Ancestral Reconstruction Review

1. In the figure below, the “bottom-up” phase of Fitch’s algorithm has been completed. Perform the “top-down” phase to assign a state to each internal vertex, and show where mutations have occurred on the tree. What is the total mutation score?

![Tree Diagram](image)

2. In the figure below, the “bottom-up” phase has again been completed, but for Sankoff’s algorithm with the scoring matrix $\sigma$. Perform the traceback phase to assign a state to each internal vertex, and show where mutations have occurred on the tree. What is the total mutation score?

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>$b$</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

![Tree Diagram with scoring matrix](image)

3. What is the runtime of Fitch’s algorithm in terms of the number of samples $n$ and the number of character states $k$? What is the runtime of Sankoff’s algorithm?

4. Is there any way to relate Fitch’s algorithm and Sankoff’s algorithm? Is one a special case of the other?
Viterbi Algorithm Practice

Suppose we have an HMM with $K = 2$ hidden states representing two weighted coins (coin 1 and coin 2). Our emissions are represented as the observed outcomes ($H$ or $T$) of coin tosses. At first, say we are given the following transition and emission probabilities:

$$
\begin{pmatrix}
  a_{11} = \frac{1}{2} & a_{12} = \frac{1}{2} \\
  a_{21} = \frac{1}{5} & a_{22} = \frac{4}{5}
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
  e_1(H) = \frac{2}{3} & e_1(T) = \frac{1}{3} \\
  e_2(H) = \frac{1}{4} & e_2(T) = \frac{3}{4}
\end{pmatrix}
$$

Note that the rows sum to 1. Also say we are given the initial state probabilities $\pi_1 = \frac{1}{2}$ and $\pi_2 = \frac{1}{2}$. Now we want to find the most likely path (Viterbi path) of hidden states for a given dataset using dynamic programming. Let $V_k(i)$ be the probability of the most probable path that ends in hidden state $k$ at position $i$ in the data. We will initialize the Viterbi recursive data structure with:

$$V_k(1) = \pi_k \cdot e_k(x_1)$$

And fill in each subsequent column using the previous column:

$$V_k(i) = e_k(x_i) \cdot \max_l \{ V_l(i-1) \cdot a_{lk} \}$$

1. Given the observed sequence $\vec{x} = (H, T, H)$ and the probabilities above, fill in the table for $V$ below, then use backpointers to find the most likely sequence of hidden states.

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>T</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Now suppose we have the opposite information - we are given the hidden state sequence $\vec{z}$ and want to estimate the probabilities. What are the new transition and emission probabilities $a_{kl}$ and $e_k(b)$?

<table>
<thead>
<tr>
<th>hidden state sequence $\vec{z}$</th>
<th>2</th>
<th>1</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>1</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>observed sequence $\vec{x}$</td>
<td>T</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>T</td>
<td>H</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>H</td>
<td>H</td>
<td>T</td>
</tr>
</tbody>
</table>

Page 2 of 2