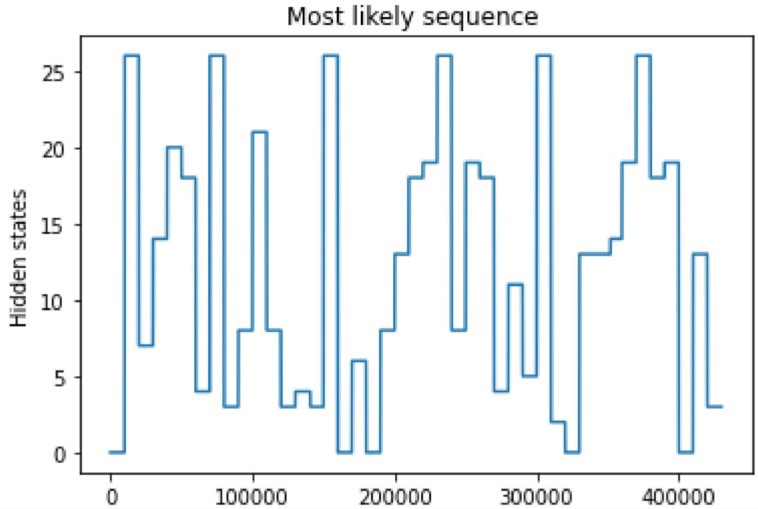
Viterbi Algorithm

```
!unzip /content/hw7files.zip
```

```
import numpy as np
N = 27
transition = np.loadtxt("/content/hw7files/transitionMatrix.txt" , dtype = "float")
observation = np.loadtxt("/content/hw7files/observations.txt" , dtype = "int").reshape(-1,
ISD = np.loadtxt("/content/hw7files/initialStateDistribution.txt" , dtype = "float").resha
eM = np.loadtxt("/content/hw7files/emissionMatrix.txt" , dtype = "float")
T = len(observation)
L_list = np.zeros((27,T))

for i in range(27):
  tmp1 = np.log(ISD[i])
  tmp2 = np.log(eM[i][observation[0]])
  L_list[i][0] = tmp1 + tmp2

final = np.zeros((27,T))
for i in range(27):
  L_list[i][0] = np.log(ISD[i]) + np.log(eM[i][observation[0]])

for t in range(1,T):
  for j in range(27):
    final[j][t] = np.argmax(L_list[:,t-1] + np.log(transition[:,j]))
    L_list[j][t] = np.amax(L_list[:,t-1] + np.log(transition[:,j])) + np.log(eM[j,observat



```

```
Viterbi = []
for i in (range(T-1,-1,-1)):
  if i != T-1:
    Viterbi.append(final[int(Viterbi[-1])][i+1])
  else:
    Viterbi.append(np.argmax(L_list[:,T-1]))

Viterbi = Viterbi[::-1]

import matplotlib.pyplot as plt
plt.plot(Viterbi)
plt.title('Most likely sequence ')
plt.xlabel('time')
plt.ylabel('Hidden states')
```

```
Text(0, 0.5, 'Hidden states')
```




```
foo = 'abcdefghijklmnopqrstuvwxyz '
dick = dict(enumerate(foo))
```

```
verify = ""
count = 0
for t in range(T-1):
  if Viterbi[t] == Viterbi[t+1]:
    count += 1
  else:
    count = 0
  if count > 8000:
    verify += (dick.get(Viterbi[t]))
    count = 0
print(verify)
```

```
a house divided against itself cannot stand
```

Hw 7 Andrew Ghafari CSE 250A - A59020215.

7.1 - Python Code.

7.2.

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

$$= \frac{P(S_{t+1}=j, S_t=i, o_1, o_2, \_\_\_ o_T) \quad ①}{P(S_t=i, o_1, o_2 \_\_\_ o_T) \quad ②}$$

for ①:

$= P(o_1 \_\_ o_t, S_t=i, o_{t+1} \_\_ o_T, S_{t+1}=j)$

$= P(o_1 \_ o_t, S_t=i) \times P(o_{t+1} \_\_ o_T, S_{t+1}=j \mid o_1 \_ o_t, S_t=i)$

$o_1 \_ o_t \perp\!\!\!\perp S_{t+1}, o_{t+1} \_ o_T$ given S_T di sep.

