COMPTE-RENDU DU DEVOIR 3

Vincent Matthys

1 HMM implementation

On notera $\boldsymbol{u} = \{u_i\}_{i\in[\![1;T]\!]}$ et $\boldsymbol{q} = \{q_i\}_{i\in[\![1;T]\!]}$

1.1

Le problème de l'inférence dans la chaine de Markov modélisée, à savoir le calcul de $p(\boldsymbol{q} \mid \boldsymbol{u})$ est appréhendable par le calcul de la marginale $p(q_t \mid \boldsymbol{u})$ suivant :

$$
\begin{aligned}
p(q_t \mid \boldsymbol{u}) &= \frac{p(\boldsymbol{u} \mid q_t)p(q_t)}{p(\boldsymbol{u})} \\
&= \frac{p(u_1, \ldots, u_t \mid q_t)p(u_{t+1}, \ldots, u_T \mid q_t)p(q_t)}{p(\boldsymbol{u})} \quad u_1, \ldots, u_t \perp\!\!\!\perp u_{t+1}, \ldots, u_T \mid q_t \\
&= \frac{p(u_1, \ldots, u_t, q_t)p(u_{t+1}, \ldots, u_T \mid q_t)}{p(\boldsymbol{u})} \\
&= \frac{\alpha_t(q_t)\beta_t(q_t)}{p(\boldsymbol{u})} \\
&= \frac{\alpha_t(q_t)\beta_t(q_t)}{\sum_{q_t} \alpha_t(q_t)\beta_t(q_t)} \\
&= \frac{\alpha_t(q_t)\beta_t(q_t)}{\sum_{q_T} \alpha_T(q_T)} \quad \text{puisque} \quad p(\boldsymbol{u}) = \sum_{q_T} p(q_T, \boldsymbol{u}) = \sum_{q_T} \alpha_T(q_T)
\end{aligned} \tag{1}
$$

avec $\alpha_t(q_t) = p(u_1, \ldots, u_t, q_t)$ et $\beta_t(q_t) = p(u_{t+1}, \ldots, u_T \mid q_t)$
On peut alors aussi exprimer $p(q_t, q_{t+1} \mid \boldsymbol{u})$ en fonction de α_t et β_t :

$$
\begin{aligned}
p(q_t, q_{t+1} \mid \boldsymbol{u}) &= \frac{p(\boldsymbol{u} \mid q_t, q_{t+1})p(q_t, q_{t+1})}{p(\boldsymbol{u})} \\
&= \frac{p(\boldsymbol{u} \mid q_t, q_{t+1})p(q_{t+1} \mid q_t)p(q_t)}{p(\boldsymbol{u})} \\
&= \frac{p(u_1, \ldots, u_t \mid q_t, q_{t+1})p(u_{t+1}, \ldots, u_T \mid q_t, q_{t+1})p(q_{t+1} \mid q_t)p(q_t)}{p(\boldsymbol{u})} \\
&= \frac{p(u_1, \ldots, u_t \mid q_t)p(u_{t+1}, \ldots, u_T \mid q_{t+1})p(q_{t+1} \mid q_t)p(q_t)}{p(\boldsymbol{u})} \\
&= \frac{p(u_1, \ldots, u_t \mid q_t)p(u_{t+1} \mid q_{t+1})p(u_{t+2}, \ldots, u_T \mid q_{t+1})p(q_{t+1} \mid q_t)p(q_t)}{p(\boldsymbol{u})} \\
&= \frac{p(u_1, \ldots, u_t, q_t)p(u_{t+1} \mid q_{t+1})p(u_{t+2}, \ldots, u_T \mid q_{t+1})p(q_{t+1} \mid q_t)}{p(\boldsymbol{u})} \\
&= \frac{\alpha_t(q_t)p(u_{t+1} \mid q_{t+1})\beta_t(q_{t+1})A_{q_t,q_{t+1}}}{p(\boldsymbol{u})}
\end{aligned} \tag{2}
$$

où $A_{q_t,q_{t+1}} = p(q_{t+1} \mid q_t)$ et $p(u_{t+1} \mid q_{t+1})$ suit une loi gaussienne.

