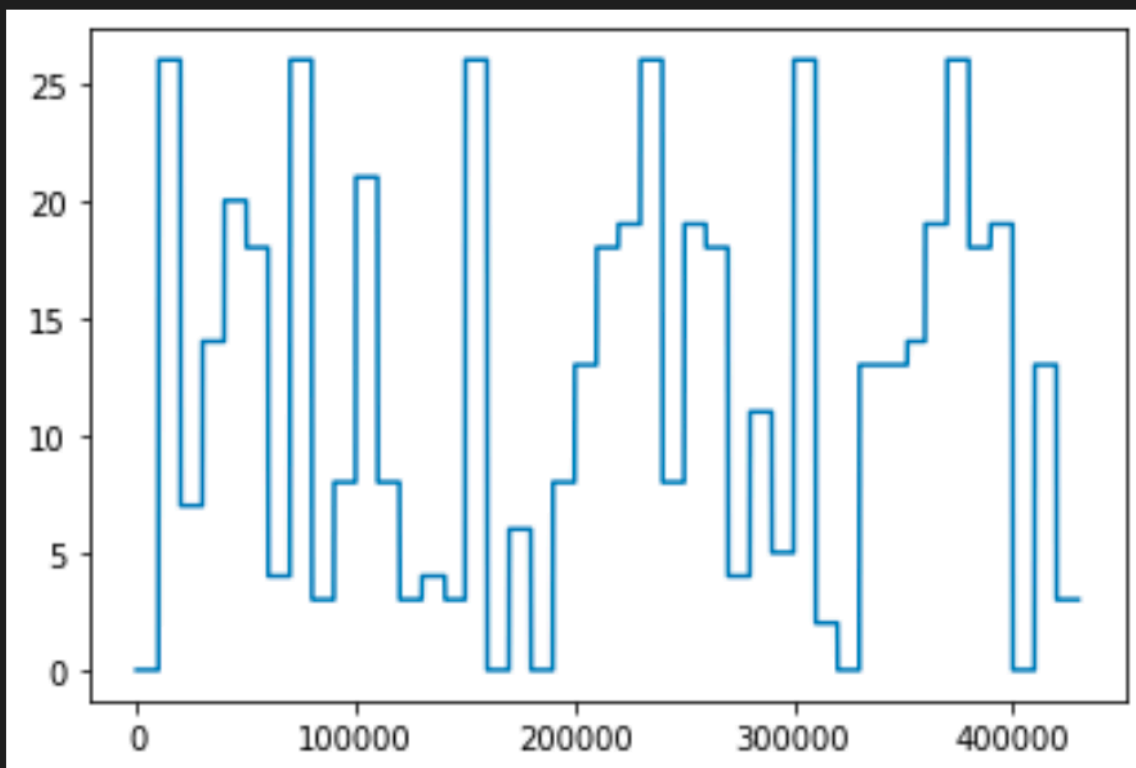
CSE 250A. Assignment 7 — ESKIL BERG OULD-SAADA

7.1 Viterbi algorithm

Source code is on next page

Plot with correct message:



a house divided against itself canot stand

SOURCE CODE

```
import numpy as np
import matplotlib.pyplot as plt
import string
```
✓ 0.4s

```
initial = np.loadtxt("initialStateDistribution.txt", dtype=float)
transition = np.loadtxt("transitionMatrix.txt", dtype=float)
emission = np.loadtxt("emissionMatrix.txt", dtype=float)
observations = np.loadtxt("observations.txt", dtype=int)
```
✓ 0.3s

```
n, m, T = 27, 2, 430000
alphabet = dict(zip(range(1,28), string.ascii_lowercase + ' '))
```
✓ 0.3s

```
L = np.zeros((n,T))
F = np.zeros((n,T))
L[:,0] = np.log(initial[0]) + np.log(emission[:,observations[0]])
F[:,0] = initial
s = np.full(T, -1, dtype=int)
```
✓ 0.1s

```
def viterbi(F, L, transition, emission, observations, s):
    # Fill F and L
    for t in range(1,T):
        for i in range(n):
            logexp = L[:,t-1] + np.log(transition[:,i])
            max = np.argmax(logexp)
            max_logexp = logexp[max]
            F[i,t] = max
            L[i,t] = max_logexp + np.log(emission[i,observations[t]])
    for t in range(T-1,-1,-1):
        if t == T-1:
            s[t] = np.argmax(L[:,T-1])
        else:
            s[t] = F[s[t+1], t+1]
    word = []
    for t in range(T-1):
        if s[t] != s[t+1]:
            word.append(alphabet.get(s[t]+1))
    word.append(alphabet.get(s[T-1]+1))
    word = ''.join(word)
    return word, s
```
✓ 0.5s

```
word, S = viterbi(F, L, transition, emission,observations, s)
plt.plot(S)
plt.show()
print(word)
```
✓ 1m 47.4s

