

*Modèles à variables latentes discrètes*

**Devoir maison facultatif**

L'objectif de ce devoir-maison est de vous aider à préparer l'examen en révisant le cours. Les règles sont

1. **Vous ne serez pas notés et n'avez aucune obligation de m'envoyer votre composition.**
2. La correction sera mise en ligne le 6 janvier 2025 sur le site web.
3. Je ne pourrais pas vous faire de retour entre le 22/12/24 et le 6/1/25. Si vous m'envoyez quelque chose avant, merci de précisez exactement le(s) point(s) précis qui vous posent question. Un mail sans question précise ne sera pas traité.
4. Au moins un exercice de l'examen sera similaire à un des exercices de ce devoir maison. **Attention**, cela ne veut **pas** dire que le reste du cours n'est pas au programme de l'examen.

Vous n'avez donc aucun intérêt à tricher ou utiliser un chatbot pour vous assister, puisque vous n'en disposerez pas le jour de l'examen.

## Exercice 1 : Équivalence entre modèles de mélanges Gaussiens contraints, K-Means et EM Classification

Le Modèle de Mélange Gaussien (GMM) et l'algorithme de K-means visent tous deux à regrouper les données en clusters. Dans cet exercice, vous explorerez l'équivalence entre leurs objectifs et algorithmes sous certaines hypothèses.

**Rappel de cours** – L'algorithme Classification-EM (CEM), est un algorithme EM standard où l'étape E est modifiée. Dans CEM, l'étape E est remplacée par une affectation stricte (étape CE) basée sur les valeurs maximales des probabilités a posteriori  $\tau_{ik}^{(t)}$ . On attribue chaque point  $x_i$  à un seul cluster, via l'estimateur

$$\hat{z}_{ik}^{(t)} = \begin{cases} 1 & \text{if } k = \operatorname{argmax}_l \tau_{il}^{(t)}, \\ 0 & \text{otherwise} \end{cases}$$

L'étape CE utilise ensuite le critère de la vraisemblance complète  $Q(\theta | \theta^{(t)}) = \log p_{\theta}(X, \hat{Z}^{(t)})$ . Cela contraste avec l'EM standard où chaque  $x_i$  contribue à chaque cluster à hauteur des probabilités  $\tau_{ik}$ .

- (a) **Log-vraisemblance d'un mélange Gaussien:** Supposons que les données soient  $\{x_i\}_{i=1}^N$  et que les paramètres du modèle soient  $\theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K$ . On note  $z_i \in \{0, 1\}^K$  les variables latentes. Écrivez les fonctions de log-vraisemblance et de log-vraisemblance complète pour un Modèle de Mélange Gaussien avec  $K$  composantes.

**Réponse :** La log-vraisemblance est donnée par :

$$\log p_{\theta}(X) = \sum_{i=1}^N \log \left( \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \right),$$

où  $\mathcal{N}(x_i | \mu_k, \Sigma_k)$  est la densité gaussienne multivariée.

- (b) **Objectif de K-Means :** Écrivez la fonction objectif minimisée par l'algorithme de K-means en fonctions des  $z_{ik}$ .

**Réponse :** La fonction objectif de K-means est :

$$J = \sum_{i=1}^N \sum_{k=1}^K z_{ik} \|x_i - \mu_k\|^2,$$

- (c) **GMM contraint :** Considérez le GMM auquel on impose les contraintes suivantes sur les paramètres

- Les matrices de covariance sont isotropes et identiques :  $\Sigma_k = \sigma^2 I$ .
- Les poids des mélanges sont uniformes :  $\pi_k = \frac{1}{K}$ .

Montrez que le critères K-means et la log-vraisemblances **complète** sont proportionnels à une constante négative près (dépendant de  $\sigma^2$ ).

**Réponse :** Simplifier la densité Gaussienne quand la covariance diagonale.

- (d) **Équivalence entre algorithme du K-Means et CEM :**

- (i) Écrivez la règle de mise à jour pour l'étape CE dans l'algorithme CEM dans le cadre du GMM contraint. Expliquez comment cette étape est équivalente à l'étape d'affectation des clusters dans K-means.

**Réponse :** La probabilité a posteriori du modèle GMM contraint s'écrit

$$\tau_{il} \propto \frac{1}{K} \mathcal{N}(x_i | \mu_k, \sigma^2 I) \propto -\frac{1}{2\sigma^2} \|x_i - \mu_l\|^2$$

La règle de mise à jour de l'étape E dans CEM est :

$$z_{ik} = \begin{cases} 1 & \text{si } k = \arg \max_l \tau_{il}, \\ 0 & \text{sinon.} \end{cases}$$

Cela correspond à affecter chaque point au cluster le plus proche en termes de probabilité à posteriori ce qui est analogue à l'étape d'affectation dans K-means.

- (ii) Écrivez la règle de mise à jour pour l'étape M dans l'algorithme CEM dans le cadre du GMM contraint. Montrez que cette étape correspond à la mise à jour des centroïdes des clusters dans K-means.