$= \alpha_{it} \cdot P(S_{t+1}, o_{t+1} \_ o_T \mid S_t)$

$= \frac{\alpha_{it}}{P(S_t)} \cdot P(S_{t+1}, o_{t+1} \_ o_T, S_t)$

$= \frac{\alpha_{it}}{P(S_t)} \cdot P(S_t, o_{t+1} \mid S_{t+1}, o_{t+2} \_\_ o_T) \times \underbrace{P(o_{t+2} \_ o_T \mid S_{t+1})}_{\beta_{j,t+1}} \cdot P(S_{t+1})$

$O_{t+2} — O_T \perp S_t, O_{t+1}$ given S_{t+1} di. sep.

$$= \frac{\alpha_{it}}{P(S_t)} \times P(S_t, O_{t+1} | S_{t+1}) \times \beta_{j,t+1} \times P(S_{t+1})$$

$$= \frac{\alpha_{it}}{P(S_t)} \times \frac{P(S_t, O_{t+1}, S_{t+1})}{P(\cancel{S_{t+1}})} \times \beta_{j,t+1} \times P(\cancel{S_{t+1}})$$

$$= \frac{\alpha_{it}}{\cancel{P(S_t)}} \times P(O_{t+1} / S_{t+1}) . P(S_{t+1} | S_t) . \cancel{P(S_t)} . \beta_{j,t+1}$$

$$= \alpha_{it} \times P(O_{t+1} | S_{t+1}) . P(S_{t+1} | S_t) . \beta_{j,t+1} .$$

$$= \alpha_{it} \times b_{j O_{t+1}} \cdot a_{ij} \cdot \beta_{j,t+1} .$$

<u>for (2):</u>

$P(S_t = i, O_1 — O_T)$.

$= P(O_{t+1} — O_T | S_t, O_1 — O_t) \times P(S_t, O_1 — O_t)$

$O_1 — O_t \perp O_{t+1} — O_T$ given S_t di. sep. $\to$ $= P(O_{t+1} — O_T | S_t) \times \alpha_{it}$

$= \beta_{it} . \alpha_{it} .$

$\Rightarrow$ we get $P(S_{t+1} = j | S_t = i, O_1 — O_T)$

$$= \frac{\cancel{\alpha_{it}} . a_{ij} . \beta_{j,t+1} . b_{j,O_{t+1}}}{\beta_{it} . \cancel{\alpha_{it}}} = \frac{a_{ij} . \beta_{j,t+1} . b_{j,O_{t+1}}}{\beta_{it}}$$

b) $P(S_t = i \mid S_{t+1} = j, O_1 \_\_ O_T)$

$$= \frac{P(S_t = i, S_{t+1} = j, O_1 \_\_\_ O_T) \quad (1)}{P(S_{t+1} = j, O_1 \_\_\_ O_T) \quad (2)}$$

(1) is same in part (a) $= a_{ij} \cdot \beta_{j,t+1} \cdot \alpha_{it} \cdot b_{j,O_{t+1}}$

(2) $= P(S_{t+1} = j, O_1 \_\_\_ O_T)$

$= P(O_{t+2} \_ O_T \mid S_{t+1}, O_1 \_ O_{t+1}) \cdot P(O_1 \_ O_{t+1}, S_{t+1})$

same reason as above $\rightarrow$ $= P(O_{t+2} \_ O_T \mid S_{t+1}) \cdot \alpha_{j,t+1}$

$= \beta_{j,t+1} \cdot \alpha_{j,t+1}$.

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

c) $P(S_{t-1} = i, S_t = k, S_{t+1} = j \mid O_1 \_ O_T)$

$$= \frac{P(S_{t-1} = i, S_t = k, S_{t+1} = j, O_1 \_\_\_ O_T) \quad (1)}{P(O_1 \_\_\_ O_T) \quad (2)}$$

(1): $P(S_t, S_{t+1}, O_t - O_T \mid S_{t-1}, O_1 - O_{t-1}) \cdot P(S_{t-1}, O_1, - O_{t-1})$

$O_1 - O_{t-1}$
$\perp S_t, S_{t+1}, O_t - O_T$
given S_{t-1}.
d-sep.

