1. In *Viterbi training*, the most probable path is used, as opposed to using the entire forward and backward tables as in Baum-Welch EM. Suppose we have an HMM with hidden state space $S = \{1, 2\}$ representing two weighted coins, and emission state space $\Sigma = \{H, T\}$ representing the observed outcomes of coin tosses. At a certain iteration, suppose you have obtained the following transition and emission probabilities

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

(a) Now we want to find the most likely path (Viterbi path) of hidden states for a given dataset using dynamic programming. Let $v_t(k)$ be the probability of the most probable path that ends in hidden state k at position t in the data. What is the base case ($t = 1$) and recursive call?

Solution: Let π_k be the probability of beginning in state k. Although the initial distribution is often denoted π, it does not necessarily need to be the stationary distribution of the transition matrix, although this is often a good choice in the absense of other information about the start state of the Markov chain. If our observed sequence of emitted states is x, then the base case for the Viterbi algorithm is then,

$$v_1(k) = \pi_k \cdot e_k(x_1),$$

and the recursion is

$$v_t(k) = e_k(x_t) \cdot \max_{j \in S}\{v_{t-1}(j) \cdot a_{jk}\}.$$

(b) Given the observed sequence $x = (H, T, H)$ and the probabilities above, fill in the table for v below, then use backpointers to find the most likely sequence of hidden states.

Solution: In this case, assume that the initial probabilities of both states are equal, so $\pi_1 = \pi_2 = \frac{1}{2}$ (it would also be fine to use the stationary distribution of the transition matrix). So for $t = 1$, the Viterbi probabilities are

$$v_1(1) = \pi_1 \cdot e_1(H) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}, \quad \text{and} \quad v_2(1) = \pi_2 \cdot e_2(H) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}.$$

We can then plug these into the recursion to obtain the Viterbi probabilities for $t = 2$ and $t = 3$:

$$v_2(1) = e_1(T) \cdot \max\{v_1(1) \cdot a_{11}, v_1(2) \cdot a_{21}\} = \frac{1}{3} \max\left\{\frac{1}{6}, \frac{1}{40}\right\} = \frac{1}{18}$$

$$v_2(2) = e_2(T) \cdot \max\{v_1(1) \cdot a_{12}, v_1(2) \cdot a_{22}\} = \frac{3}{4} \max\left\{\frac{1}{6}, \frac{1}{10}\right\} = \frac{1}{8}$$

$$v_3(1) = e_1(H) \cdot \max\{v_2(1) \cdot a_{11}, v_2(2) \cdot a_{21}\} = \frac{2}{3} \max\left\{\frac{1}{36}, \frac{1}{40}\right\} = \frac{1}{54}$$

$$v_3(1) = e_2(H) \cdot \max\{v_2(1) \cdot a_{12}, v_2(2) \cdot a_{22}\} = \frac{1}{4} \max\left\{\frac{1}{36}, \frac{1}{10}\right\} = \frac{1}{40}$$

Filling in the table with these values and adding backpointers as we go, we obtain:

	H	T	H
1	1/3	1/18	1/54
2	1/8	1/8	1/40

To find the most probable hidden state sequence, we look for the maximum value in the last column, then use the backpointers to read off the path. In this case we get

$$(Q_1, Q_2, Q_3) = (1, 2, 2).$$

(c) Now suppose you have a longer observed sequence, and have completed the Viterbi decoding to obtain the most probable hidden state sequence below. What are the updated transition and emission probabilities a_{kl} and $e_k(\sigma)$?

hidden state	2	1	2	1	1	1	2	2	2	2	2	2	2	2	1	2	2
emitted state	T	H	H	H	T	H	T	T	H	T	T	T	T	H	H	T	T

Solution: We first obtain the counts of transitions between each pair of states, and emissions of H and T from each state:

$$A_{11} = 2, A_{12} = 3, A_{21} = 3, A_{22} = 8$$

$$E_1(H) = 4, E_1(T) = 1, E_2(H) = 3, E_2(T) = 9$$

After normalizing (see Homework 4, #4), we get our new maximum likelihood estimates for the transition and emission probabilities:

$$\begin{pmatrix} \hat{a}_{11} = \frac{2}{5} & \hat{a}_{12} = \frac{3}{5} \\ \hat{a}_{21} = \frac{3}{11} & \hat{a}_{22} = \frac{8}{11} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \hat{e}_1(H) = \frac{4}{5} & \hat{e}_1(T) = \frac{1}{5} \\ \hat{e}_2(H) = \frac{1}{4} & \hat{e}_2(T) = \frac{3}{4} \end{pmatrix}$$

2. The Viterbi algorithm finds the most probable path ending in state k at position t in the sequence x, i.e.

$$v_t(k) = \max_{q_{1:t-1}} \left\{ \mathbb{P}(x_{1:t}, Q_{1:t-1} = q_{1:t-1}, Q_t = k) \right\}.$$

Prove the Viterbi recursion from question 1.

Solution: We can use the definition of conditional probability, and the conditional independence properties implied by the HMM structure:

$$\begin{aligned}
v_t(k) &= \max_{q_{1:t-1}} \left\{ \mathbb{P}(x_t, Q_t = k, x_{1:t-1}, Q_{1:t-1} = q_{1:t-1}) \right\} \\
&= \max_{q_{1:t-1}} \left\{ \mathbb{P}(x_t, Q_t = k | x_{1:t-1}, Q_{1:t-1} = q_{1:t-1}) \mathbb{P}(x_{1:t-1}, Q_{1:t-1} = q_{1:t-1}) \right\} \\
&= \max_{q_{1:t-1}} \left\{ \mathbb{P}(x_t, Q_t = k | Q_{t-1} = q_{t-1}) \mathbb{P}(x_{1:t-1}, Q_{1:t-1} = q_{1:t-1}) \right\} \\
&= \max_{q_{1:t-1}} \left\{ \mathbb{P}(x_t | Q_t = k) \mathbb{P}(Q_t = k | Q_{t-1} = q_{t-1}) \mathbb{P}(x_{1:t-1}, Q_{1:t-1} = q_{1:t-1}) \right\} \\
&= e_k(x_t) \cdot \max_{q_{t-1}} \left\{ a_{q_{t-1},k} \cdot \max_{q_{1:t-2}} \{ \mathbb{P}(x_{1:t-1}, Q_{1:t-2} = q_{1:t-2}, Q_{t-1} = q_{t-1}) \} \right\}
\end{aligned}$$

We have now obtained $v_t(k)$ in terms of $v_{t-1}(q_{t-1})$. Since q_{t-1} can be any $j \in S$, we can rewrite the final equation and obtain the Viterbi recursion:

$$v_t(k) = e_k(x_t) \cdot \max_{j \in S} \{ v_{t-1}(j) \cdot a_{jk} \}.$$