**Réponse :** La règle de mise à jour de l'étape M est :

$$\mu_k = \frac{\sum_{i=1}^N z_{ik} x_i}{\sum_{i=1}^N z_{ik}}.$$

Cela calcule le centroïde du cluster  $k$ , tout comme K-means.

(iii) Qu'en déduisez-vous ?

**Réponse :** En combinant les étapes CE et M, K-means réalise une version de CEM où les  $\tau_{ik}$  sont binaires, et les covariances sont fixes et isotropes ( $\Sigma_k = \sigma^2 I$ ).

(e) **Bilan:** Selon vous quels sont les avantages et inconvénients d'une affectation stricte (comme dans K-means et CEM) par rapport aux affectations souples (comme dans EM standard).

**Réponse :** Les affectations strictes sont plus rapides car beaucoup de termes peuvent être "oubliés" dans fonction objectif ( $z_{il} = 0$ ). Cependant, elles peuvent conduire à des solutions sous-optimales lorsque les clusters se chevauchent, car elles négligent l'incertitude. Les affectations souples, en revanche, tiennent compte de cette incertitude, ce qui peut améliorer la robustesse et les performances globales.

## Exercice 2 : Algorithme EM pour le Mélange de Distributions Gamma

Dans cet exercice, vous dériverez et implémenterez l'algorithme EM pour le modèle de mélange de distribution Gamma. La loi Gamma est une loi continue à 2 paramètres défini par la densité de probabilité suivante :

$$\text{Gamma}(x | \alpha_k, \beta_k) = \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} x^{\alpha_k-1} e^{-\beta_k x} \mathbb{1}_{x>0},$$

où  $\alpha_k$  est le paramètre de forme et  $\beta_k$  le paramètre d'échelle.

(a) **Configuration du Modèle :** Considérez un ensemble de données positives  $\{x_i\}_{i=1}^N$  tirées d'un mélange de  $K$  distributions Gamma. Écrivez la densité de probabilité pour le modèle de mélange et définissez les paramètres  $\gamma_k$  de chaque composante.

**Réponse :** La fonction de densité de probabilité pour un mélange de distributions Gamma est donnée par :

$$p(X | \theta) = \prod_{i=1}^n \sum_{k=1}^K \pi_k \text{Gamma}(x_i | \alpha_k, \beta_k),$$

où :

- $\pi_k$  est le poids du mélange pour la  $k$ -ième composante, avec  $\sum_{k=1}^K \pi_k = 1$ .
- $\text{Gamma}(x | \alpha_k, \beta_k) = \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} x^{\alpha_k-1} e^{-\beta_k x}$ , où  $\alpha_k$  est le paramètre de forme et  $\beta_k$  le paramètre d'échelle.

Les paramètres du modèle sont donc  $\theta = \{\pi_k, \alpha_k, \beta_k\}_{k=1}^K$ .

(b) **Étape E :** Dérivez l'expression pour les probabilités a posteriori  $\tau_{ik}$  indiquant qu'un point de données  $x_i$  appartient à la composante  $k$ .

**Réponse :** Les probabilités a posteriori  $\tau_{ik}$  sont données par :

$$\tau_{ik} = \frac{\pi_k \text{Gamma}(x_i | \alpha_k, \beta_k)}{\sum_{j=1}^K \pi_j \text{Gamma}(x_i | \alpha_j, \beta_j)}.$$

Cette expression repose sur la règle de Bayes. Aucune approximation supplémentaire n'est nécessaire à ce stade.

- (c) **Étape M :** Dérivez les équations de mise à jour pour les paramètres des distributions Gamma ( $\alpha_k$  et  $\beta_k$ ) et les poids du mélange ( $\pi_k$ ). Expliquez comment ces mises à jour sont calculées de manière itérative.

**Réponse :** Les mises à jour sont les suivantes :

- Mise à jour des poids du mélange :

$$\pi_k = \frac{\sum_{i=1}^N \tau_{ik}}{N}.$$

- La mise à jour du paramètre de forme  $\alpha_k$  est obtenue en résolvant numériquement

$$\log(\alpha_k) - \psi(\alpha_k) = \log\left(\frac{1}{\sum_{i=1}^N \tau_{ik}} \sum_{i=1}^N \tau_{ik} \log(x_i)\right) - \frac{\sum_{i=1}^N \tau_{ik} \log(x_i)}{\sum_{i=1}^N \tau_{ik}},$$

où  $\psi = \Gamma'$  est la fonction digamma.

- Mise à jour du paramètre d'échelle  $\beta_k$  :

$$\beta_k = \frac{\alpha_k}{\frac{\sum_{i=1}^N \tau_{ik} x_i}{\sum_{i=1}^N \tau_{ik}}}.$$

Ces mises à jour sont calculées de manière itérative jusqu'à convergence.

- (d) **Implémentation de l'Algorithme :** Décrivez les étapes de l'algorithme EM pour le mélange de distributions Gamma sous forme de pseudocode.

