Introduction aux processus stochastiques

Chaînes de Markov cachées

Inférence à partir d'observations d'état bruitées.

Soit une chaîne de Markov cachée (π, Π, Σ) à 3 états S_1, S_2, S_3 et 3 observations E_1, E_2, E_3.
On note $\mathcal{X}_t$ la séquence des états cachés et $\mathcal{Y}_t$ la séquence des observations, pour t = $1, 2, \ldots, N$.
Pour rappel,

$$\pi_i = P(\mathcal{X}_1 = S_i) \qquad \Pi_{ij} = P(\mathcal{X}_{t+1} = S_j | \mathcal{X}_t = S_i) \qquad \Sigma_{ij} = P(\mathcal{Y}_t = E_j | \mathcal{X}_t = S_i).$$

On considère le cas particulier

$$\pi = [\ 1 \quad 1 \quad 1\ ]/3 \qquad \Pi = \begin{bmatrix} .7 & .3 & 0 \\ 0 & .7 & .3 \\ .3 & 0 & .7 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} .8 & .1 & .1 \\ .1 & .8 & .1 \\ .1 & .1 & .8 \end{bmatrix}.$$

On dispose d'une séquence $Y_1\ Y_2 \ldots Y_N$ d'observations, où chaque $Y_i \in \{E_1, E_2, E_3\}$.
On considère le cas particulier où $N = 3$ et la séquence observée est $E_1\ E_3\ E_3$.

1. Dessiner le diagramme de transition d'états de la chaîne de Markov (π, Π).
2. Dessiner le réseau bayesien de la chaîne de Markov cachée (π, Π, Σ) pour les 3 premiers instants.
3. *Propagation α/Estimation d'état.*
   Calculer de façon efficace les probabilités $P(\mathcal{X}_N = S_i | \mathcal{Y}_1 = Y_1, \ldots, \mathcal{Y}_N = Y_N)$, pour $i \in \{1, 2, 3\}$ et $N = 3$.
   En notation abrégée, la question revient à demander $P(\mathcal{X}_N | Y_1 \ldots Y_N)$ pour $N = 3$.
   Suggestion : se ramener au calcul de $P(\mathcal{X}_t = S_i, \mathcal{Y}_1 = Y_1, \ldots, \mathcal{Y}_t = Y_t)$, c'est-à-dire en notation abrégée $P(\mathcal{X}_t, Y_1 \ldots Y_N)$, pour t valant successivement 1,2,3.
4. *Propagation β/Filtrage.*
   Calculer de façon efficace $P(\mathcal{X}_t | Y_1 \ldots Y_N)$ pour $t = 1$ et $t = 2$.
   Suggestion : se ramener au calcul de $P(\mathcal{Y}_{t+1} = Y_{t+1}, \ldots, \mathcal{Y}_N = Y_N | \mathcal{X}_t = S_i)$, c'est-à-dire $P(Y_{t+1} \ldots Y_N | \mathcal{X}_t)$, et exploiter le calcul de $P(\mathcal{X}_t, Y_1 \ldots Y_N)$ du point précédent.
5. *Prédiction.*
   Calculer $P(\mathcal{X}_t | Y_1 \ldots Y_N)$. pour $N = 3$ et $t \in \{4, 5\}$.
6. Supposer à présent qu'on n'observe que Y_1 et Y_3.
   Calculer $P(\mathcal{X}_2 | Y_1, Y_3)$.
   Suggestion : Modifier les algorithmes de propagation établis aux points précédents.
7. Si on n'observe que $Y_1 = E_1$ et $Y_2 = E_3$, et qu'on recherche comme précédemment

   $$p_1 = \left[\ P(\mathcal{X}_1 = S_1 | Y_1, Y_2) \quad P(\mathcal{X}_1 = S_2 | Y_1, Y_2) \quad P(\mathcal{X}_1 = S_3 | Y_1, Y_2)\ \right]$$
   $$p_2 = \left[\ P(\mathcal{X}_2 = S_1 | Y_1, Y_2) \quad P(\mathcal{X}_2 = S_2 | Y_1, Y_2) \quad P(\mathcal{X}_2 = S_3 | Y_1, Y_2)\ \right],$$

   on trouve $p_1 = \left[\ 0.4706 \quad 0.1824 \quad 0.3471\ \right]$, $p_2 = \left[\ 0.3471 \quad 0.1824 \quad 0.4706\ \right]$. Mais si S_1 est l'état le plus probable en $t = 1$ étant donné les observations, et si S_3 l'état le plus probable en $t = 2$, par contre la *séquence* d'états la plus probable n'est pas S_1, S_3 puisque la transition $S_1 - S_3$ est interdite selon Π !
   Quel critère devrait-on maximiser pour trouver la séquence d'états la plus probable ?
   Donner le critère le plus simple possible.

Reconnaissance vocale.

