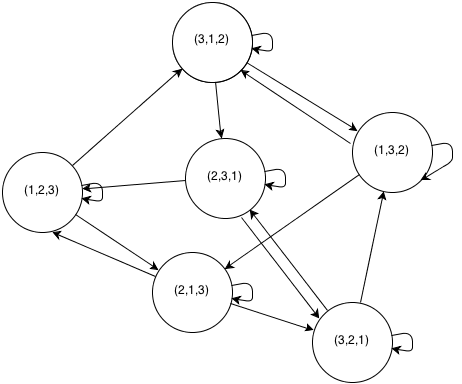
**Question 1:**

1. **Write down the Markov Chain, including the set of states, transition probabilities, and initial distribution, for the case of N = 3.**

**Solution:**

**Markov Chain:**



**Set of States:**

{(1,2,3), (2,1,3), (3,1,2), (1,3,2), (2,3,1), (3,2,1)}

**Transition Probabilities:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (1,2,3) | (2,1,3) | (3,1,2) | (3,2,1) | (1,3,2) | (2,3,1) |
| (1,2,3) |  |  |  |  |  |  |
| (2,1,3) |  |  |  |  |  |  |
| (3,1,2) |  |  |  |  |  |  |
| (3,2,1) |  |  |  |  |  |  |
| (1,3,2) |  |  |  |  |  |  |
| (2,3,1) |  |  |  |  |  |  |

**Initial Distribution:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (1,2,3) | (2,1,3) | (3,1,2) | (3,2,1) | (1,3,2) | (2,3,1) |
| 1 | 0 | 0 | 0 | 0 | 0 |

1. **Prove that the Markov Chain is ergodic and hence a stationary distribution exists.**

**Solution:**

A Markov chain is ergodic if it is possible to move from every state to every state (not necessary in one move) i.e. **irreducible** and **aperiodic**.

* 1. Over here each state can be reached other state in max N=2. Therefore the above Markov Chain is **irreducible**.
  2. The period of a state ‘i’ is given by

As we can see in the transition diagram every state loops to itself with N=1 moves. Therefore the probability >0 . Therefore =1 for all states and thus the above Markov Chain is **aperiodic**.

Thus we can say that the above Markov Chain is Ergodic and hence a stationary distribution exists.

1. **What is the stationary distribution in the case of N = 3? (You can figure this out by hand, write a program, or use mathematical software like Matlab or Mathematica. In any case, please submit your work (your hand calculations, your program, or your interactive Matlab session.)**

**Solution:**

Here is the MATLAB code to calculate the stationary distribution:

B = [1/6,1/3,1/2,0,0,0;

1/18,1/9,0,1/6,0,0;

0,0,1/4,0,1/12,1/6;

0,0,0,1/12,1/36,1/18;

0,1/36,1/24,0,1/72,0;

1/36,0,0,1/12,0,1/18];

[v d] = eigs(B',1);

t = v/sum(v);

Which gives the following vector:

t =

0.0655

0.1026

0.3645

0.1355

0.1011

0.2308

1. **What is the stationary distribution for N in general?**

**Solution:**

Let n = N! be the number of states. Let each of the states be {X1,X2,…,Xn}. Let xij denote the transition probability from Xi to Xj . The transition matrix would look something like:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | … | … | Xn |
| X1 | x11 | X12 | … | … | X1n |
| X2 | x21 | X22 | … | … | X2n |
| : | : | : | … | … | : |
| : | : | : | … | … | : |
| Xn | xn1 | xn2 | … | … | xnn |

Using the formula BTπ=π,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x11 | x21 | … | … | xn1 |
| x12 | X22 | … | … | X2n |
| : | : | … | … | : |
| : | : | … | … | : |
| x1n | x2n | … | … | xnn |

|  |
| --- |
| P1 |
| P2 |
| : |
| : |
| Pn |

**\* =**

|  |
| --- |
| P1 |
| P2 |
| : |
| : |
| Pn |

Converting to equation form,

x11P1 + x21P2 + … + xn1Pn = P1

:

:

x1nP1 + x2nP2 + …+ xnnPn = Pn

where, P1+P2+ … + Pn = 1 and x11 + x12 + … + x1n = 1

Solution to this set of equation can be computed using eigen value decomposition as before.

For n =1, there is only 1 state,hence P1 = P1

**Answer 2**

**(a)** We have used simple variable elimination to compute the conditional probability distributions in the E-step. Most of the variables are seen, and hence running variable elimination is intuitive and easy to compute. The only situation is where more computation was involved was when more than two variables were missing in any exemplar. But this situation was also handled efficiently as each variable was binary valued.

**(b)** We do get different learned probability distributions when (i) the parameters at t=0 are initialozed at random and (ii) the parameters are set using intuition. We show the results of these two cases in the following two tables.

All the distribution tables have been written in the text file attached with this document, named “distributions.txt”.

The probabilites that were set using “intuition” are present in the code under the function named “intuition”.

**(c)** Using random values for initializing the distribution tables, the accuracy comes to 80.6% using the method mentioned in the assignment document.

Using the intuitive values for initializing the distribution tables, the accuracy is calculated as 84.6%.

Notes:

1. As we increase the number of iterations of the estimation and maximization steps, the accuracy calculated in step (c) increases. At around 10 iterations, the accuracy reaches its maximum at the values mentioned above. As we increase the number of iterations, this value does not increase any more until a certain number. From then on the value tends to decrease. We think this is because EM tends to over-fit the data as the number of iterations increases above a certain value.
2. When the value of Intelligence being 0 is set somewhere between 0.3 and 0.65, the accuracy remains constant and is at a peak.
3. When the value of intelligence is set as 0.5 (for 0 and 1), it tends to remain the same even after 10 iterations of both the expectation and maximization steps.

References:

1. http://www.mathematik.uni-ulm.de/stochastik/lehre/ss06/markov/skript\_engl/node12.html