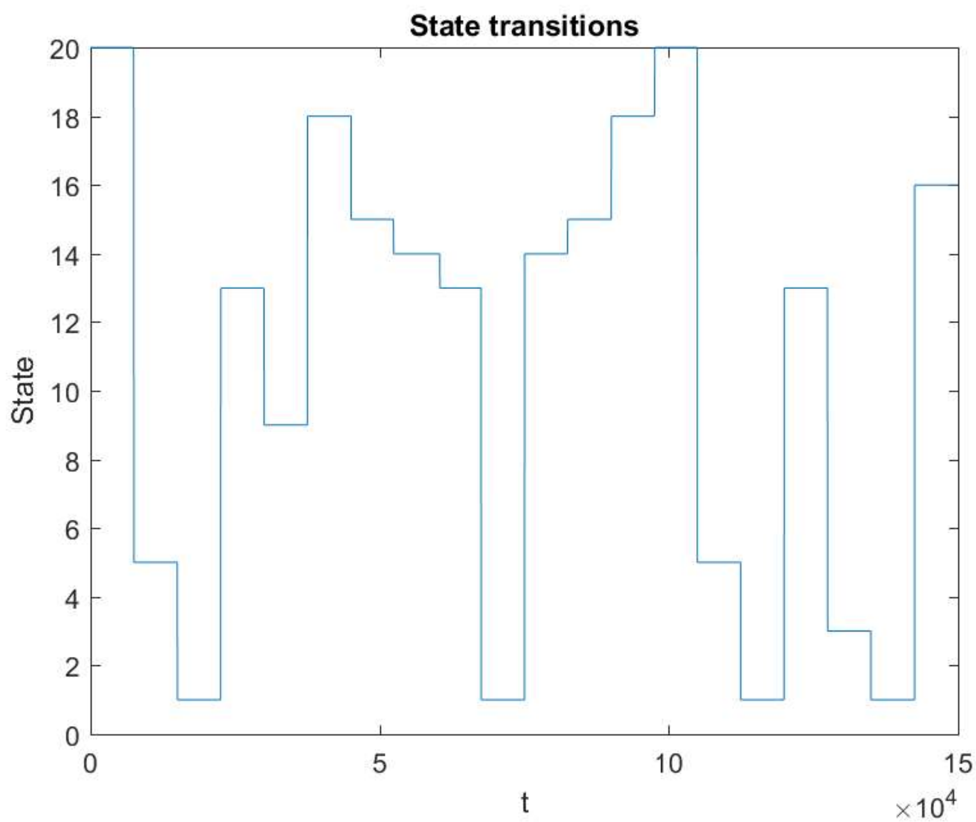
```
% Problem 6.1 - Viterbi Algorithm
clear; clc; close all

load('emissionMatrix.mat')
load('initialStateDistribution.mat')
load('observations.mat')
load('transitionMatrix.mat')

% Initialize
transitionMatrix(:,27) = [];
A = transitionMatrix;
B = emissionMatrix;
O = observations;
T = length(O);
n = length(A);
max_P = zeros(1,T);
P = zeros(n,T);
P0 = initialStateDistribution;

tree = zeros(n,T);
path = zeros(1,T);

% Calculate the first probability vector
P(:,1) = (A*P0).*B(:,O(1)+1)/sum((A*P0).*B(:,O(1)+1));

% Using probabilities, construct a matrix containing indices of connected
% node
for t = 2:T
    for j = 1:n % j is the index of the 2nd column
        best_val = 0;
        for i = 1:n % i is the index of the 1st column
            new = P(i,t-1)*A(i,j);
            if new > best_val
                best_val = new;
                tree(j,t) = i;
                P(j,t) = new*B(j,O(t)+1);
            end
        end
    end
    P(:,t) = P(:,t)/sum(P(:,t));
end

% Construct path backwards
[~,path(T)] = max(P(:,T));
for t = T-1:-1:1
    path(t) = tree(path(t+1),t+1);
end
plot(path)
title('State transitions')
xlabel('t')
ylabel('State')
```

```
% print out message converted into a string
% 97 is a on ASCII table
str = char(path(1)+96);
for t = 2:T
    if path(t) ~= path(t-1)
        str = [str char(path(t)+96)];
    end
end
disp(['The hidden message is: ' str])
```



