DM PGM #2

ATTAIKI SOUHAIB

Exercice 1: Conditional independance and factorisations

1) Pour $p \in \mathcal{L}(G)$, on a $p(x,y,z,t) = p(x)\,p(y)\,p(z|x,y)\,p(t|z)$

ⓐ $X \perp\!\!\!\perp Y | T$, $\forall p \in \mathcal{L}(G)$ est une expression fausse. Contre exemple :

Soient X, Y deux V.A I.I.D qui suivent une loi de Bernoulli, soit Z la variable binaire tq $Z = 0$ ssi $X = 0$ et $Y = 1$ et $Z = 1$ sinon. et $T = Z$

donc on a clairement que X et Y sont dependant à T.

2) ⓐ la proposition est vraie (P.S : les calculs sont simplifiés pour optimiser l'espace). on a

$p(x,y) = \sum_z p(x,y,z) = \sum_z p(x,y|z)\,p(z) = \sum_z p(x|z)\,p(y|z)\,p(z)$ ①. et on a

$p(x,y) = p(x)\,p(y) = \left(\sum_z p(x|z)\,p(z)\right)\left(\sum_z p(y|z)\,p(z)\right)$ ②. or Z est binaire, donc $z=0$ ou $z=1$,

on pose $p(z=1) = \pi$ et on note $p(a|z=i) = p(a|i)$, et on ① = ②, donc :

$p(x|0)\,p(y|0)(1-\pi) + p(x|1)\,p(y|1)\,\pi = \left(p(x|0)(1-\pi) + p(x|1)\,\pi\right)\left(p(y|0)(1-\pi) + p(y|1)\,\pi\right)$

$\Leftrightarrow p(x|0)\,p(y|0)(1-\pi)\pi + p(x|1)\,p(y|1)\,\pi(1-\pi) - \pi(1-\pi)\left(p(x|1)\,p(y|0) + p(x|0)\,p(y|1)\right) = 0$

$\Leftrightarrow p(x|0)\left[p(y|0) - p(y|1)\right] - p(x|1)\left[p(y|0) - p(y|1)\right] = 0$

$\Leftrightarrow \left[p(x|0) - p(x|1)\right]\left[p(y|0) - p(y|1)\right] = 0 \Rightarrow p(x|0) = p(x|1)$ ou $p(y|0) = p(y|1)$

si $p(x|0) = p(x|1)$, alors $p(x) = p(x,0) + p(x,1) = p(x|0)(1-\pi) + \pi\,p(x|1) = p(x|1) = p(x|0)$

donc $p(x) = p(x|z)$ d'où $X \perp\!\!\!\perp Z$, de même pour le deuxième cas. d'où le resultat.

ⓑ La proposition est fausse en général. Contre exemple.

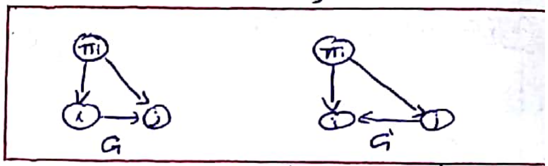
Soient X, Y deux V.A tel que $X \perp\!\!\!\perp Y$. on pose $Z = (X, Y)$.

on a $X \perp\!\!\!\perp Y | Z$ et $X \perp\!\!\!\perp Y$, mais clairement, on a pas $X \perp\!\!\!\perp Z$

ou $Y \perp\!\!\!\perp Z$ car en mesurant Z, on a mesuré X et Y.

Exercice 2 : Distributions factorizing in a graph.

1) soit $p \in \mathcal{L}(G)$. on a $p(x) = \prod_{k=1}^{n} p(x_k \mid x_{\pi_k})$. on note par π'_m les parents de m dans $\mathcal{L}(G')$. on a $\pi_j = \pi_i \cup \{i\}$ et $\pi'_j = \pi_i$ et $\pi'_i = \pi_i \cup \{j\}$



et on a $p(x) = p(x_i \mid \pi_{x_i})\, p(x_j \mid \pi_{x_j}) \prod_{\substack{k=1 \\ k\neq i,j}}^{n} p(x_k \mid \cancel{x}_{\pi_k})$

$= p(x_i \mid \pi_{x_i})\, p(x_j \mid \pi_{x_j}) \prod_{\substack{k=1 \\ k\neq i,j}}^{n} p(x_k \mid x_{\pi'_k})$. donc il suffit de montrer

que $p(x_i \mid \pi_{x_i})\, p(x_j \mid \pi_{x_j}) = p(x_i \mid \pi'_{x_i})\, p(x_j \mid \pi'_{x_j})$. (Parce que les deux graphes ne diffèrent que dans i et j)

On a
$$p(x_i \mid \pi_{x_i})\, p(x_j \mid \pi_{x_j}) = \frac{p(x_i, \pi_{x_i})}{p(\pi_{x_i})} \cdot \frac{p(x_j, \pi_{x_j})}{p(\pi_{x_j})} = \frac{p(x_i, \pi_{x_i})}{p(\pi_{x_i})} \cdot \frac{p(x_j, \pi_i, x_i)}{p(\pi_i, x_i)}$$

$$= \frac{p(x_j, \pi_{x_i})}{p(\pi_{x_i})} \cdot \frac{p(x_i, \pi_i, x_j)}{p(x_j, \pi_{x_i})} = p(x_j \mid \pi'_j) \cdot p(x_i \mid \pi'_i)$$

d'où $p \in \mathcal{L}(G')$. par symmetrie on montre que $p \in \mathcal{L}(G') \Rightarrow p \in \mathcal{L}(G)$

d'où $\mathcal{L}(G) = \mathcal{L}(G')$.

