Probabilistic Graphical Models : Homework 2

Raphael Avalos
raphael@avalos.fr

2/11/18

1 Exercise 1

1.1 Question 1

The implied factorization for any joint distribution $p \in \mathcal{L}(G)$ is :

$$p(x, y, z, t) = p(x)p(y)p(z|x, y)p(t|z)$$

Lets take $X \sim \mathcal{B}(p), Y \sim \mathcal{B}(p), Z = X \oplus Y, T = Z$. It is clear that $X \perp\!\!\!\perp Y$ and that $X \perp\!\!\!\perp Y \not| Z$ because with X and Z we can determine Y. Therefore since $Z = T$, $X \perp\!\!\!\perp Y \not| T$

1.2 Question 2

1.2.a

We consider $Z \sim \mathcal{B}(\pi)$ with $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Y$. We can write $p(x, y)$ in two ways:

$$\begin{aligned} p(x, y) &= p(x, y \mid z = 0)p(z = 0) + p(x, y \mid z = 1)p(z = 1) \\ &= p(x \mid z = 0)p(y \mid z = 0)p(z = 0) + p(x \mid z = 1)p(y \mid z = 1)p(z = 1) \end{aligned}$$

And

$$\begin{aligned} p(x, y) &= p(x)p(y) \\ &= [p(x \mid z = 0)p(z = 0) + p(x \mid z = 1)p(z = 1)][p(y \mid z = 0)p(z = 0) + p(y \mid z = 1)p(z = 1)] \end{aligned}$$

Then we take the difference between those two expressions of $p(x, y)$ and factorize by $p(z = 0)p(z = 1) \neq 0$

$$\begin{aligned} 0 &= -p(x \mid z = 0)p(y \mid z = 0) - p(x \mid z = 1)p(y \mid z = 1) + p(x \mid z = 0)p(y \mid z = 1) + p(x \mid z = 1)p(y \mid z = 0) \\ 0 &= [p(x \mid z = 0) - p(x \mid z = 1)][p(y \mid z = 0) - p(y \mid z = 1)] \end{aligned}$$

Therefore, $X \perp\!\!\!\perp T$ or $Y \perp\!\!\!\perp T$

1.2.b

2 Exercise 2

2.1 Question 1

Let $G = (V, E)$ be a DAF, and $i \to j$ be a covered edge of G. We consider $G = (V, E')$ where $E' = (E \setminus \{i \to j\}) \cup \{j \to i\}$.

$$\begin{aligned} p(x_j \mid x_{\pi_j^G}) p(x_i \mid x_{\pi_i^G}) &= p(x_j \mid x_{\pi_i^G}, x_i) p(x_i \mid x_{\pi_i^G}) \\ &= p(x_i \mid x_{\pi_i^G}, x_j) p(x_j \mid x_{\pi_i^G}) && \textit{(Bayes)} \\ &= p(x_i \mid x_{\pi_i^{G'}}) p(x_j \mid x_{\pi_j^{G'}}) \end{aligned}$$

Since we haven't modified any other edges, we have proven that $\mathcal{L}(G) = \mathcal{L}(G')$

2.2 Question 2

Let $G = (V, E)$ a directed tree and $\tilde{G}$ the symmetrized graph (which is equal to moralized graph). The cliques of $\tilde{G}$ are by the definition of a tree the set $\mathcal{C} = \{(x, \pi_x) \mid x \in V\} \cup V$. Now let $p \in \mathcal{L}(\tilde{G})$ and consider the ψ such that $\sum_x \prod_{c \in \mathcal{C}} \psi_c(x_c) = 1$, and x_r be the root node.

$$\begin{aligned} p(x) &= \prod_{c \in \mathcal{C}} \psi_c(x_c) \\ &= \prod_{x_i \in V \setminus \{x_r\}} \psi_{x_i}(x_i) \psi_{x_i, \pi_{x_i}}(x_i, \pi_{x_i}) \psi_{x_r}(x_r) \end{aligned}$$

We can define $f(x_r, x_{\pi_{x_r}}) \propto \psi_{x_r}(x_r)$, $f(x_i, x_{\pi_{x_i}}) \propto \psi_{x_i}(x_i) \psi_{x_i, \pi_{x_i}}(x_i, \pi_{x_i})$ such that $\forall i, x_{\pi_i}, \sum_{x_i} f(x_i, x_{\pi_{x_i}}) = 1$ and therefore $p \in \mathcal{L}(G)$. So $\mathcal{L}(\tilde{G}) \subset \mathcal{L}(G)$
Now let $p \in \mathcal{L}(G)$ and x_r be the root node.

$$\begin{aligned} p(x) &= \prod_{x_i \in V \setminus \{x_r\}} p(x_i \mid x_{\pi_{x_i}}) p(x_r) \\ &\propto \prod_{x_i \in V \setminus \{x_r\}} p(x_{\pi_{x_i}} \mid x_i) p(x_i) p(x_r) \end{aligned}$$

We can define $\psi_{x_r}(x_r) = p(x_r)$, $\psi_{x_i}(x_i) = p(x_i)$ and $\psi_{x_i, \pi_{x_i}}(x_i, \pi_{x_i}) = p(\pi_{x_i} \mid x_i)$ and therefore $p \in \mathcal{L}(\tilde{G})$. So $\mathcal{L}(G) \subset \mathcal{L}(\tilde{G})$
Finally $\mathcal{L}(G) = \mathcal{L}(\tilde{G})$

3 Exercise 3

3.1 K-mean

We programmed k-mean++.