7.2 Inference in HMMs

Consider a discrete HMM with hidden states S_t, observations O_t, transition matrix $a_{ij} = P(S_{t+1} = j | S_t = i)$ and emission matrix $b_{ik} = P(O_t = k | S_t = i)$. In class, we defined the forward-backward probabilities:

$$\begin{aligned} \alpha_{it} &= P(o_1, o_2, \ldots, o_t, S_t = i), \\ \beta_{it} &= P(o_{t+1}, o_{t+2}, \ldots, o_T | S_t = i), \end{aligned}$$

for a particular observation sequence $\{o_1, o_2, \ldots, o_T\}$ of length T. In terms of these probabilities, which you may assume to be given, as well as the transition and emission matrices of the HMM, show how to (efficiently) compute the following posterior probabilities.

The key tools, as usual, are the product rule, conditional independence (CI), and marginalization.

(a) $P(S_{t+1} = j | S_t = i, o_1, o_2, \ldots, o_T)$

$$\overset{\text{P.R.}}{=} \frac{P(S_{t+1} = j, S_t = i, O_1, \ldots, O_T)}{P(S_t = i, O_1, \ldots, O_T)}$$

$$\overset{\text{CI}}{=} \frac{P(S_t = i, O_1, \ldots, O_t) P(S_{t+1} = j | S_t = i) P(O_{t+1} | S_{t+1} = j) P(O_{t+2}, \ldots, O_T | S_{t+1} = j)}{P(S_t = i, O_1, \ldots, O_t) P(O_{t+1}, \ldots, O_T | S_t = i)}$$

$$= \frac{\cancel{\alpha_{it}}\, a_{ij}\, b_j(O_{t+1}) \beta_{j\,t+1}}{\cancel{\alpha_{it}}\, \beta_{it}}$$

$$= \frac{a_{ij}\, b_j(O_{t+1}) \beta_{j\,t+1}}{\beta_{it}}$$

(b) $P(S_t = i | S_{t+1} = j, o_1, o_2, \ldots, o_T)$

$$\overset{\text{P.R.}}{=} \frac{P(S_{t+1} = j, S_t = i, O_1, \ldots, O_T)}{P(S_{t+1} = j, O_1, \ldots, O_T)}$$

$$\overset{\text{CI}}{=} \frac{P(S_t = i, O_1, \ldots, O_t) P(S_{t+1} = j | S_t = i) P(O_{t+1} | S_{t+1} = j) P(O_{t+2}, \ldots, O_T | S_{t+1} = j)}{P(S_{t+1} = j, O_1, \ldots, O_{t+1}) P(O_{t+2}, \ldots, O_T | S_{t+1} = j)}$$

$$= \frac{\alpha_{it}\, a_{ij}\, b_j(O_{t+1}) \cancel{\beta_{j\,t+1}}}{\alpha_{j\,t+1} \cancel{\beta_{j\,t+1}}}$$

$$= \frac{\alpha_{it}\, a_{ij}\, b_j(O_{t+1})}{\alpha_{j\,t+1}}$$

(c) $P(S_{t-1}=i, S_t=k, S_{t+1}=j | o_1, o_2, \ldots, o_T) \overset{\text{P.R.}}{=} \dfrac{P(S_{t-1}=i, S_t=k, S_{t+1}=j, o_1 o_2, \ldots, o_T)}{P(o_1, \ldots, o_T)}$

$$\overset{\text{P.R.}}{=} \frac{P(S_{t-1}=i, o_1, \ldots, o_{t-1}) P(S_t=k, S_{t+1}=j, o_t, \ldots, o_T | S_{t-1}=i, o_1, \ldots, o_{t-1})}{P(o_1, \ldots, o_T)}$$

$$\overset{\text{P.R.}}{=} \frac{P(S_{t-1}=i, o_1, \ldots, o_{t-1}) P(S_t=k | S_{t-1}=i, o_1, \ldots, o_{t-1}) P(S_{t+1}=j, o_t, \ldots, o_T | S_{t-1}=i, o_1, \ldots, o_{t-1}, S_t=k)}{P(o_1, \ldots, o_T)}$$

