0.2,0.2) (0.2,0.8).

- Analyser le programme de la fonction `gmm` et préciser comment est représentée la matrice de covariance dans le cas sphérique, diagonal et général.

Utiliser la fonction `gmmsamp` pour générer N=1000 échantillons du mélange. Visualiser les données graphiquement à l'aide de la fonction `plot`.

**2. Calcul des probabilités**

Ecrire les différentes fonctions qui calculent les probabilités du Modèles sur les données:

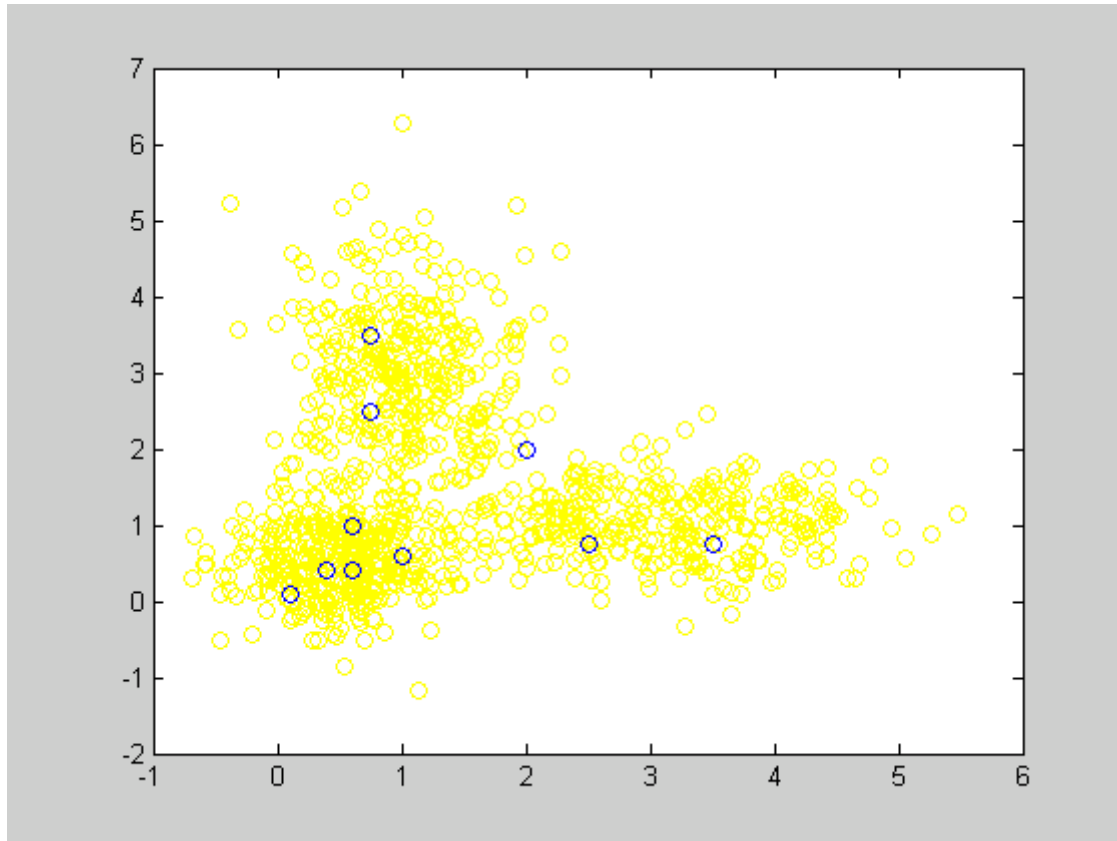
La vraisemblance de chaque composante : 
$$f(x / j) = \frac{1}{(2\pi)^{p/2} \prod_i \sigma_{ji}} \exp\left(-\sum_{i=1}^p \frac{(x_i - \mu_{ji})^2}{2\sigma_{ji}^2}\right)$$

La vraisemblance du modèle : 
$$f(x) = \sum_j f(x / j)P(j)$$

La probabilité *a posteriori* de chaque composante :  $P(j/x) = \frac{f(x/j)P(j)}{\sum_j f(x/j)P(j)}$

- Pour chaque fonction on donnera le rôle de chaque ligne de programme :
- Donner les valeurs de ses probabilités pour les 10 points suivants : ( 2.5 0.75) (0.4 0.4) (0.75 2.5) (3.5 0.75) (0.6 0.4) (0.75 3.5) (0.6 1) (1 0.6) (2 2) (0.1 0.1)
- Comment interpréter les probabilités *a posteriori* ?

Réponse :



| Echantillons | Vraisemblance des composantes |        |        | Vraisemblance du modèle | Proba a posteriori |        |        |
|--------------|-------------------------------|--------|--------|-------------------------|--------------------|--------|--------|
|              |                               |        |        |                         |                    |        |        |
| ( 2.5 0.75)  | 0.2911                        | 0.0000 | 0.0001 | 0.0971                  | 0.9997             | 0.0001 | 0.0002 |
| (0.4 0.4)    | 0.0024                        | 0.7570 | 0.0024 | 0.2539                  | 0.0031             | 0.9938 | 0.0031 |
| (0.75 2.5)   | 0.0001                        | 0.0000 | 0.2911 | 0.0971                  | 0.0002             | 0.0001 | 0.9997 |
| (3.5 0.75)   | 0.2911                        | 0.0000 | 0.0000 | 0.0970                  | 1.0000             | 0.0000 | 0.0000 |
| (0.6 0.4)    | 0.0044                        | 0.7570 | 0.0039 | 0.2551                  | 0.0058             | 0.9891 | 0.0051 |
| (0.75 3.5)   | 0.0000                        | 0.0000 | 0.2911 | 0.0970                  | 0.0000             | 0.0000 | 1.0000 |
| (0.6 1)      | 0.0109                        | 0.4154 | 0.0219 | 0.1494                  | 0.0243             | 0.9269 | 0.0488 |
| (1 0.6)      | 0.0219                        | 0.4154 | 0.0109 | 0.1494                  | 0.0488             | 0.9269 | 0.0243 |
| (2 2)        | 0.0175                        | 0.0000 | 0.0175 | 0.0117                  | 0.4999             | 0.0003 | 0.4999 |
| (0.1 0.1)    | 0.0003                        | 0.3576 | 0.0003 | 0.1194                  | 0.0008             | 0.9985 | 0.0008 |

- Les probabilités *a posteriori* permettent de connaître la composante la plus probable et donc de prendre une décision d'affectation d'un point de l'espace à l'une des composantes (sous-classes) modélisées.