Homework 2

Reda Ouhamma, Mossaab Wiklandour

November 2018

1 Exercice 1

1.1 a)

For the graph G, $p \in L(G)$ if and only if:

$$p(x_x, x_y, x_z, x_t) = p(x_x) * p(x_y) * p(x_z|x_y, x_x) * p(x_t|x_z)$$

1.1.1 b)

The statement $X \perp\!\!\!\perp Y|T$ is false, one counter example is the following:
X, Y two independant variables, we consider $Z = X + Y$ and $T = Z$ The distribution that we get does belong to $L(G)$ but $X \not\perp\!\!\!\perp Y|T$

1.2 a)

Let's consider that Z is a binary variable such that: $Z \in \{0, 1\}$, $\quad p(Z = 0) = q$ And such that : $X \perp\!\!\!\perp Y|Z \quad and \quad X \perp\!\!\!\perp Y$

$$\begin{aligned} p(x_x, x_y) =& q.p(x_x, x_y|0) + (1 - q).p(x_x, x_y|1) \\ =& q.p(x_x|0).p(x_y|0) - (1 - q).p(x_x|1).p(x_y|1) \end{aligned}$$

And we also have :

$$\begin{aligned} p(x_x, x_y) =& p(x_x) * p(x_y) \\ =& (q.p(x_x|0) + (1 - q).p(x_x|1)) \times (q.p(x_y|0) + (1 - q).p(x_y|1)) \end{aligned}$$

By extracting the equations we get : $(p(x_x|0)-p(x_x|1))(p(x_y|0)-p(x_y|1)) = 0$ Which means that either $X \perp\!\!\!\perp Z \quad or \quad Y \perp\!\!\!\perp Z$ and the statement is True !

2 Exercice 2

2.1

Let $p \in L(G)$,

$$p(x) = \prod_1^n p(x_k|\pi_{x_k}) = p(x_i|\pi_{x_i}) * p(x_j|\pi_{x_j}) * \prod_{1,k\neq i,k\neq j}^n p(x_k|\pi_{x_k})$$
$$p(x_j|x_i, \pi_{x_i}) * p(x_i|\pi_{x_i}) = \frac{p(x_i, x_j, \pi_{x_i})}{p(\pi_{x_i})} \quad = p(x_i|(\pi_{x_i} \cup x_j)) * p(x_j|(\pi_{x_j} \backslash i))$$

From the latter we find the same expression for p as for $p' \in L(G^{'})$,
Since the operations we did are equivalences, we get that $L(G) = L(G^{'})$

2.2

Let $p^{'} \in L(G^{'})$, and let's arrange the vertexes in G in a topological order.
The only cliques in the resulting graph are the singletons $\{i\}, i \in \{1, ..., n\}$ and $\{i, i-1\}, i \in \{2, ..., n\}$ so the probability distributions are given by:

$$p(x) = \frac{1}{Z} * \prod_1^n \psi_{0,k}(x_k) * \prod_2^n \psi_{1,k}(x_k, \pi_{x_k})$$

$$s.t: \quad Z = \sum_x \psi_{0,1}(x_1) * \prod_2^n \psi_{0,k}(x_k) * \psi_{1,k}(x_k, \pi_{x_k})$$

$$= (\sum_{x_n} \psi_{0,n}(x_n) * \psi_{1,n}(x_n, \pi_{x_n})) * \sum_{\bar{x_n}} \psi_{0,1}(x_1) * \prod_2^{n-1} \sum_{x_k} \psi_{0,k}(x_k) * \psi_{1,k}(x_k, \pi_{x_k}) \quad \text{marginalizing by leaf } x_n$$

$$= (\sum_{x_1} \psi_{0,1}(x_1)) * \prod_2^n \sum_{x_k} \psi_{0,k}(x_k) * \psi_{1,k}(x_k, \pi_{x_k})$$

According to proposition 4.4 we get $p(x) = p(x_1) * \prod_2^n p(x_k|\pi_{x_k})$ hence $p \in L(G)$, we proceeded by equivalences hence: $L(G) = L(G^{'})$

3 Exercice 3

3.a We run Kmeans on 100 random initialization for different norms. The folllowing table shows the mean, min and max of the number of iterations needed in each case. The algorithm converges well for all the norms.

	Mean	Max	Min
L1	15.70	38.0	1.0
L2	12.45	29.0	1.0
Linf	10.95	25.0	4.0

The L1 norm has more difficulty to converge and yeilds very different results from the other norms. A possible difference might be that the shape of the L1 norm ball is a diamond.

Centers remain nearly the same after different random initializations, and distortion is relatively stable as it only varies between 3238 and 3241.

3.b Let $k \in [1, 4]$ and $i \in [1, N]$, where N is the size of the data sample.

$$w_{ki} = \frac{\alpha_k p_{ki}}{\sum_j \alpha_j p_{ji}}$$
$$\alpha_k = \frac{\sum_i w_{ki}}{N}$$

The E-step calculation is:

$$\mu_k = \frac{N}{\alpha_k} \sum i w_{ki} x_i$$
$$\sigma_k^2 = \frac{N}{\alpha_k} \sum i w_{ki} (x_i - \mu_k)^2$$

3.c The E-step calculation changes using:

$$\Sigma_k = \frac{N}{\alpha_k} \sum_i w_{ki} (x_i - \mu_k)(x_i - \mu_k)^T$$

3.d The following table shows the likelihood of each set for the guassian mixture:

Method	Training	Test
Isotropic	-3643	-3621
General	-3103	-3304

The general case fits betters and is more adapted to the data as it is not isotropic.