$$\overset{\text{CI}}{=} \frac{\overset{\alpha_{it-1}}{P(S_{t-1}=i, o_1, \ldots, o_{t-1})} \overset{a_{ik}}{P(S_t=k | S_{t-1}=i)} \overset{b_k(o_t)}{P(o_t | S_t=k, S_{t-1}=i, o_1, \ldots, o_{t-1})} P(o_{t+1}, \ldots, o_T, S_{t+1}=j | S_t=k, S_{t-1}=i, o_1, \ldots, o_t)}{P(o_1, \ldots, o_T)}$$

$$= \frac{\alpha_{it-1} a_{ik} b_k(o_t) P(o_{t+1}, \ldots, o_T, S_{t+1}=j | S_t=k, S_{t-1}=i, o_1, \ldots, o_{t-1})}{P(o_1, \ldots, o_T)}$$

$$\overset{\text{CI}}{=} \frac{\alpha_{it-1} a_{ik} b_k(o_t) P(S_{t+1}=j | S_t=k) P(o_{t+1} | S_{t+1}=j, S_t=k, S_{t-1}=i, o_1, \ldots, o_{t-1}) P(o_{t+2}, \ldots, o_T | o_{t+1}, S_{t+1}=j, S_t=k, S_{t-1}=i, o_1, \ldots, o_{t-1})}{P(o_1, \ldots, o_T)}$$

$$= \frac{\alpha_{it-1} a_{ik} b_k(o_t) a_{kj} b_j(o_{t+1}) \beta_{jt+1}}{P(o_1, \ldots, o_T)} \overset{\text{marg}}{=} \frac{\alpha_{it-1} a_{ik} b_k(o_t) a_{kj} b_j(o_{t+1}) \beta_{jt+1}}{\sum_x P(o_1, \ldots, o_T, S_t=x)} \overset{\text{P.R.}}{=} \frac{\alpha_{it-1} a_{ik} b_k(o_t) a_{kj} b_j(o_{t+1}) \beta_{jt+1}}{\sum_x P(o_1, \ldots, o_t, S_t=x) P(o_{t+1}, \ldots, o_T | S_t=x, o_1, \ldots, o_t)}$$

$$= \frac{\alpha_{it-1} a_{ik} b_k(o_t) a_{kj} b_j(o_{t+1}) \beta_{jt+1}}{\sum_x \alpha_{xt} \beta_{xt}}$$

(d) $P(S_{t+1}=j | S_{t-1}=i, o_1, o_2, \ldots, o_T)$

$$\overset{\text{P.R.}}{=} \frac{P(S_{t-1}=i, S_{t+1}=j, o_1 o_2, \ldots, o_T)}{P(S_{t-1}=i, o_1, \ldots, o_T)} \underset{\text{CI}}{\overset{\text{marg}}{=}} \frac{\sum_k P(S_{t-1}=i, S_t=k, S_{t+1}=j, o_1 o_2, \ldots, o_T)}{P(S_{t-1}=i, o_1, \ldots, o_{t-1}) P(o_t, \ldots, o_T | S_{t-1}=i)}$$

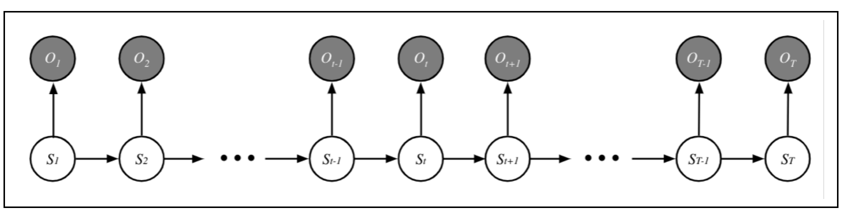
$$\overset{\text{from c)}}{=} \frac{\sum_k \cancel{\alpha_{it-1}} a_{ik} b_k(o_t) a_{kj} b_j(o_{t+1}) \beta_{jt+1}}{\cancel{\alpha_{it-1}} \beta_{it-1}}$$

$$= \frac{\sum_k a_{ik} b_k(o_t) a_{kj} b_j(o_{t+1}) \beta_{jt+1}}{\beta_{it-1}}$$

7.3 Conditional independence

Consider the hidden Markov model (HMM) shown below, with hidden states S_t and observations O_t for times $t \in \{1, 2, \ldots, T\}$. State whether the following statements of conditional independence are true or false.