On échantillonne un signal vocal. Les échantillons à l'instant $t = i\Delta$ constituent un processus stochastique. Après diverses opérations de traitement de signal sur ce processus, on obtient un processus $\mathcal{Y}_i$ qui correspond à l'observation de vecteurs du spectre $\mathcal{X}_i$.
A chaque mot à reconnaître, on peut associer un modèle de chaîne de Markov cachée où $\mathcal{X}_i$ correspond aux états et $\mathcal{Y}_i$ aux observations. Ces modèles sont obtenus en calibrant les paramètres des chaînes à partir d'une base de données contenant les mots prononcés par différents locuteurs.

Dans le problème simplifié qui vient, on suppose qu'on veut pouvoir distinguer les mots "if" et "Yves", qu'il n'existe que 3 états i, f, v notés $1, 2, 3$, plus un état terminal noté 4. Une réalisation pour "if" serait du type $1112224\ldots$, tandis qu'une réalisation pour "Yves" serait plutôt du type $111111113334\ldots$
On suppose que "if" suit la chaîne de Markov cachée HMM$_1 = (\pi, \Pi_1, \Sigma)$,
et que "Yves" suit la chaîne de Markov cachée HMM$_2 = (\pi, \Pi_2, \Sigma)$:

$$\pi = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\Pi_1 = \begin{bmatrix} .7 & .3 & 0 & 0 \\ 0 & .7 & 0 & .3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \Pi_2 = \begin{bmatrix} .9 & 0 & .1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & .7 & .3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & .6 & .4 & 0 \\ 0 & .4 & .6 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

On observe la séquence

$$Y_1 \ldots Y_N = 1111233444\ldots$$

Reconnaître le mot revient à trouver lequel des 2 modèles explique le mieux les observations, i.e. trouver le modèle qui maximise $P(Y_1 \ldots Y_N)$.

1. Calculer $p_1 = P(Y_1 \ldots Y_8 | \text{HMM}_1)$.
2. Calculer $p_2 = P(Y_1 \ldots Y_8 | \text{HMM}_2)$.
3. En déduire le mot le plus probable pour la séquence observée.

Décodage de messages.

On considère des séquences $\mathcal{X}_t$ avec $t = 1, 2, \ldots$ et $\mathcal{X}_t = \{1, 2\}$.
Les séquences suivent un modèle de Markov (π, Π) :

$$\pi = \begin{bmatrix} .5 & .5 \end{bmatrix} \qquad\qquad \Pi = \begin{bmatrix} .8 & .2 \\ .4 & .6 \end{bmatrix}.$$

Les séquences sont reçues bruitées par un récepteur. Le bruit est tel que les S_1 sont vus comme des S_2 avec une probabilité 0.2, et les S_2 comme des S_1 avec une probabilité 0.2.
Soit $[Y_1 \ldots Y_4] = [2\ 1\ 2\ 2]$ la séquence observée au récepteur (processus $\mathcal{Y}_t$).

Quelle est la séquence originale $[X_1 \ldots X_4]$ la plus probable ?

1. Modéliser la situation par une chaîne de Markov cachée (π, Π, Σ).
2. Trouver la séquence $X_1 \ldots X_4$ la plus probable.
   Pour ce faire, maximiser

   $$P(\mathcal{X}_1 = X_1, \ldots, \mathcal{X}_4 = X_4 | Y_1 \ldots Y_4)$$

   en fonction des $X_t \in \{1, 2\}$, $1 \le t \le 4$.

   Suggestion :

   (a) Montrer que le problème revient à minimiser

   $$f(X_1, \ldots, X_4) = -\ln[(P(X_1)P(Y_1|X_1)] - \sum_{k=2}^{4} \ln[(P(X_k|X_{k-1})P(Y_k|X_k)]$$

   en choisissant la séquence $X_1, \ldots X_4$.

   (b) Montrer ensuite que ce problème revient à chercher le chemin de coût minimum dans un graphe à 9 noeuds, 1 noeud initial plus 8 noeuds étiquetés $\mathcal{X}_k = j$, avec $1 \le k \le 4$ et $1 \le j \le 2$. Le graphe comporte des arcs orientés du noeud initial vers les noeuds $\mathcal{X}_1 = j$, qui ont un coût $b_j^{(2)}$, et des arcs orientés allant des noeuds $\mathcal{X}_{k-1} = i$ aux noeuds $\mathcal{X}_k = j$, qui ont un coût $a_{ij}^{(q)}$ où q correspond à la valeur $Y_k = q$ observée.
   Un chemin part du noeud initial et arrive à un des 2 noeuds $\mathcal{X}_4 = j$; le coût du chemin est la somme des coûts des arcs qui le composent.
   Les coûts sont

   $$\begin{aligned} b_j^{(2)} &= -\ln[P(X_1 = j)P(Y_1 = 2|X_1 = j)] \\ a_{ij}^{(q)} &= -\ln[P(X_k = j|X_{k-1} = i)P(Y_k = q|X_k = j)] \end{aligned}$$

   pour $1 \le i, j, q \le 2$.

   (c) Calculer numériquement les coûts et trouver le chemin de coût minimum, en utilisant la propriété suivante :
   Si $N_0 \ldots N_k$ est un chemin de coût minimum d'un noeud N_0 à un noeud N_k, alors $N_0 \ldots N_{k-1}$ doit être un chemin de coût minimum du noeud N_0 au noeud N_{k-1}.