CS123a Statistical Machine Learning (Spring 2013): Homework Assignment #6

Chuan Wang

April 20, 2013

(1) Calculate $P(ACGT \mid \lambda)$

According to Forward Algorithm,

$$\alpha_t(i) = P(o_1, \ldots, o_t, q_t; \lambda) = [\sum_{j=1}^{N} \alpha_{t-1} a_{ji}] b_{io_t}$$

we have,

$$\begin{aligned} \alpha_1(1) &= \pi_1 b_{11} = 0.5 \times 0.4 = 0.2 \\ \alpha_1(2) &= \pi_2 b_{21} = 0.5 \times 0.2 = 0.1 \\ \boldsymbol{\alpha}_1 &= \begin{bmatrix} 0.2 & 0.1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \boldsymbol{\alpha}_2 &= \boldsymbol{\alpha}_1 \mathbf{A} \mathbf{b_2} \\ &= \begin{bmatrix} 0.2 & 0.1 \end{bmatrix} \begin{bmatrix} 0.96 & 0.04 \\ 0.1 & 0.9 \end{bmatrix} * \begin{bmatrix} 0.1 \\ 0.3 \end{bmatrix} = \begin{bmatrix} 0.0202 & 0.0294 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \boldsymbol{\alpha}_3 &= \boldsymbol{\alpha}_2 \mathbf{A} \mathbf{b_3} \\ &= \begin{bmatrix} 0.0202 & 0.0294 \end{bmatrix} \begin{bmatrix} 0.96 & 0.04 \\ 0.1 & 0.9 \end{bmatrix} * \begin{bmatrix} 0.1 \\ 0.3 \end{bmatrix} = \begin{bmatrix} 0.0022 & 0.0081 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \boldsymbol{\alpha}_4 &= \boldsymbol{\alpha}_3 \mathbf{A} \mathbf{b_4} \\ &= \begin{bmatrix} 0.0022 & 0.0081 \end{bmatrix} \begin{bmatrix} 0.96 & 0.04 \\ 0.1 & 0.9 \end{bmatrix} * \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.0012 & 0.0015 \end{bmatrix} \end{aligned}$$

$$P(ACGT \mid \lambda) = 0.0012 + 0.0015 = 0.0027$$

(2) How many possible explanations (hidden state sequences) for ACATCGTCGGTAGT ?

2^{14}

(3) Calculate the best hidden state sequence for ACATCGTCGGTAGT.

The best hidden state sequences are all non-coding.

All the results are from the program vertibi.py. To run it just type in:
python vertibi.py