Exercice 1.1.

- Si $X \perp Y | T$ alors X et Y sont d-séparables par tous T.

Montrons que cela n'est pas le cas avec l'algorithme de Bayes ball :

- Imaginons une balle partant de X. Puisque Z n'est pas observé mais que (X, Z, Y) est une v-structure, alors la balle passe et arrive en Y ;

- Ainsi X et Y ne sont pas d-séparables par T donc ne sont pas indépendants conditionnellement à T

Exercice 1.2

ⓐ D'une part : $p(x, y) = p(x)p(y)$ car $X \perp\!\!\!\perp Y$.

D'autre part, $p(x y) = p(xy, z=1) + p(xy, z=0)$ car $Z \sim Ber(\pi)$

$$= p(x/z=1)\, p(y/z=1)\, \underbrace{p(z=1)}_{\pi} + p(x/z=0)\, p(y/z=0)\, \underbrace{p(z=1)}_{1-\pi} \quad \text{car } X \perp\!\!\!\perp Y | T$$

$$= p(x)p(y) \cdot \left[\frac{p(z=1/x)\, p(z=1/y)}{\pi} + \frac{p(z=0|x)\, p(z=0|y)}{1-\pi} \right]$$

(règle de Bayes si $\pi \in]0,1[$.

- En notant $p_x = p(z=1/x)$ et $p_y = p(z=1/y)$:

$$1 = \frac{p_x p_y}{\pi} + \frac{(1-p_x)(1-p_y)}{1-\pi}$$

$$= p_x \cdot \frac{p_y - \pi}{\pi(1-\pi)} + \frac{1-p_y}{1-\pi}$$

La fonction $f : t \mapsto t \cdot \frac{p_y - \pi}{\pi(1-\pi)} + \frac{1-p_y}{1-\pi}$ est affine donc si $\frac{p_y-\pi}{\pi(1-\pi)} \neq 0$, alors $\exists! \, t, f(t) = 1$.

Or $f(\pi) = 1$ donc $p_x = p(z=1/x) = \pi = p(z=1)$
et $1 - p_x = p(z=0/x) = 1 - \pi = p(z=0)$ } donc $X \perp\!\!\!\perp Z$

Si $\frac{p_y - \pi}{\pi(1-\pi)} = 0$, alors $p_y = \pi$ et de même $Y \perp\!\!\!\perp Z$.

- Ainsi, si $\pi \in]0,1[$, $X \perp\!\!\!\perp Z$ ou $Y \perp\!\!\!\perp Z$.
- Si $\pi = 0$, ou $\pi = 1$, alors Z est déterministe donc $X \perp\!\!\!\perp Z$ et $X \perp\!\!\!\perp Z$.

ⓑ Si $X \not\!\perp\!\!\!\perp Y$. Prenons $Z = (X, Y)$. On a bien $X \perp\!\!\!\perp Y | Z$

mais pourtant $X \not\!\perp\!\!\!\perp Z$ et $Y \not\!\perp\!\!\!\perp Z$ en général (par exemple $X \sim Unif(0,1)$, $Y \sim Unif(1,2)$)

La propriété est fausse dans un cas général.

Exercice 2

① Soit $p_1 \in \mathcal{L}(G)$ et $p_2 \in \mathcal{L}(G')$.

Alors
$$p_1(x) = \left[\prod_{\substack{k\neq j\\ k\neq i}} p(x_k \mid x_{\pi_k})\right] \times p(x_i \mid x_{\pi_i})\, p(x_j \mid x_{\pi_i \cup \{i\}})$$

$$p_2(x) = \left[\prod_{\substack{k\neq j\\ k\neq i}} p(x_k / x_{\pi_k})\right] \times p(x_j \mid x_{\pi_j})\, p(x_i \mid x_{\pi_j \cup \{j\}})$$

Or
$$p(x_i \mid x_{\pi_i})\, p(x_j \mid x_{\pi_i \cup \{i\}}) = p(x_i \mid x_{\pi_i}) \cdot \frac{p(x_j\, x_{\pi_i \cup \{i\}})}{p(x_{\pi_i}\, x_i)} = \frac{p(x_j\, x_{\pi_i \cup \{i\}})}{p(x_{\pi_i})}$$