State transitions
State
t
$\times 10^4$
0
5
10
15
2
4
6
8
10
12
14
16
18
20

6.2

a) $P(S_{t+1}=j \mid S_t=i, O_{1:T}) = \dfrac{P(S_{t+1}=j, O_{1:T} \mid S_t=i)}{P(O_{1:T} \mid S_t=i)} = \dfrac{P(S_{t+1}=j, O_{1:T} \mid S_t=i)}{P(O_{1:t}, O_{t+1:T} \mid S_t=i)}$

$$= \frac{P(O_{1:T} \mid S_t=i, S_{t+1}=j)\, P(S_{t+1}=j \mid S_t=i)}{P(O_{1:t} \mid S_t=i)\, P(O_{t+1:T} \mid S_t=i)}$$

$$= \frac{\cancel{P(O_{1:t} \mid S_t=i)}\, P(O_{t+1:T} \mid S_{t+1}=j)\, P(S_{t+1}=j \mid S_t=i)}{\cancel{P(O_{1:t} \mid S_t=i)}\, P(O_{t+1:T} \mid S_t=i)}$$

$$= \boxed{\frac{\beta_{jt}\, a_{ij}}{\beta_{it}}}$$

b) $P(S_t=i \mid S_{t+1}=j, O_{1:T}) = \dfrac{P(S_t=i, O_{1:T} \mid S_{t+1}=j)}{P(O_{1:T} \mid S_{t+1}=j)} = \dfrac{P(S_t=i, O_{1:T} \mid S_{t+1}=j)}{P(O_{1:t+1} \mid S_{t+1}=j)\, P(O_{t+2:T} \mid S_{t+1}=j)}$

$$= \frac{P(O_{1:T} \mid S_t=i, S_{t+1}=j)\, P(S_t=i \mid S_{t+1}=j) \cdot P(S_{t+1}=j)}{P(O_{1:t+1} \mid S_{t+1}=j)\, P(O_{t+2:T} \mid S_{t+1}=j) \cdot P(S_{t+1}=j)}$$

$$= \frac{P(O_{1:t} \mid S_t=i)\, P(O_{t+1:T} \mid S_{t+1}=j)\, P(S_t=i, S_{t+1}=j)}{P(O_{1:t+1}, S_{t+1})\, P(O_{t+2:T} \mid S_{t+1}=j)}$$

$$= \frac{P(O_{1:t}, S_t=i)\, P(O_{t+1:T} \mid S_{t+1}=j)\, P(S_{t+1}=j \mid S_t=i)}{P(O_{1:t+1}, S_{t+1}=j)\, P(O_{t+2:T} \mid S_{t+1}=j)} = \boxed{\frac{\alpha_{it}\, b_{j,o_t}\, \beta_{j,t+1}\, a_{ij}}{\alpha_{j,t+1}\, \beta_{j,t+1}}}$$

c) $P(S_{t-1}=i, S_t=j, S_{t+1}=k \mid O_{1:T}) = P(S_{t+1}=k \mid S_t=j, S_{t-1}=i, O_{1:T})\, P(S_t=j, S_{t-1}=i \mid O_{1:T})$

$$= P(S_{t+1}=k \mid S_t=j, O_{t+1:T})\, P(S_t=j \mid S_{t-1}=i, O_{1:T})\, P(S_{t-1}=i \mid O_{1:T})$$

$$= P(S_{t+1}=k \mid S_t=j, O_{t+1:T})\, P(S_t=j \mid S_{t-1}=i, O_{t:T}) \sum_h P(S_{t-1}=i \mid S_{t-2}=h, O_{1:T})\, P(S_{t-2}=h \mid O_{1:T})$$

$$= P(S_{t+1}=k \mid S_t=j, O_{t+1:T})\, P(S_t=j \mid S_{t-1}=i, O_{t:T}) \sum_h P(S_{t-1}=i \mid S_{t-2}=h, O_{t-1:T})\, P(S_{t-2}=h \mid O_{1:T})$$

$$= \boxed{\frac{\beta_{kt}\, a_{jk}}{\beta_{jt}} \cdot \frac{\beta_{j,t-1}\, a_{ij}}{\beta_{i,t-1}} \cdot \sum_h \frac{\beta_{i,t-2}\, a_{hi}}{\beta_{h,t-2}} \cdot \sum \ldots}$$

d) $P(S_t=k \mid S_{t+1}=i, O_{1:T}) = \dfrac{P(S_{t-1}=k, O_{1:T} \mid S_{t+1}=i)}{P(O_{1:t+1} \mid S_{t+1}=j)\, P(O_{t+2:T} \mid S_{t+1}=i)} =$

