Project 4 – HMMs & Customer Journey

During the time that a customer becomes familiar with your product and to the time that they buy, the customer goes through multiple steps, which can be described as

- Zero: Step zero – customer doesn't even know you exist.
- Aware: Customer is aware of your product
- Considering: Customer considers that your product can provide it some value (solve some problem for them)
- Experiencing: Customer is trying out your product via a trial run and is experiencing what it is like
- Ready (to buy): Customer is excited enough that the customer is ready to pay for the product.
- Satisfied: Customer is now a satisfied, paying customer. You can recognize this state simply from the fact that they make a payment from the credit card page, so you don't need to infer this state, it is merely included here for completeness of the journey.

At any time, the customer may drop off and go to "will not buy" category (the deal is lost). The customer may also stay in the same stage. Here is a transition probability between these stages:

Initial State	New State	Probability
Zero	Aware	0.4
Aware	Considering	0.3
Aware	Ready	0.01
Aware	Lost	0.2
Considering	Experiencing	0.2
Considering	Ready	0.02
Considering	Lost	0.3
Experiencing	Ready	0.3
Experiencing	Lost	0.3
Ready	Lost	0.2

You have a variety of links on your website, belonging to different categories, such as product demos, how to videos, customer testimonials, pricing calculators, etc. You can see what the customer clicks on (evidence/emissions). You need to identify which stage the customer is in. At any stage, the customer has a different probability of being interested in category of link. (Customer may also not click on any link, or multiple links!)

Stage	Demo	Video	Testimonial	Pricing	Blog	Payment
Zero	0.1	0.01	0.05	0.3	0.5	0.0
Aware	0.1	0.01	0.15	0.3	0.4	0.0
Considering	0.2	0.3	0.05	0.4	0.4	0.0
Experiencing	0.4	0.6	0.05	0.3	0.4	0.0

Ready	0.05	0.75	0.35	0.2	0.4	0.0
Lost	0.01	0.01	0.03	0.05	0.2	0.0
Satisfied	0.4	0.4	0.01	0.05	0.5	1.0

Structure of the Input File

Lines that begin with # (hash/pound sign) are comments and can be ignored.
#
There is one line per time slot for up to 40 timeslots (interactions)
In each timeslot there is a list of links that the user clicked on.
For any one timeslot, the list may be empty
Demo,Video

Pricing

Demo
Video,Blog
Video,Payment

For some files, the actual states are included as well.

Expected Output

- Give an HMM for this situation. Specify the states, the transition probabilities and the emission probabilities.
- Output the most likely explanation of the state given the observations.