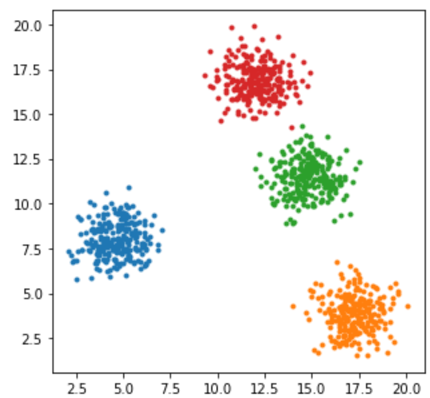
et
$$p(x_j \mid x_{\pi_j})\, p(x_i \mid x_{\pi_j \cup \{j\}}) = p(x_j \mid x_{\pi_j}) \cdot \frac{p(x_i\, x_{\pi_j \cup \{j\}})}{p(x_{\pi_j}\, x_j)} = \frac{p(x_j \overbrace{x_{\pi_j}}^{= x_{\pi_i}} x_i)}{\underbrace{p(x_{\pi_j})}_{= x_{\pi_i}}}$$

D'où $p_1(x) = p_2(x)$ et finalement $\underline{\mathcal{L}(G) = \mathcal{L}(G')}$

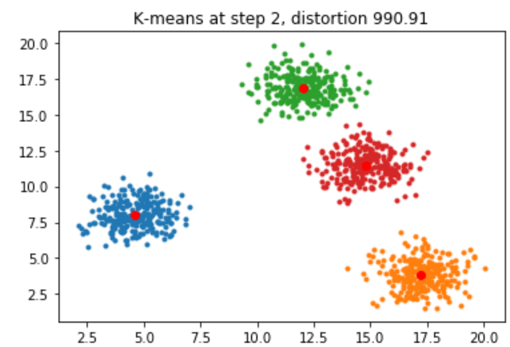
② Puisque chaque noeud n'a qu'un seul parent dans un arbre, la version "moralisée" d'un arbre est le même arbre.
Ainsi, si G est un arbre dirigé, et G' son arbre correspondant non dirigé, alors $\underline{\mathcal{L}(G) = \mathcal{L}(G')}$

Exercice 4.a.

Pour tester les performances de l'algorithme, j'ai construit des données synthétiques labellisées :



Données labellisées générées avec des gaussiennes



Résultat du k-means

La convergence dépend cependant de l'initialisation. J'ai ensuite pu comparer la distorsion des résultats avec l'algorithme classique du K-Means et celui de K-Means++.
J'ai obtenu le tableau suivant en simulant l'algorithme 1000 fois :

Algorithme	Distorsion		
	Moyenne	**Écart-type**	**Max**
Kmeans	1665.29	1796.92	23645.19
Kmeans++	1162.58	590.24	3203.85

L'algorithme Kmeans++ fonctionne non seulement mieux en moyenne, mais est nettement plus robuste (la distorsion maximale est nettement plus faible).

Exercice 4.b

Sous l'hypothèse d'une variance proportionnelle à l'identité, la vraisemblance d'une distribution du mélange de gaussienne peut s'écrire :

$$p(\boldsymbol{x}, \boldsymbol{z} | \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_i^k \ln \pi_k + \sum_{n=1}^{N} \sum_{k=1}^{K} z_i^k \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu_k}, \boldsymbol{\sigma_k I})$$

Avec $\pi_k = p(z_k = 1)$, $\boldsymbol{z}$ étant le vecteur indicateur aléatoire qui attribue un élément à un cluster.
E-Step : l'espérance de la vraisemblance en Z conditionnellement à Xest :

$$\sum_{n=1}^{N} \sum_{k=1}^{K} \tau_n^k \ln \pi_k + \sum_{n=1}^{N} \sum_{k=1}^{K} \tau_n^k \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu_k}, \boldsymbol{\sigma_k I})$$

avec $\tau_i^k = p_{\sigma\mu\pi}(z_i = k | x_i) = \frac{\pi_k \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \sigma_k)}{\sum_j \pi_j \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_j, \sigma_j)}$. On calcule donc chaque τ_i^k

M-Step :

Le problème de maximisation du logarithme de la vraisemblance étant concave sur un domaine convexe, le maximiseur annule la dérivant. La vraisemblance en deux termes. D'un côté, on dérive par rapport à $\boldsymbol{\mu}_k$:

$$0 = \frac{1}{\sigma_k} \sum_{n=1}^{N} \tau_n^k \, (\boldsymbol{x}_n - \boldsymbol{\mu}_k)$$

Soit :

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \tau_n^k \, \boldsymbol{x}_n$$

En notant :

$$N_k = \sum_{n=1}^{N} \tau_n^k$$

De la même manière :

$$\sigma_k^2 \boldsymbol{I} = \frac{1}{dN_k} \sum_{n=1}^{N} \tau_n^k \, (x_n - \boldsymbol{\mu}_k)(x_n - \boldsymbol{\mu}_k)^T$$

L'autre terme dépend de π_k et en le dérivant par rapport à π_k :