À des fins d'implémentation, on considèrera les logarithmes des α et β, $\tilde{\alpha}$ et $\tilde{\beta}$ de sorte que les formules de récursion peuvent se réecrire :

$$
\begin{cases}
\ln \alpha_t(q_t) = \ln p(u_t \mid q_t) + \ln \sum_{q_{t-1}} \left(e^{\ln \alpha_{t-1}(q_{t-1})} A_{q_{t-1},q_t}\right) \\
\ln \beta_t(q_t) = \ln \sum_{q_{t+1}} \left(e^{\ln \beta_{t+1}(q_{t+1})} A_{q_t,q_{t+1}} p(y_{t+1} \mid q_{t+1})\right)
\end{cases} \tag{3}
$$

$$
\begin{cases}
\tilde{\alpha}_t(q_t) = \ln p(u_t \mid q_t) + \ln \sum_{q_{t-1}} \left(e^{\tilde{\alpha}_{t-1}(q_{t-1})} A_{q_{t-1},q_t}\right) \\
\tilde{\beta}_t(q_t) = \ln \sum_{q_{t+1}} \left(e^{\tilde{\beta}_{t+1}(q_{t+1})} A_{q_t,q_{t+1}} p(y_{t+1} \mid q_{t+1})\right)
\end{cases} \tag{4}
$$

et que la tâche d'inférence dite de filtrage en équation (1) s'écrit :

$$
\begin{aligned}
\ln p(q_t \mid \boldsymbol{u}) &= \ln \alpha_t(q_t) + \ln \beta_t(q_t) - \ln \sum_{q_T} \alpha_T(q_T) \\
\ln p(q_t \mid \boldsymbol{u}) &= \tilde{\alpha}_t(q_t) + \tilde{\beta}_t(q_t) - \ln \sum_{q_T} e^{\tilde{\alpha}_T(q_T)}
\end{aligned} \tag{5}
$$

2

2.1

$$
\begin{aligned}
X \perp\!\!\!\perp Y \mid Z &\Leftrightarrow p(x, y \mid z) = p(x \mid z)p(y \mid z) \qquad \forall x, y, z \quad \text{t.q.} \quad p(z) > 0 \\
&\Leftrightarrow p(x, y, z) = p(x \mid z)p(y \mid z)p(z) \qquad \forall x, y, z \quad \text{t.q.} \quad p(z) > 0 \\
&\Leftrightarrow \frac{p(x, y, z)}{p(y, z)} = p(x \mid z)\frac{p(y \mid z)p(z)}{p(y, z)} \qquad \forall x, y, z \quad \text{t.q.} \quad p(y, z) > 0 \\
&\Leftrightarrow p(x \mid y, z) = p(x \mid z) \qquad \forall x, y, z \quad \text{t.q.} \quad p(y, z) > 0
\end{aligned} \tag{6}
$$

2.2

Etant donné le modèle graphique orienté G :

$$p \in \mathcal{L}(G) \Leftrightarrow \forall x, y, z, t \quad p(x,y,z,t) = p(x)p(y)p(z \mid x,y)p(t \mid z) \tag{7}$$

2.3

Non traité.

3 Distributions factorisant sur un graphe

3.1

Non traité.

3.2

Non traité.

4 Entropie et information mutuelle

4.1 Entropie

4.1.a)

Avec les conventions définies :

$$
\begin{gathered}
p_X(x) = \mathbb{P}(X = x) < 1 \Rightarrow p_X(x)\log(p_X(x)) < 0 \Rightarrow -\sum_{x\in\mathcal{X}} p_X(x)\log(p_X(x)) = H(X) \geq 0 \\
H(X) = 0 \Rightarrow \forall x \in \mathcal{X},\, p_X(x)\log(p_X(x)) = 0 \Rightarrow \forall x \in \mathcal{X},\, p_X(x) = 1 \Rightarrow \text{X is constant with probability 1}
\end{gathered}
\tag{8}
$$

4.1.b)