	Centroid 1		Centroid 2		Centroid 2		Centroid 4		Distortion
1	3.78809	4.99905	-3.79520	-4.24816	3.48330	-2.84991	-2.14180	3.97338	3241.28275
2	3.80280	5.10467	-3.81879	-4.27423	3.33557	-2.64452	-2.240347	4.12744	3237.77959
3	3.80280	5.10467	-3.81879	-4.27423	3.33557	-2.64452	-2.24034	4.12744	3237.77959
4	3.78809	4.99905	-3.66286	-4.11101	3.60401	-2.88772	-2.15095	4.04338	3239.87631
5	3.80280	5.10467	-3.81879	-4.27423	3.33557	-2.64452	-2.24034	4.12744	3237.77959

The algorithm is quite stable, in 5 runs we had 3 times the same result and the other two have a 2 and 4 increase in distortion.

3.2 EM: Covariance proportional to identity

By applying the method of the EM algorithm we find the following.
The E step consist in the update of the latent varialble q

$$q_{k,n}^{t+1} = \frac{\pi_k^t p(x_n \mid \mu_k^t, {\sigma_k^t}^2)}{\sum_x \pi_k^t p(x \mid \mu_k^t, {\sigma_k^t}^2)} = \frac{\frac{\pi_k^t}{\sigma_k^t} exp(-\frac{1}{2{\sigma_k^t}^2}(x_n - \mu_k^t)^T(x_n - \mu_k^t))}{\sum_x \frac{\pi_k^t}{\sigma_k^t} exp(-\frac{1}{2{\sigma_k^t}^2}(x - \mu_k^t)^T(x - \mu_k^t))}$$

The M step consist in the update of pi, σ, μ

$$\pi_k^{t+1} = \frac{\sum_n q_{k,n}^{t+1}}{N}$$
$$\mu_k^{t+1} = \frac{\sum_n q_{k,n}^{t+1} x_n}{\sum_n q_{k,n}^{t+1}}$$
$${\sigma_k^{t+1}}^2 = \frac{\sum_n q_{k,n}^{t+1}(x_n - \mu_k^t)^T(x_n - \mu_k^t)}{d\sum_n q_{k,n}^{t+1}}$$

3.3 EM: General

$$q_{k,n}^{t+1} = \frac{\pi_k^t p(x_n \mid \mu_k^t, \Sigma_k^t)}{\sum_x \pi_k^t p(x \mid \mu_k^t, \Sigma_k^t)} = \frac{\frac{\pi_k^t}{\sqrt{|\Sigma_k^t|}} exp(-\frac{1}{2}(x_n - \mu_k^t)^T {\Sigma_k^t}^{-1}(x_n - \mu_k^t))}{\sum_x \frac{\pi_k^t}{\sqrt{|\Sigma_k^t|}} exp(-\frac{1}{2}(x - \mu_k^t)^T {\Sigma_k^t}^{-1}(x - \mu_k^t))}$$
$$\Sigma_k^{t+1} = \frac{\sum_n q_{k,n}^{t+1}(x_n - \mu_k^t)(x_n - \mu_k^t)^T}{\sum_n q_{k,n}^{t+1}}$$

3.4 Evaluation

Log Likelihood	Iso	General
Train	-2682	-2345
Test	-2733	-2426
[Train,Test]	$-5326 = -2663 * 2$	$-4744 = -2372 * 2$

First, we made sure that the test data and the train data both have the same size, and therefore the comparaison of the log likelihood makes sense. Based on the log likelihoods we can deduce that by taking Σ proportional to the identity we produce results that are worst compared to the general case. This is not a surprise considering that the clusters aren't spheres. The other result that we can deduce from this comparaison is that the general case seems to have overfitted more than the iso one.

4 Plots

Figure 1: K-mean ++

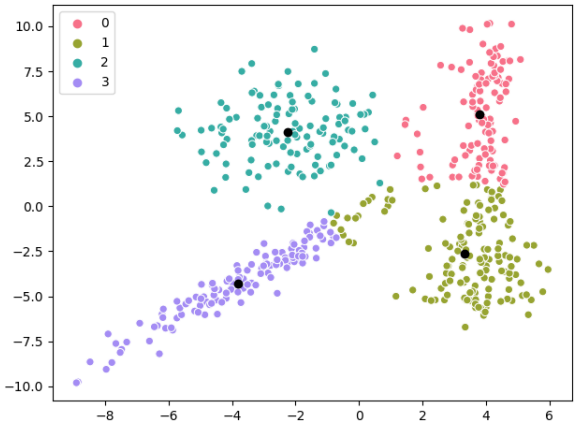

Figure 2: EM ISO

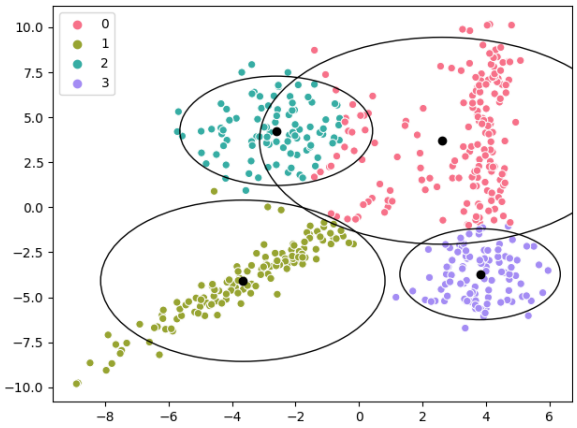

Figure 3: EM General