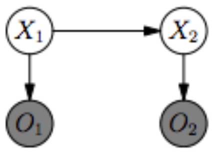
HMM QUESTIONS

1. Given below is a HMM: (Forward probability question)



X_1	$\Pr(X_1)$
0	0.3
1	0.7

X_t	X_{t+1}	$\Pr(X_{t+1}\|X_t)$
0	0	0.4
0	1	0.6
1	0	0.8
1	1	0.2

X_t	O_t	$\Pr(O_t\|X_t)$
0	A	0.9
0	B	0.1
1	A	0.5
1	B	0.5

You observe that the value of O_1 =A and O_2 = B.

a. Calculate the probability Pr(X_1=0, O_1=A) i.e. α_1^0

b. Calculate the probability Pr(X_1=1, O_1=A) i.e. α_1^1

c. Use the forward algorithm to compute the probability distribution of:
Pr(X_2 = 0, O_1 =A, O_2 = B) i.e. α_2^0
Hint: Since nothing is said about X1, you will have to take the two different cases – X1=0 and X1=1

d. Use the forward algorithm to compute the probability distribution of:
Pr(X_2 = 1, O_1 =A, O_2 = B) i.e. α_2^1

e. What is the probability of the observables: P(O_1 = A, O_2 = B) Hint: It is $\alpha_2^0 + \alpha_2^1$

f. Compute the value of Pr(X1 = 0, O_1 = A, O_2 = B) Hint: It is $\alpha_1^0\beta_1^0$

2. You are given the following HMM for a bioinformatics project: (Joint probability)

	0	S_1	S_2	S_3	S_4	S_5	S_6	f	A	C	G	T
0	0	1	0	0	0	0	0	0				
S_1	0	0	1	0	0	0	0	0	0.5	0.3	0	0.2
S_2	0	0	0	0.3	0	0.7	0	0	0.1	0.1	0.2	0.6
S_3	0	0	0	0	1	0	0	0	0.2	0	0.1	0.7
S_4	0	0	0	0	0	0	0	1	0.1	0.3	0.4	0.2
S_5	0	0	0	0	0	0	1	0	0.1	0.3	0.3	0.3
S_6	0	0	0	0	0	0	0	1	0.2	0.3	0	0.5

This indicates that there are 6 states and one final state. The starting probability of S1 is 1, indicating that the model always starts at S1.
The states emit the four genome characters A, C, T, G and their emission probabilities are given on the right hand side.

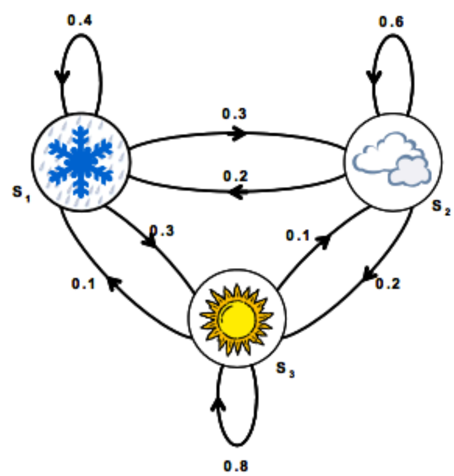
a. You see the following sequence of characters (observations)
O = ACTA
Also, you know the value of first two states:
$Q = S_1S_2$
Compute the following:
P(O Q)

b. You observe the same sequence as in part a, but this time you have knowledge about the third and fourth states: $Q_{34} = S_3S_4$
Compute $P(O \mid Q_{34})$ using HMM approach.

3. Joint Probability

- Consider a simple three-state Markov model of the weather
- Any given day, the weather can be described as being
  - State 1: precipitation (rain or snow)
  - State 2: cloudy
  - State 3: sunny
- Transitions between states are described by the transition matrix

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$



** Note: The questions below are about Markov Model and not Hidden Markov Model **

a. Given that the weather on day t=1 is sunny, what is the probability that the weather for the next 7 days will be “sun, sun, rain, rain, sun, clouds, sun”? You can assume a first order Markov model for the states.

b. What is the probability that the weather stays in the <u>same known state Si</u> for exactly T consecutive days?

c. What is the probability that the weather stays in <u>any state Si</u> for exactly T consecutive days?

4. Suppose you have two coins (i.e. states) that are both biased. You can only observe the outcome from both the dice and not which one was used for each outcome. You would like to create a sequential Hidden Markov Model for predicting which coin was used at each step. How many **parameters (which could be vector or scalar)** are needed to describe this model and how many minimum variables are needed for each parameter. (Parameter estimation question)

5. Consider a village where all villagers are either healthy or have a fever and only the village doctor can determine whether each has a fever. The doctor diagnoses fever by asking patients how they feel. The villagers may only answer that they feel normal, dizzy, or cold.
The complete model can be explained as: (Viterbi Question)

```
states = ('Healthy', 'Fever')

observations = ('normal', 'cold', 'dizzy')

start_probability = {'Healthy': 0.6, 'Fever': 0.4}

transition_probability = {
   'Healthy' : {'Healthy': 0.7, 'Fever': 0.3},
   'Fever' : {'Healthy': 0.4, 'Fever': 0.6}
   }

emission_probability = {
   'Healthy' : {'normal': 0.5, 'cold': 0.4, 'dizzy': 0.1},
   'Fever' : {'normal': 0.1, 'cold': 0.3, 'dizzy': 0.6}
   }
```

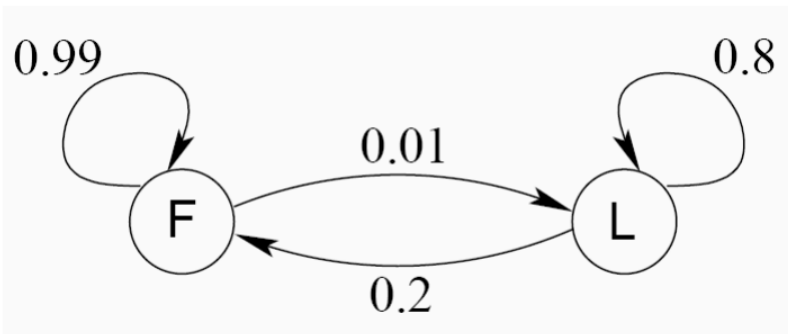
You get three observations in three consecutive days as:

Normal, Cold, Dizzy.