2) $\Rightarrow$) soit $p \in \mathcal{L}(G)$, donc $p(x) = \prod_{i \in V} p(x_i \mid \pi_{x_i})$. puisque un arbre ne contient pas de V-structure, donc $\#(\pi_{x_i}) \leq 1$, donc en posant $p(x_i \mid \pi_{x_i}) = \psi_i(x_i, \pi_{x_i})$, on a $p \in \mathcal{L}(G')$

$\Leftarrow$) on montre la réciproque récurrence sur n, le nombre de nœud. pour $n=1$, le résultat est trivial ($p(x_i) = \psi(x_i)$). soit $n>1$, soit G' un arbre non orienté tel que $|V| = n$ et soit G son arbre orienté correspondant, on numérote les noeuds de telle façon que v_1 soit la racine dans G, et v_n soit une feuille avec un seul parent v_{n-1}. on suppose le résultat vrai pour $n-1$.

soit $p \in \mathcal{L}(G')$, donc $p(x) = \frac{1}{Z} \prod_{(i,j)\in E} \psi_{i,j}(x_i, x_j)$. soit $\tilde{p}$ le marginal sur $(1, \dots, n-1)$.

$$\tilde{p} = \frac{1}{Z} \tilde{\psi}_{n-1}(x_{n-1}) \prod_{\substack{(i,j)\in E \\ (i,j)\neq(n-1,n)}} \psi_{i,j}(x_i, x_j) \text{ tel que } \tilde{\psi}_{n-1}(x_{n-1}) = \sum_{x_n} \psi(x_{n-1}, x_n).$$

non orienté

soit $\tilde{G}' = (V \setminus \{n\}, E \setminus \{(n-1, n)\})$, $\tilde{G}'$ est arbre tel que $|\tilde{V}| = n-1$, en plus $\tilde{p} \in \mathcal{L}(\tilde{G}')$, donc, d'après l'hypothèse de récurrence, $\tilde{p} \in \mathcal{L}(\tilde{G})$ ($\tilde{G}$ est l'arbre orienté correspondant), donc

$\tilde{p}(x) = \prod_{i=1}^{n-1} p(x_i \mid \pi_{x_i})$. On a dans G, $\pi_{x_n} = \{n-1\}$. on pose

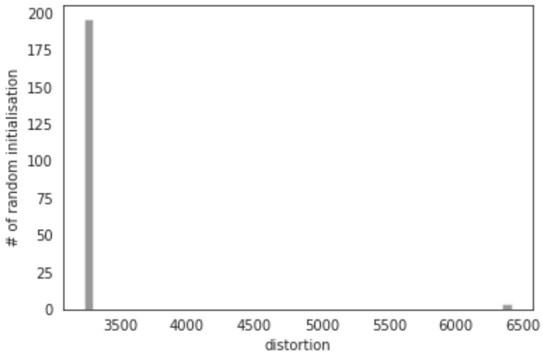
$$f(x_{n-1}, x_n) = \begin{cases} \dfrac{\psi_{n-1,n}(x_{n-1}, x_n)}{\tilde{\psi}_{n-1}} & \text{si } \tilde{\psi}_{n-1} \neq 0 \\ \dfrac{1}{C_{x_n}} & \text{sinon} \end{cases}$$

avec C_{x_n}, le nombre de valeur que peut prendre x_n.

on a $\sum_{x_n} f(x_{n-1}, x_n) = 1$, en plus $p = f(x_{n-1}, x_n)\, \tilde{p}$ car si $\tilde{\psi}_{n-1} = 0 \Rightarrow \psi_{n-1,n}(x_{n-1}, x_n) = 0, \forall x_n$

on a $p \in \mathcal{L}(G)$. donc le resultat est vrai pour n. C.Q.F.D

Exercice 3.a



On peut voir que différents résultats sont obtenus en utilisant différentes initialisations. Cependant, la plupart des initialisations donnent pratiquement le même résultat (centres et distorsion), et un très faible pourcentage de résultats est très loin du minimum.