**Réponse :** Voici les étapes principales de l'algorithme EM :

- Initialiser les paramètres  $\theta = \{\pi_k, \alpha_k, \beta_k\}$  aléatoirement ou via une méthode heuristique.
- Répéter jusqu'à convergence :
  - Étape E :** Calculer les responsabilités  $\tau_{ik}$  pour chaque  $i, k$  :

$$\tau_{ik} = \frac{\pi_k \text{Gamma}(x_i | \alpha_k, \beta_k)}{\sum_{j=1}^K \pi_j \text{Gamma}(x_i | \alpha_j, \beta_j)}.$$

- Étape M :** Mettre à jour les paramètres :

- $\pi_k = \frac{\sum_{i=1}^N \tau_{ik}}{N},$
- Résoudre numériquement  $\alpha_k$  à l'aide de l'équation de l'étape précédente,
- $\beta_k = \frac{\alpha_k}{\frac{\sum_{i=1}^N \tau_{ik} x_i}{\sum_{i=1}^N \tau_{ik}}}.$

- (e) **Considérations Pratiques :** Discutez des défis liés à l'implémentation de l'algorithme EM pour ce modèle, en particulier en ce qui concerne l'initialisation, les critères de convergence et la stabilité numérique.

**Réponse :** Les défis incluent :

- **Initialisation :** Une mauvaise initialisation peut entraîner la convergence vers un mauvais optimum local. L'utilisation de techniques comme K-means pour initialiser les paramètres peut aider.
- **Critères de Convergence :** Le critère de convergence se base sur la variation de la log-vraisemblance (garantie de converger par EM), ou bien des paramètres (non-garanti en général).
- **Stabilité Numérique :** Tout les quantités doivent être calculées en log, et les  $\log \tau_{ik}$  doivent être normalisés en utilisant la version numériquement stable de log-sum-exp. De plus, le calcul de la fonction digamma et l'optimisation de  $\alpha_k$  peuvent être sensibles aux valeurs des données, nécessitant une mise à l'échelle ou une régularisation.

### Exercice 3: VEM for binary Stochastic Block Model (SBM)

The stochastic block model (SBM) assumes a graph with  $n$  nodes divided into  $K$  groups. The presence or absence of an edge between any two nodes is modeled as a Bernoulli random variable, where the probability of an edge depends on the groups to which the nodes belong.

Let:

- Let  $q(Z) = \prod_{i=1}^n \prod_{k=1}^K \tau_{ik}^{z_{ik}}$  denote a probability distribution over all latent variable,
- $\gamma_{kl}$  denote the probability of an edge between nodes in groups  $k$  and  $l$ ,
- $\pi_k$  denote the prior group proportion for group  $k$ , where  $\sum_{k=1}^K \pi_k = 1$ .

#### Questions

- (a) How do we call a fully factorized distribution like  $q$  ?
- (b) Write the Evidence lower bound (ELBO) for the binary SBM with the variational distribution  $q$  as above.

**Réponse :**

$$\mathcal{L}(\tau, \gamma, \pi) = \mathbb{E}_q[\log P(\mathbf{A}, \mathbf{Z}, \pi | \gamma)] - \mathbb{E}_q[\log q(\mathbf{Z})].$$

The ELBO can be expanded as:

$$\mathcal{L}(\tau, \gamma, \pi) = \mathbb{E}_q[\log P(\mathbf{A} | \mathbf{Z}, \gamma)] + \mathbb{E}_q[\log P(\mathbf{Z} | \pi)] + \log P(\pi) - \mathbb{E}_q[\log q(\mathbf{Z})].$$

- (c) **Variational E-step (VE-step)** What optimization has to be solved in the VE-step ? Does it have an exact solution ? If not, write down an iterative algorithm to solve it and explicit its stopping criterion(s).

**Réponse :** The posterior membership  $\tau_{ik}$  is updated via fixed point iterations:

$$\tau_{ik} \propto \exp \left( \log \pi_k + \sum_{j \neq i} \sum_{l=1}^K \tau_{jl} \left[ A_{ij} \log \gamma_{kl} + (1 - A_{ij}) \log (1 - \gamma_{kl}) \right] \right),$$

- (d) **M-step updates** Write down the updates for the parameters  $\gamma_{kl}$  and  $\pi_k$  using the current posterior membership probabilities.

**Réponse :** The connectivity matrix  $\gamma_{kl}$  is updated as:

$$\gamma_{kl} = \frac{\sum_{i=1}^n \sum_{j=1}^n A_{ij} \tau_{ik} \tau_{jl}}{\sum_{i=1}^n \sum_{j=1}^n \tau_{ik} \tau_{jl}}.$$

The group proportion  $\pi_k$  is updated as:

$$\pi_k = \frac{\sum_{i=1}^n \tau_{ik}}{n}.$$

- (e) How would you cluster the nodes of the network after the VEM ?