Calculate the likely states for each day.

6. A casino has two types of dice in stock – fair dice and loaded dice. The probability of picking a loaded dice is 0.01 i.e. 1% of dice are loaded. A loaded dice outputs the number 6 with a 50% probability and the other numbers with 10% probability each. You observe three outcomes of one game (using the same dice) and all 3 are sixes. What is the probability that the dice was loaded? (Conditional probability question)

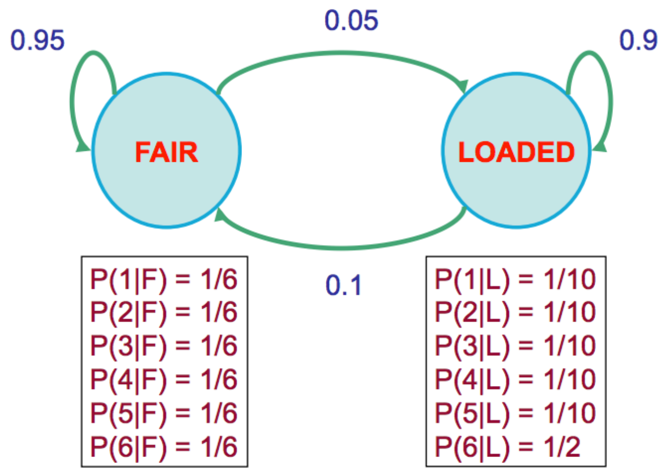
7. Using the same scenario as shown above, but now the transition probability between fair and loaded dice is given as: (Joint probability question)



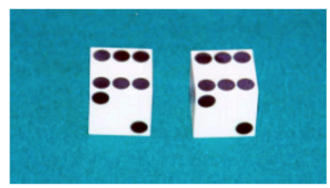
What is the probability of the following: P({326}, {FFL}) i.e. first observation is 3 using fair dice and so on.

8. Consider again the dishonest casino model: (Joint probability question)

The dishonest casino model



What is the probability of a sequence of rolls



x = 1, 2, 1, 5, 6, 2, 1, 6, 2, 4

and the parse

π = Fair, Fair, Fair, Fair, Fair, Fair, Fair, Fair, Fair, Fair?

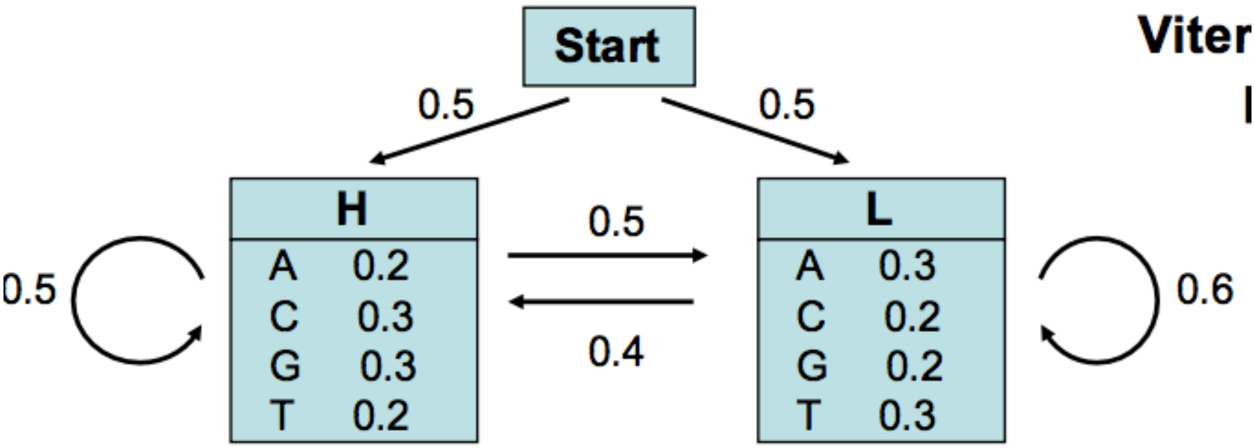
(say initial probs $a_{0,\text{Fair}} = ½$, $a_{0,\text{Loaded}} = ½$)

9. Decoding problem: (Viterbi algorithm)

Given the following observations: Observed Sequence: $x = 1,2,1,6,6$

Find the most likely path using Viterbi algorithm (as explained in class)

10. In human genome, there are two regions (states in HMM) – high GC content and low GC. The emissions are four possible characters – A, C, T, or G. The HMM is specified as below:



You observe the following sequence:

G G C $\boxed{A}$ C T G A A

a. Write down the Viterbi algorithm the transitions up to the character A (marked in red) to find the most probable states.

b. In the second scenario, find out the probability that the sequence S=GGCA was generated using the above HMM using the forward algorithm.

c. What is the probability that the state of the 4^{th} sequence is H?