Exercice 3.b

D'après ce qu'on vu dans le cours, les valeurs de π_t et μ_t ne dépendent pas de la modélisation et donc leurs formules ne changent pas. Pour trouver la valeur de Σ dans le cas isotropique, on maximise l'expression du vraisemblance obtenue à l'E-step en prenant en compte que $\Sigma_j = \alpha_j \mathbb{I}$. La partie de la vraisemblance qui dépend de Σ est :

$$L = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{k} \tau_i^j \left[d \log(\alpha_j) + \frac{1}{\alpha_j} \|x_i - \mu_{j,t}\|^2 \right]$$

En dérivant L par rapport à α_j, on trouve :

$$\frac{\partial L}{\partial \alpha_j} = 0 \Leftrightarrow \alpha_j^{(t+1)} = \frac{1}{d} \frac{\sum_{i=1}^{n} \tau_i^j \|x_i - \mu_{j,t+1}\|^2}{\sum_{i=1}^{n} \tau_i^j}$$

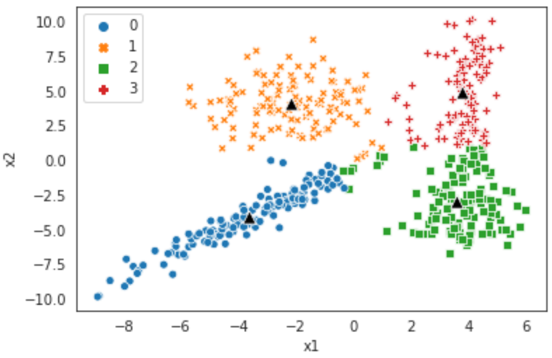
Exercice 3.d

Log-likelihood

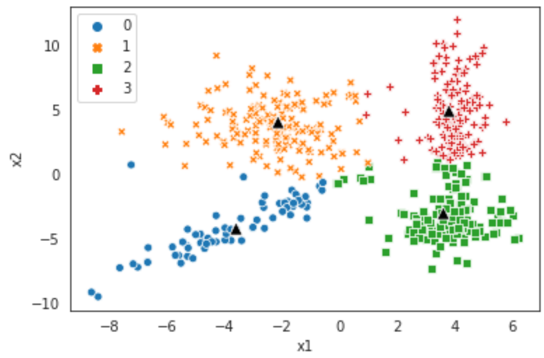
Covariance type	Training set	Test set
Isotropic	-2645.52	-2694.09
General	-2327.71	-2408.97

- Il est clair que les deux modèles donnent des résultats satisfaisants. Cependant, la deuxième modélisation (general gaussian mixture) produit des cluster plus fins et une log-likelihood plus faible que la première modélisation. Ce résultat est cohérent parce que la deuxième modélisation met moins d'hypothèses sur la matrice de covariance et, par conséquent, elle a une plus grande puissance de modélisation.
- Comme prévu, le log-likelihood de l'ensemble de données d'entraînement est plus petit que le log-likelihood de l'ensemble de données de test.

Kmeans

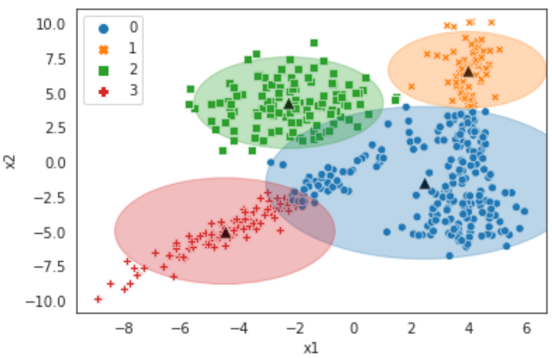


“Train set”

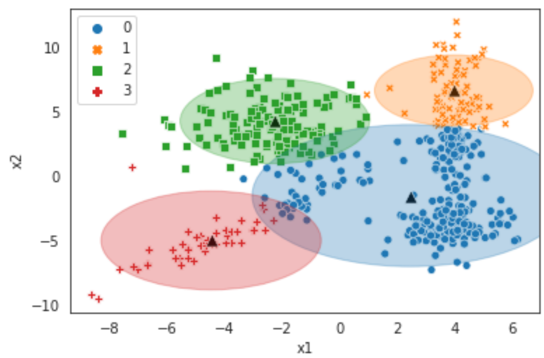


“Test set”

EM - Isotropic

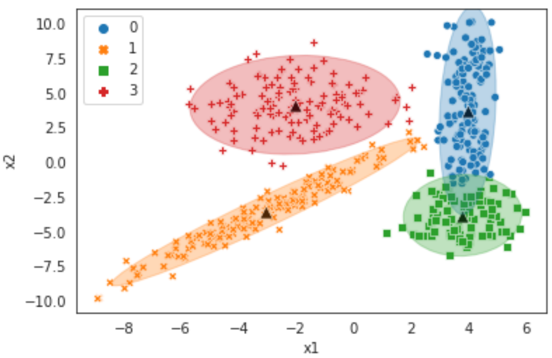


“Train set”

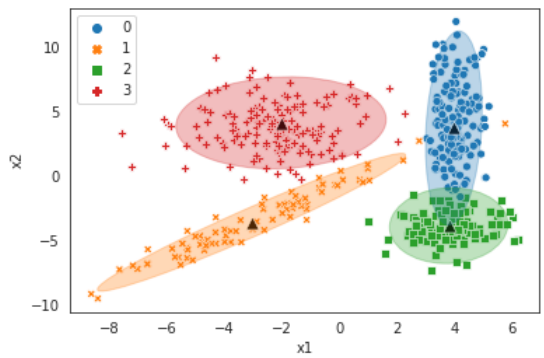


“Test set”

EM - General



“Train set”



“Test set”