6.3

a) $$P(S_t = j \mid o_{1:t}) = \sum_i P(S_t = j \mid S_{t-1} = i, o_{1:t}) P(S_{t-1} = i \mid o_{1:t})$$ * conditioning on S_{t-1}

$$= \sum_i \frac{P(S_t = j, o_{1:t} \mid S_{t-1} = i)}{P(o_{1:t} \mid S_{t-1} = i)} P(S_{t-1} = i \mid o_{1:t})$$ * product rule

$$= \sum_i \frac{P(S_t = j, \mid S_{t-1} = i) P(o_{1:t-1} \mid S_{t-1} = i) \cdot P(S_{t-1} = i \mid o_{1:t})}{P(o_{1:t-1} \mid S_{t-1} = i) P(o_t \mid S_{t-1} = i)}$$ *split up observations

$$= \boxed{\frac{\sum_i a_{ij}\, b_j(o_t)\, q_{i,t-1}}{\sum_{ij} b_j(o_t)\, a_{ij}\, q_{i,t-1}}}$$ * substitute

b) $$P(x_t \mid y_{1:t}) = \int P(x_t \mid dx_{t-1}, y_{1:t}) P(dx_{t-1} \mid y_{1:t})$$

$$= \int \frac{P(x_t, y_{1:t} \mid dx_{t-1}) P(dx_{t-1} \mid y_{1:t})}{P(y_{1:t} \mid dx_{t-1})}$$

$$= \int \frac{P(x_t \mid dx_{t-1}) P(y_{1:t-1} \mid dx_{t-1}) P(dx_{t-1} \mid y_{1:t})}{P(y_{1:t-1} \mid dx_{t-1}) P(y_t \mid dx_{t-1})}$$

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

6.4

a) $P(X_t \mid X_{t-1}) = \sum_{i=1}^{n} P(X_t \mid S_t, X_{t-1}) P(S_t \mid X_{t-1})$

$= \sum_{i=1}^{n} P(X_t \mid S_t) P(S_t \mid X_{t-1})$

mixture of n Gaussians
univariate

b) $P(X_1, X_T \mid S_1, S_T) = P(X_1 \mid S_1) P(X_T \mid S_T)$

2d Gaussian

c) $P(X_1, \dots, X_t) = P(X_1 \mid X_2, \dots, X_t) P(X_2 \mid X_3, \dots, X_t) \dots P(X_{t-1} \mid X_t)$

t dimensional Gaussian, each dimension is a mixture of n
multivariate

d) $P(X_t \mid X_{t-1}, X_{t+1}) = \sum_{i=1}^{n} P(X_t \mid S_t = i, X_{t-1}, X_{t+1}) P(S_t = i \mid X_{t-1}, X_{t+1})$

mixture of n Gaussians
univariate

e) $P(X_t) = \sum_{i=1}^{n} P(X_t \mid S_t = i) P(S_t = i)$

mixture of n Gaussians
univariate

f) $P(X_1, \dots, X_T \mid S_1, \dots, S_T) = P(X_1 \mid S_1) \dots P(X_T \mid S_T)$