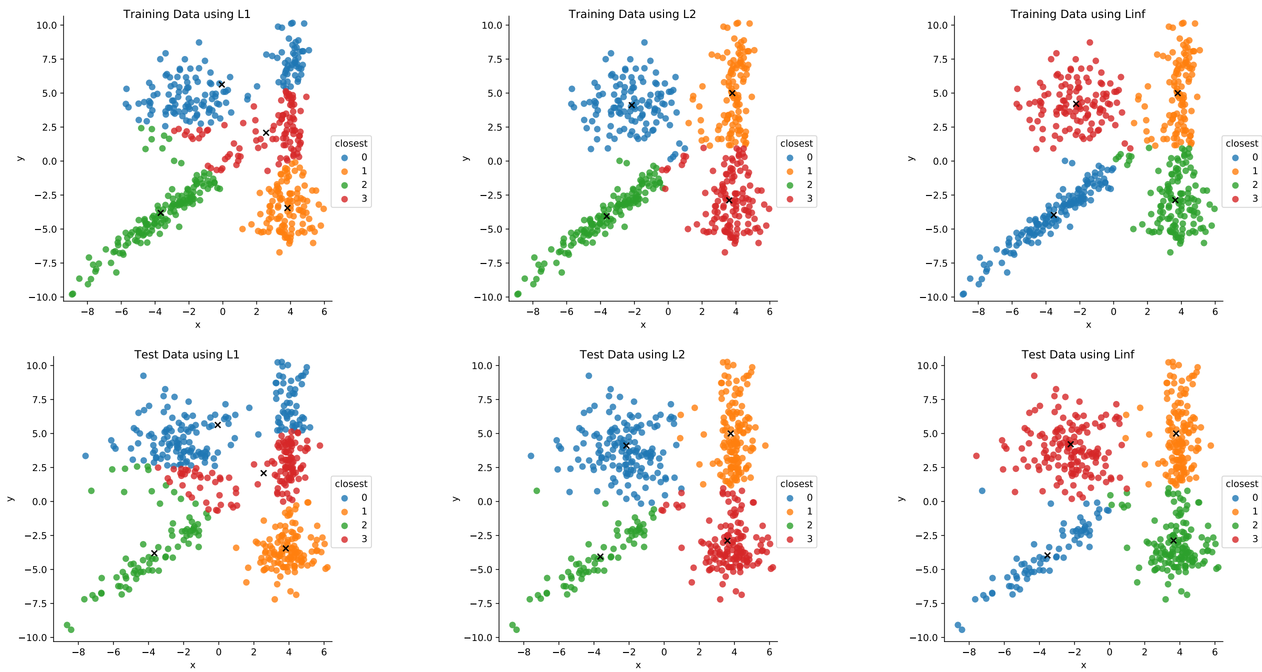

Figure 1: Plot of the training data and test data K-means.

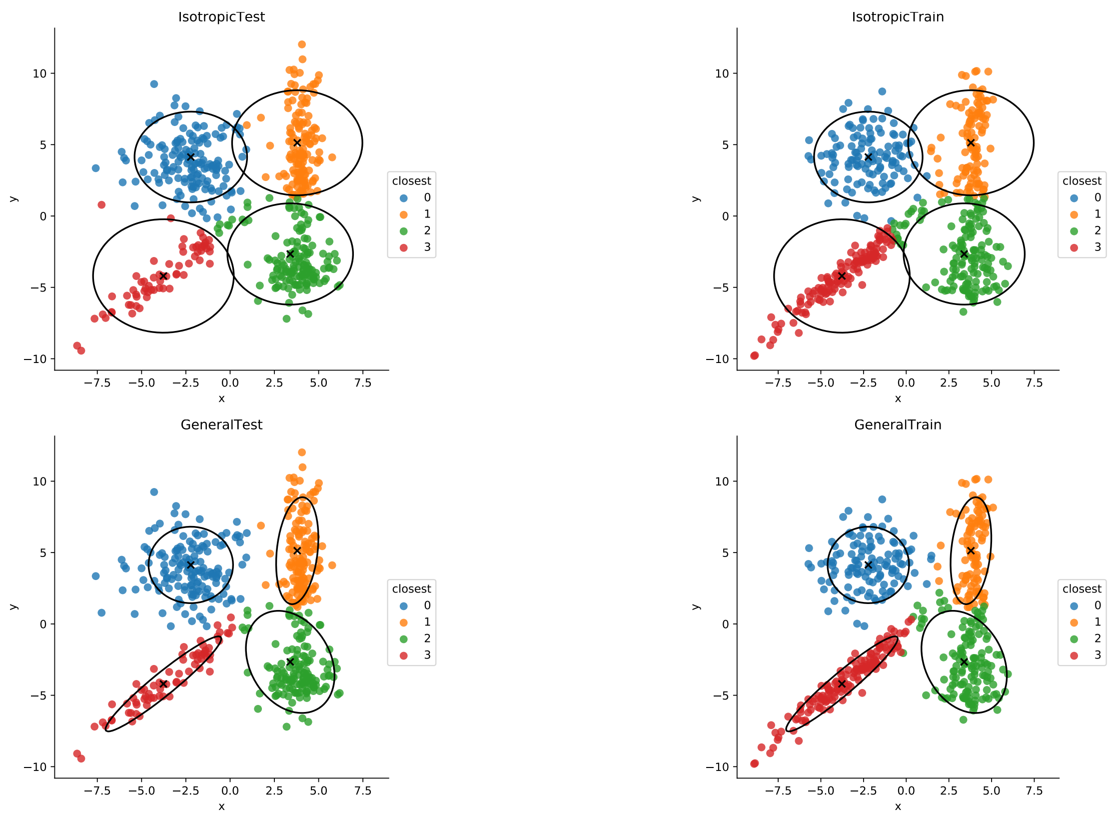

Figure 2: Plot of the training data and test data Gaussian mixture.