$= P(S_t, S_{t+1}, O_t - O_T \mid S_{t-1}) \cdot \alpha_{i\,t-1}$.

$= P(S_t, S_{t+1}, S_{t-1}, O_t - O_T) \cdot \dfrac{\alpha_{i\,t-1}}{P(S_{t-1})}$.

$= P(S_{t-1} = i, S_t, O_t, O_{t+1} \mid S_{t+1}, O_{t+2} - O_T) \cdot P(S_{t+1}, O_{t+2} - O_T) \cdot \dfrac{\alpha_{i\,t-1}}{P(S_{t-1})}$

$= \dfrac{\alpha_{i\,t-1}}{P(S_{t-1})} \cdot P(S_{t-1}, S_t, O_t, O_{t+1} \mid S_{t+1}) \cdot P(O_{t+2} - O_T \mid S_{t+1}) \cdot P(S_{t+1})$

$= \dfrac{\alpha_{i\,t-1}}{P(S_{t-1})} \cdot P(S_{t-1}, S_t, S_{t+1}, O_t, O_{t+1}) \cdot \beta_{j,t+1} \cdot \dfrac{\cancel{P(S_{t+1})}}{\cancel{P(S_{t+1})}}$

$= \dfrac{\alpha_{i\,t-1}}{\cancel{P(S_{t-1})}} \cdot P(S_t \mid S_{t-1}) \cdot P(S_{t+1} \mid S_t) \cdot P(O_t \mid S_t) \cdot P(O_{t+1} \mid S_{t+1}) \cdot \cancel{P(S_{t-1})} \cdot \beta_{j,t+1}$

$= \alpha_{i\,t-1} \cdot a_{ik} \cdot a_{kj} \cdot b_k(O_t) \cdot b_j(O_{t+1}) \cdot \beta_{j,t+1}$

(2) $= P(O_1 - O_T)$

$= \sum_k P(S_t = k, O_1 - O_T)$

$= \sum_k P(O_1 - O_t, S_t = k) \cdot P(O_{t+1} - O_T \mid S_t = k)$

$= \sum_k \alpha_{kt} \cdot \beta_{kt}$.

$$\Rightarrow (C) = \frac{\alpha_{i,t-1} \cdot a_{ik} \cdot a_{kj} \cdot b_{kO_t} \cdot b_j(O_{t+1}) \cdot \beta_{j,t+1}}{\sum_k \alpha_{k,t} \cdot \beta_{k,t}}$$

d) $P(S_{t+1} = j \mid S_{t-1} = i, O_1 \ldots O_T)$

$$= \frac{\sum_k P(S_{t+1}, S_{t-1}, S_t = k, O_1 \ldots O_T)}{P(S_{t-1}, O_1 \ldots O_T)}$$ ← same as we did for $P(S_t \ldots O_T)$ & $P(S_{t+1}, O_1 \ldots O_T)$.

$$= \frac{\sum_k a_{ik} \cdot a_{kj} \cdot b_k(O_T) \cdot \beta_{j,t+1} \cdot b_j(O_{t+1}) \cdot \cancel{\alpha_{i,t-1}}}{\beta_{j,t-1} \cdot \cancel{\alpha_{i,t-1}}}$$

$$= \frac{\sum_k a_{ik} \cdot a_{kj} \cdot b_k(O_T) \cdot \beta_{j,t+1} \cdot b_j(O_{t+1})}{\beta_{j,t-1}}$$

7.3)

1) False 3) False 5) True 7) True 9) False 11) True.