F	$P(S_t \mid S_{t-1}) = P(S_t \mid S_{t-1}, S_{t+1})$
T	$P(S_t \mid S_{t-1}) = P(S_t \mid S_{t-1}, O_{t-1})$
F	$P(S_t \mid S_{t-1}) = P(S_t \mid S_{t-1}, O_t)$
F	$P(S_t \mid O_{t-1}) = P(S_t \mid O_1, O_2, \ldots, O_{t-1})$
T	$P(O_t \mid S_{t-1}) = P(O_t \mid S_{t-1}, O_{t-1})$
F	$P(O_t \mid O_{t-1}) = P(O_t \mid O_1, O_2, \ldots, O_{t-1})$
T	$P(S_2, S_3, \ldots, S_T \mid S_1) = \prod_{t=2}^{T} P(S_t \mid S_{t-1})$
T	$P(S_1, S_2, \ldots, S_{T-1} \mid S_T) = \prod_{t=1}^{T-1} P(S_t \mid S_{t+1})$
F	$P(S_1, S_2, \ldots, S_T \mid O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(S_t \mid O_t)$
F	$P(S_1, S_2, \ldots, S_T, O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(S_t, O_t)$
T	$P(O_1, O_2, \ldots, O_T \mid S_1, S_2, \ldots, S_T) = \prod_{t=1}^{T} P(O_t \mid S_t)$
T	$P(O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(O_t \mid O_1, \ldots, O_{t-1})$



7.4 Belief updating

(a) Discrete HMMs

$$q_{jt} = P(S_t=j \mid O_1, O_2, \dots, O_t)$$

$$\overset{\text{Bayes}}{=} \frac{P(S_t=j \mid O_1,\dots,O_{t-1})P(O_t \mid S_t=j, O_1,\dots,O_{t-1})}{P(O_t \mid O_1,\dots,O_{t-1})}$$

Numerator: $P(S_t=j \mid O_1,\dots,O_{t-1})P(O_t \mid S_t=j, O_1,\dots,O_{t-1}) \overset{\text{C.I.}}{=} P(S_t=j \mid O_1,\dots,O_t)P(O_t \mid S_t=j)$

$$\overset{\text{marg}}{=} \sum_i P(S_{t-1}=i, S_t=j \mid O_1,\dots,O_t)P(O_t \mid S_t=j)$$

$$\overset{\text{P.R.}}{=} \sum_i P(S_{t-1}=i \mid O_1,\dots,O_t)P(S_t=j \mid S_{t-1}=i, O_1,\dots,O_t)\underbrace{P(O_t \mid S_t=j)}_{b_j(O_t)}$$

$$\overset{\text{C.I.}}{=} b_j(O_t)\sum_i \underbrace{P(S_{t-1}=i \mid O_1,\dots,O_t)}_{q_{i,t-1}}\underbrace{P(S_t=j \mid S_{t-1}=i)}_{a_{ij}}$$

$$= b_j(O_t)\sum_i q_{i,t-1}a_{ij}$$

$$\Rightarrow P(S_t=j \mid O_1,\dots,O_t) = \frac{b_j(O_t)\sum_i q_{i,t-1}a_{ij}}{Z}$$

Since $\sum_j P(S_t=j \mid O_1,\dots,O_t) = 1$, we have $Z = \sum_j b_j(O_t)\sum_i q_{i,t-1}a_{ij}$

$$\Rightarrow P(S_t=j \mid O_1,\dots,O_t) = \frac{b_j(O_t)\sum_i q_{i,t-1}a_{ij}}{\sum_j b_j(O_t)\sum_i q_{i,t-1}a_{ij}} = q_{jt}$$

QED

(b) Continuous dynamical systems

$$P(X_t \mid y_1,\dots,y_t) = \frac{P(X_t \mid y_1,\dots,y_{t-1})P(y_t \mid X_t, y_1,\dots,y_{t-1})}{Z}$$

Emission matrix becomes $P(y_t \mid X_t)$

Transition matrix becomes $P(X_t \mid X_{t-1})$

Numerator: $P(X_t \mid y_1,\dots,y_{t-1})P(y_t \mid X_t, y_1,\dots,y_{t-1}) \overset{\text{C.I.}}{=} P(X_t \mid y_1,\dots,y_{t-1})P(y_t \mid X_t)$

$$\overset{\text{marg}}{=} \int P(X_{t-1}, X_t \mid y_1,\dots,y_{t-1})P(y_t \mid X_t)\,dX_{t-1}$$

$$\overset{\text{CI \& P.R.}}{=} \int P(X_{t-1} \mid y_1,\dots,y_{t-1})P(X_t \mid X_{t-1})P(y_t \mid X_t)\,dX_{t-1}$$

$$\Rightarrow P(X_t \mid y_1,\dots,y_t) = \frac{P(y_t \mid X_t)\int P(X_{t-1} \mid y_1,\dots,y_{t-1})P(X_t \mid X_{t-1})\,dX_{t-1}}{Z}$$