T dimensional Gaussian
multivariate

6.5

$$(y) \longrightarrow (x)$$

$$P(y=i)\ \pi_i \qquad P(\vec{x}|y=i) = (2\pi)^{-\frac{d}{2}} |\Sigma_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}-\vec{\mu}_i)}$$

a) $$P(y=1|\vec{x}) = \frac{P(\vec{x}|y=1)\,P(y=1)}{P(\vec{x})} = \frac{P(\vec{x}|y=1)\,P(y=1)}{\sum_i P(\vec{x}|y=i)\,P(y=i)}$$

$$= \frac{(2\pi)^{-\frac{d}{2}} |\Sigma_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma_1^{-1} (\vec{x}-\vec{\mu}_1)} \cdot \pi_1}{\text{same term} + (2\pi)^{-\frac{d}{2}} |\Sigma_0|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x}-\vec{\mu}_0)} \cdot \pi_0}$$

$$\boxed{= \frac{\pi_1 |\Sigma_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma_1^{-1} (\vec{x}-\vec{\mu}_1)}}{\pi_1 |\Sigma_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma_1^{-1} (\vec{x}-\vec{\mu}_1)} + \pi_0 |\Sigma_0|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x}-\vec{\mu}_0)}}}$$

b) If $\Sigma_1 = \Sigma_0 = \Sigma$, then

$$P(y=1|\vec{x}) = \frac{\pi_1 \cancel{|\Sigma|^{-\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_1)}}{\pi_1 \cancel{|\Sigma|^{-\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma^{-1} (\vec{x}-\mu_1)} + \pi_0 \cancel{|\Sigma|^{-\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_0)}}$$

* divide each term by numerator

$$= \frac{1}{1 + \frac{\pi_0}{\pi_1} \frac{e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_0)}}{e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_1)}}} = \frac{1}{1 + e^{\pi_0 - \frac{1}{2}(\vec{x}-\vec{\mu}_0)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_0) - \pi_1 + \frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_1)}}$$

* Now simplifying the exponential term, Z,

$$Z = \pi_0 - \pi_1 - \frac{1}{2}\left(\cancel{\vec{x}^T \Sigma^{-1} \vec{x}} - {}^T\Sigma^{-1}\vec{\mu}_0 - \vec{\mu}_0^T \Sigma^{-1} \vec{x} + \vec{\mu}_0^T \Sigma^{-1} \mu_0\right) + \frac{1}{2}\left(\cancel{\vec{x}^T \Sigma^{-1} \vec{x}} - \vec{x}^T \Sigma^{-1} \vec{\mu}_1 - \vec{\mu}_1^T \Sigma^{-1} \vec{x} + \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1\right.$$

$$Z = \pi_0 - \pi_1 + \frac{1}{2} \vec{\mu}_0^T \Sigma^{-T} \vec{x} + \frac{1}{2} \vec{\mu}_0 \Sigma^{-1} \vec{x} - \frac{1}{2} \vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0 - \frac{1}{2} \vec{\mu}_1^T \Sigma^{-T} \vec{x} - \frac{1}{2} \vec{\mu}_1^T \Sigma^{-1} \vec{x} + \frac{1}{2} \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1$$

So $P(y=1|\vec{x}) = \sigma(\vec{w} \cdot \vec{x} + b)$ where

$$\vec{w} = -\frac{1}{2}\left(\vec{\mu}_0^T \Sigma^{-T} + \vec{\mu}_0^T \Sigma^{-1} - \vec{\mu}_1^T \Sigma^{-T} - \vec{\mu}_1^T \Sigma^{-1}\right)$$

$$b = -\pi_0 + \pi_1 + \frac{1}{2} \vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0 - \frac{1}{2} \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1$$

6.5

c) When does $\dfrac{P(y=1|\vec{x})}{P(y=0|\vec{x})} = n$, $n > 0$

$$\frac{P(y=1|\vec{x})}{P(y=0|\vec{x})} = \frac{\dfrac{1}{1+e^{-\vec{w}\cdot\vec{x}-b}}}{1-\dfrac{1}{1+e^{-\vec{w}\cdot\vec{x}-b}}} = \frac{\dfrac{1}{\cancel{1+e^{-\vec{w}\cdot\vec{x}-b}}}}{\dfrac{+e^{-\vec{w}\cdot\vec{x}-b}}{\cancel{1+e^{-\vec{w}\cdot\vec{x}-b}}}}$$

$$\frac{P(y=1|\vec{x})}{P(y=0|\vec{x})} = e^{\vec{w}\cdot\vec{x}+b} = n$$

$$\boxed{\vec{w}\,\vec{x}+b = \ln(n)}$$