2) True 4) False 6) ~~True~~ False 8) True 10) False 12) True

7.4

a) $q_{jt} = P(S_t = j \mid O_1 \_\_ O_t)$

$$= \frac{P(S_t = j, O_1 \_\_ O_t)}{P(O_1 \_\_ O_t)}$$ Product Rule

$$\overset{①}{=} \frac{\sum_i P(S_t = j, S_{t-1} = i, O_1 \_\_ O_t)}{② \sum_i \sum_j P(S_t = j, S_{t-1} = i, O_1 \_\_ O_t)}$$ Marginalization

For ①:

$$= \frac{① \sum_i P(S_t, S_{t-1}, O_t \mid O_1 \_\_ O_{t-1})}{② \sum_i \sum_j P(S_t, S_{t-1}, O_t \mid O_1 \_\_ O_{t-1})}$$

$P(S_t = j, S_{t-1} = i, O_t \mid O_1 \_\_ O_{t-1})$

conditional independence by di

$= P(S_{t-1} = i \mid O_1 - O_{t-1}) \cdot P(S_t = j \mid S_{t-1} = i) \cdot P(O_t \mid S_t = j)$

$= b_j(O_t) \cdot a_{ij} \, q_{it-1}$

The same thing happens for numerator.

$\Rightarrow$ we get $$\frac{\sum_i b_j(o_t) \cdot a_{ij} \cdot q_{it-1}}{\sum_i \sum_j b_j(o_t) \, a_{ij} \, q_{it-1}}$$

$$= \frac{b_j(o_t) \sum a_{ij} q_{it-1}}{Z_t}$$ $\Rightarrow$ Hypo. proven.

f) We can replace Σ w/ $\int$ and we'll have.

$$P(x_t \mid y_1 - y_t) = \frac{P(x_t, y_1 - y_t)}{P(y_1 - y_t)}$$

$$= \frac{\int dx_{t-1}\, P(x_t, x_{t-1}, y_1 - y_t)}{\iint dx_t\, dx_{t-1}\, P(x_t, x_{t-1}, y_1 - y_t)}$$

$$= \frac{\int dx_{t-1}\, P(y_t \mid x_{t-1}, y_1 - y_t) \cdot P(x_t \mid x_{t-1}, y_1 - y_t) \cdot P(x_{t-1} \mid y_1 - y_t)}{\int dx_t \int dx_{t-1}\, P(y_t \mid x_t, x_{t-1}, y_1 - y_t) \cdot P(x_t \mid x_{t-1}, y_1 - y_t) \cdot P(x_{t-1} \mid y_1 - y_t)}$$

using CI

$$= \frac{\int dx_{t-1}\, P(y_t \mid x_t) \cdot P(x_t \mid x_{t-1}) \cdot P(x_{t-1} \mid y_1 - y_t)}{\int dx_t \int dx_{t-1}\, P(y_t \mid x_t) \cdot P(x_t \mid x_{t-1}) \cdot P(x_{t-1} \mid y_1 - y_t)}.$$

$$= \frac{P(y_t \mid x_t) \int dx_{t-1}\, P(x_t \mid x_{t-1}) \cdot P(x_{t-1} \mid y_1 - y_t)}{\int dx_t\, P(y_t \mid x_t) \cdot \int dx_{t-1} \cdot P(x_t \mid x_{t-1}) \cdot P(x_{t-1} \mid y_1 - y_t)}$$

$$= \frac{P(y_t|x_t) \cdot \int dx_{t-1}\, P(x_t|x_{t-1}) \cdot P(x_{t-1}|y_1 \text{——} y_t)}{Z_t}$$

* The reason why Gaussian Random Variables are much easier for real time updating is.

→ one reason is it easier to integrate GRV distribution functions and it is more tractable.

. We can also take advantage of some math properties like if $P(\vec{X})$ & $P(\vec{Y})$ are GRV, so $P(a\vec{X}, b\vec{Y})$ is also a GRV. if a, b are scalar coefficients.

. Conditional probabilities w/ real values of PDF may not have tractable integrals / or it might be hard to calculate it.

. for GRV, $P(\vec{X})$ & $P(\vec{Y})$, marginals are $P(X_i)$ & $P(Y_i)$ and conditionals are $P(X_i|Y_i)$ & $P(Y_i|X_i)$ ⇒ easy to compute / tractable.