Because we know $\int P(X_t \mid y_1,\dots,y_t)\,dX_t = 1$, we get

$$P(X_t \mid y_1,\dots,y_t) = \frac{P(y_t \mid X_t)\int dX_{t-1}P(X_{t-1} \mid y_1,\dots,y_{t-1})P(X_t \mid X_{t-1})}{\int dX_t\, P(y_t \mid X_t)\int dX_{t-1}P(X_{t-1} \mid y_1,\dots,y_{t-1})P(X_t \mid X_{t-1})}$$

QED

7.5 V-chain

(a) **Base case**

$$\begin{aligned}
P(Y_1 = j, o_1) &\overset{\text{marg}}{=} \sum_i P(X_1 = i, Y_1 = j, O_1) \\
&\overset{\text{P.R.}}{=} \sum_i \underbrace{P(X_1 = i) P(Y_1 = j \mid X_1 = i)}_{P(X_1, Y_1)} P(O_1 \mid X_1 = i, Y_1 = j) \\
&\overset{X_1 \,\&\, Y_1 \text{ marg. indep.}}{=} \sum_i P(X_1 = i) \underbrace{P(Y_1 = j)}_{\pi_j} \underbrace{P(O_1 \mid X_1 = i, Y_1 = j)}_{b_{ij}(O_1)} \\
&= \sum_i P(X_1 = i) \pi_j b_{ij}(O_1)
\end{aligned}$$

(b) **Forward algorithm**

$$\begin{aligned}
\alpha_{j,t+1} &= P(O_1, \ldots, O_{t+1}, Y_{t+1} = j) \\
&\overset{\text{marg}}{=} \sum_k P(Y_t = k, Y_{t+1} = j, O_1, \ldots O_{t+1}) \\
&\overset{\text{marg}}{=} \sum_k \sum_i P(Y_t = k, Y_{t+1} = j, X_{t+1} = i, O_1, \ldots, O_{t+1}) \\
&\overset{\text{P.R.}}{=} \sum_k \sum_i \underbrace{P(O_1, \ldots, O_t, Y_t = k)}_{\alpha_{kt}} P(X_{t+1} = i \mid Y_t = k, O_1, \ldots, O_t) P(Y_{t+1} = j \mid X_{t+1} = i, Y_t = k, O_1, \ldots, O_t) P(O_{t+1} \mid Y_t = k, Y_{t+1} = j, X_{t+1} = i, O_1, \ldots, O_t) \\
&\overset{\text{C.I.}}{=} \sum_k \sum_i \alpha_{kt} \underbrace{P(X_{t+1} = i \mid Y_t = k)}_{a_{ki}} \underbrace{P(Y_{t+1} = j)}_{\pi_j} \underbrace{P(O_{t+1} \mid Y_{t+1} = j, X_{t+1} = i)}_{b_{ij}(O_{t+1})} \\
&= \sum_k \sum_i \alpha_{kt} a_{ki} \pi_j b_{ij}(O_{t+1})
\end{aligned}$$

(c) **Likelihood**

$$\begin{aligned}
P(o_1, o_2, \ldots, o_T) &\overset{\text{marg.}}{=} \sum_j P(O_1, \ldots, O_T, Y_T = j) \\
&= \sum_j \alpha_{jT}
\end{aligned}$$

(d) **Complexity**

We have the length of $X_t = n_x$, length of $Y_t = n_y$ and the length of the sequence is T.

$$\alpha_{j,t+1} = \sum_k \sum_i \alpha_{kt} a_{ki} \pi_j b_{ij}(O_{t+1})$$

$$= \pi_j \sum_k \sum_i \alpha_{kt} a_{ki} b_{ij}(O_{t+1})$$

$$= \pi_j \sum_i b_{ij}(O_{t+1}) \sum_k \alpha_{kt} a_{ki}$$

We need to do this from $t=1$ up to $t=T \Rightarrow$ complexity $O(T)$

For the red part, we sum over k. $k \in [1, n_y] \Rightarrow$ Complexity $O(n_y)$

For the blue part, we sum over i. $i \in [1, n_x] \Rightarrow$ Complexity $O(n_x)$.

$\Rightarrow$ Total complexity: $\underline{\underline{O(T \cdot n_x \cdot n_y)}}$