8.1. Suppose that the five amino acid sequences are produced by a profile hidden Markov model with length 7 (m7 is the stop state). Below are listed each amino acid sequence with the corresponding path in the HMM that produced it. Use this data to align the 5 sequences.

- AWWCCRPR \( m_0 m_1 i_1 m_2 m_3 i_3 m_4 i_4 m_5 d_6 m_7 \)
- AWCCCR \( m_0 m_1 m_2 m_3 m_4 m_5 m_6 m_7 \)
- RACCPR \( m_0 i_0 d_1 m_2 m_3 i_3 m_4 d_5 m_6 m_7 \)
- CWCCC \( m_0 m_1 d_2 d_3 m_5 m_6 i_6 m_7 \)
- CWHCCR \( m_0 m_1 i_2 m_3 d_3 m_5 m_6 m_7 \)

8.2 Consider the following hidden Markov chain model. The hidden chain itself has a start state, labeled 0, and three other states, \( \{1, 2, 3\} \). The transition matrix of the chain is

\[
\begin{pmatrix}
0 & 0.3 & 0.3 & 0.4 \\
0 & 0.6 & 0.2 & 0.2 \\
0 & 0.1 & 0.2 & 0.7 \\
0 & 0.2 & 0.3 & 0.5
\end{pmatrix}
\]

The output alphabet is \( \{A, C, G, T\} \) and the emission probabilities are

\[
\begin{pmatrix}
e_1(A) & e_1(C) & e_1(G) & e_1(T) \\
e_2(A) & e_2(C) & e_2(G) & e_2(T) \\
e_3(A) & e_3(C) & e_3(G) & e_3(T)
\end{pmatrix} = \begin{pmatrix}
.4 & .1 & .1 & .4 \\
.3 & .25 & .3 & .15 \\
0 & .4 & .5 & .1
\end{pmatrix}
\]

(a) Find the probability of \( P(AAGTC, 12231) \) and the probability of \( P(GGAC, 3132) \).

(b) Find \( P(GGA) \) using the forward algorithm.

(c) Find the path in the hidden chain that most likely produced \( GGA \) using the Viterbi algorithm.

8.3.*** Consider the example worked on pages 16 and 17 of the chapter 8 notes on hidden Markov models. For that example:

a) Find the probability of the output sequence \( ABC \).

b) Find the path \( s_1 s_2 s_3 s_4 s_5 \) that maximizes the probability of the output sequence \( ABCBB \). (Note: the example works out the algorithm for \( ABC \). You will need to continue the procedure for two more columns and find the optimal path.)