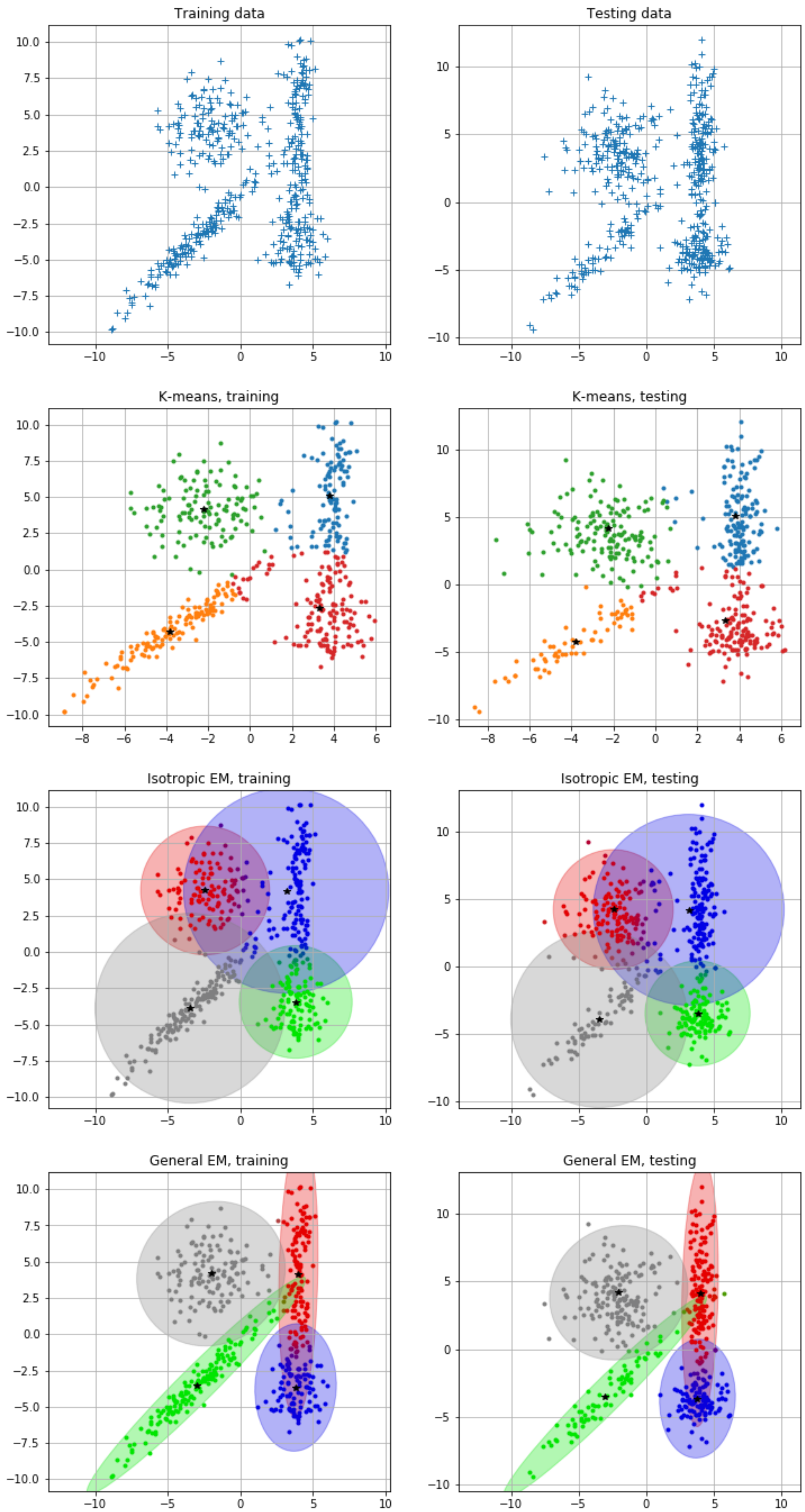
$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} \tau_n^k$$

Exercice 4.d

Set	Isotropic	General
Train	-19740.56	-69820.00
Test	-19518.09	-66032.70

Complete log-likelihood

Les données ne sont pas réparties selon des cercles dans le plan. C'est pourquoi, elles s'expliquent difficilement avec des covariances diagonales. Cela se voit visuellement, car l'ellipse de covariance du cluster en haut à gauche est très large. Ainsi, des points situés loin de lui sont attribués par erreur. Cela se voit également à travers ce tableau


Training data
Testing data
K-means, training
K-means, testing
Isotropic EM, training
Isotropic EM, testing
General EM, training
General EM, testing