Par définition de la Kullback-Leibler divergence il vient :

$$
\begin{aligned}
D(p_X \parallel q) &= \sum_{x\in\mathcal{X}} p_X(x)\log\frac{p_X(x)}{q(x)} \\
&= -\sum_{x\in\mathcal{X}} p_X(x)\log q(x) - H(X) \\
&= -\sum_{x\in\mathcal{X}} p_X(x)\log\frac{1}{|X|} - H(X) \quad \text{puisque} \quad \forall x \in \mathcal{X},\, q(x) = \frac{1}{k} = \frac{1}{k} \\
&= \log k \sum_{x\in\mathcal{X}} p_X(x) - H(X) \\
&= \log k - H(X) \quad \text{car} \quad \sum_{x\in\mathcal{X}} p_X(x) = 1
\end{aligned}
\tag{9}
$$

4.1.c)

Avec les équations (8) et (9), on a directement

$$
\log k - H(X) = D(p_X \parallel q) \leq \log k \tag{10}
$$

4.2 Information mutuelle

4.2.a)

Par définition de l'information mutuelle :

$$
\begin{aligned}
I(X_1, X_2) &= \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} p_{1,2}(x_1,x_2)\log\frac{p_{1,2}(x_1,x_2)}{p_1(x_1)\,p_2(x_2)} \\
I(X_1, X_2) &= D(p_{1,2} \parallel p_1p_2) \quad \text{par définition de la Kullback-Leibler divergence}
\end{aligned}
\tag{11}
$$

Or la Kullback-Leibler divergence est positive pour toute paire $(p_{1,2}, p_1p_2)$ de distributions, donc $I(X_1, X_2) \geq 0$.

4.2.b)

Toujours avec la défintion de l'information mutuelle :

$$
\begin{aligned}
I(X_1, X_2) &= \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} p_{1,2}(x_1,x_2)\log\frac{p_{1,2}(x_1,x_2)}{p_1(x_1)\,p_2(x_2)} \\
I(X_1, X_2) &= \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} p_{1,2}(x_1,x_2)\log p_{1,2}(x_1,x_2) \\
&\quad - \sum_{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} p_{1,2}(x_1,x_2)\log\left(p_1(x_1)p_2(x_2)\right) \\
I(X_1, X_2) &= -H(X_1, X_2) \\
&\quad - \sum_{x_1\in\mathcal{X}_1}\left(\sum_{x_2\in\mathcal{X}_2} p_{1,2}(x_1,x_2)\right)\log\left(p_1(x_1)\right) \\
&\quad - \sum_{x_2\in\mathcal{X}_2}\left(\sum_{x_1\in\mathcal{X}_1} p_{1,2}(x_1,x_2)\right)\log\left(p_2(x_2)\right) \\
I(X_1, X_2) &= -H(X_1, X_2) \\
&\quad - \sum_{x_1\in\mathcal{X}_1} p_1(x_1)\log\left(p_1(x_1)\right) \\
&\quad - \sum_{x_2\in\mathcal{X}_2} p_2(x_2)\log\left(p_2(x_2)\right) \\
I(X_1, X_2) &= H(X_1) + H(X_2) - H(X_1, X_2)
\end{aligned}
\tag{12}
$$

Et ainsi l'information mutuelle peut s'écrire uniquement à partir des entropies de X_1, X_2 et de (X_1, X_2).

4.2.c)

D'après l'éqiation (12), on peut réecrire :

$$
\begin{aligned}
&H(X_1, X_2) - (H(X_1) + H(X_2)) = -I(X_1, X_2) \\
&H(X_1, X_2) - (H(X_1) + H(X_2)) \leq 0 \quad \text{puisque, } I(X_1, X_2) \geq 0
\end{aligned}
\tag{13}
$$

D'après l'équation (13), pour p_1 et p_2 données, l'entropie maximale correspond à $I(X_1, X_2) = 0$, c'est-à-dire $D(p_{1,2} \parallel p_1 p_2) = 0$ d'après l'équation (9). Or la Kullback-Leibler divergence s'anulle si et seulement si les distributions sont identiques. On a alors $p_{1,2} = p_1 p_2$, ce qui correspond à $X_1 \perp\!\!\!\perp X_2$.

5 Gaussian Mixtures

5.1 K-means