

Apprentissage

Partiel - Mardi 29 mars 2005

Partie Cours

2 heures - Documents non autorisés

Exercice 1 - Chaînes de Markov cachées

On veut construire un HMM modélisant une famille de séquences d'acides nucléiques dont les propriétés sont les suivantes :

1. le premier acide nucléique est toujours un A, le dernier soit un C soit un G,
2. en dehors de ces deux acides nucléiques (premier et dernier), la plupart des séquences sont composées de deux types de régions distinctes (1 et 2, composées d'un ensemble d'acides nucléiques), mais certaines séquences ne comportent qu'une région 1 ou une région 2,
 - (a) dans la région 1, chacune des 4 bases A, C, G et T a la même probabilité d'apparaître,
 - (b) dans la région 2, G et C ont deux fois plus de chance d'apparaître que A et T.

Soit la chaîne de Markov à quatre états D, R_1, R_2, F de matrice de transition :

$$T = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 0 & 0.7 & 0.1 & 0.2 \\ 0 & 0.5 & 0.3 & 0.2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- 1- Préciser les éléments permettant de construire à partir de cette chaîne de Markov une HMM modélisant la famille de séquences d'acides nucléiques
- 2- Calculer le chemin le plus probable pour la séquence ATGC
- 3- À partir de la structure du HMM, que pouvez-vous dire sur la longueur des régions 1 et 2 et le nombre d'occurrences de chaque région dans les séquences ayant servies à définir ce HMM ?

RAPPEL : ALGORITHME DE VITERBI

Etant donné une séquence de symboles $O = o_1 \cdots o_T$ et un HMM $H = (S, \Sigma, T, G)$, on recherche la séquence d'états du HMM qui a la probabilité maximale de générer O . Cette dernière, et les états $\psi_t(s)$ correspondants, sont calculés de manière inductive :

1. $(\forall s \in S) \delta_1(s) = P(\text{start} \rightarrow s)P(o_1 | s)$ et $\psi_1(s) = \text{start}$
2. $(\forall s \in S)(\forall t \in \{2 \cdots T\}) \delta_t(s) = \max_{s' \in S} (\delta_{t-1}(s')P(s' \rightarrow s)) P(o_t | s)$ et $\psi_t(s) = \arg \max_{s' \in S} (\delta_{t-1}(s')P(s' \rightarrow s))$
3. $P(O|H) = \max_{s \in S} (\delta_T(s)P(s \rightarrow \text{end}))$

Le chemin de Viterbi est alors donné en lançant la procédure

1. $s_T^* = \arg \max_{s \in S} (\delta_T(s) P(s \rightarrow \text{end}))$
2. $(\forall t \in \{T-1 \dots 0\}) s_t^* = \psi_{t+1}(s_{t+1}^*)$

Exercice 2 - Réseaux de neurones

On considère un perceptron multicouche à deux entrées, une couche cachée comportant deux neurones, et une sortie. On pose :

- $X = (x, y, x_0 = 1)^T$
- $W_1 = (W_1^x, W_1^y, W_1^0)^T$, et $W_2 = (W_2^x, W_2^y, W_2^0)^T$.

La sortie du réseau est calculée par

$$c = \sigma(a_1 \sigma(W_1^T X) + a_2 \sigma(W_2^T X) + a_0)$$

avec

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad \text{et donc} \quad \sigma'(z) = \sigma(z)(1 - \sigma(z))$$

Pour une sortie calculée c lorsque la sortie désirée est o , la fonction de coût du réseau est

$$J_1(o, c) = o \log \frac{o}{c} + (1 - o) \log \frac{1 - o}{1 - c}$$

- 1- Représenter ce réseau
- 2- Calculer les gradients de la fonction de coût en fonction des paramètres du réseau
- 3- Comparer le résultat avec le gradient de la fonction de coût quadratique $J_2(o, c) = \frac{1}{2}(c - o)^2$. Quel coût vous semble le meilleur et dans quel cas ?
- 4- Que se passerait-il pour J_1 et J_2 si $c = a_1 \sigma(W_1^T X) + a_2 \sigma(W_2^T X) + a_0$?