CS151 - Written Problem 6
Solutions

1. More Taxi prediction

   (a) Look at the Taxi cab example again from lecture notes and make sure that you understand how we calculated 1) the filtered probability of $P(CarOrTaxi_2|red, yellow)$ 2) $P(CarOrTaxi_3|red, yellow)$

   (b) Let's say you see a third vehicle now that is yellow. What is the probability that this third vehicle is a Taxi?

   $$P(X_3|NY, Y, Y) = \alpha P(Y|X_3)(P(X_3|C)message(C)+P(X_3|T)message(T))$$
   $$= \alpha \begin{bmatrix} 0.1 \\ 0.75 \end{bmatrix} \left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} 0.133 + \begin{bmatrix} 0.75 \\ 0.25 \end{bmatrix} 0.867 \right)$$
   $$= \alpha \begin{bmatrix} 0.0717 \\ 0.212 \end{bmatrix} = \begin{bmatrix} 0.252 \\ 0.748 \end{bmatrix}$$

   (c) Now that you've seen the color of the third car, how does this change your probability of the second part being a cab, that is, what is $P(CarOrTaxi_2|red, yellow, yellow)$?

   In this case we're dealing with smoothing since we want to ask a question about a state where we have both information about the observed variable up to that state as well as in the future.

   $$P(X_2|NY, Y, Y) = \alpha P(X_2|NY, Y)(P(Y|C)P(|C)P(C|X_3)+P(Y|T)P(|T)P(T|X_3)$$
   $$= \alpha \begin{bmatrix} 0.133 \\ 0.867 \end{bmatrix} \left(0.1 \begin{bmatrix} 0.5 \\ 0.75 \end{bmatrix} + 0.75 \begin{bmatrix} 0.5 \\ 0.25 \end{bmatrix} \right)$$
   $$= \alpha \begin{bmatrix} 0.0565 \\ 0.228 \end{bmatrix} = \begin{bmatrix} 0.199 \\ 0.801 \end{bmatrix}$$

   (d) How did the smoothed estimate from (c) change from our filtered estimate when we only had seen a red and a yellow car? Explain qualitatively why this new estimate in (c) makes sense, given your evidence and transition models (I am looking for an English description here).

Our probability that car 2 was a taxi has decreased after receiving the information that car 3 was yellow. This makes sense because taxis are unlikely to bunch up, so seeing two of them in a row doesnt make sense ($P(T|T)$ is much lower than $P(C|T)$). So after seeing the second red car you become less sure that you saw a taxi, in effect because your data doesnt fit the model very welly-our evidence model is essentially competing with your transition model, and they cancel each other out a bit.

2. An appealing use of HMMs is for localization: in other words, given a map, and a set of observations of your environment, figure out where you are. Suppose we are walking around the Claremont Colleges campus, which is roughly a 1x1 mile square (5280 feet by 5280 feet) (OK, not quite, but close enough...). Weve been given a map, and wed like to figure out where we are every ten seconds, down to a resolution of 1 foot.

   (a) Let's start to formalize this as an HMM. What does each of the hidden states X_t represent? What is the domain of each state variable? How big is this domain?

   Each state X_t represents the location at time t. The domain is a position in the grid, the set of 1x1 foot squares on the campuses. The size of the domain is 5280^2.

   (b) Suppose that youre walking with a blindfold on, at roughly 2 miles per hour, trying to go straight. You can ignore obstacles (such as buildings) for now. Whats a reasonable transition model, $P(X_t|X_{t-1})$?

   Since we dont know what direction were walking in, the transition model should give equal probability to all of the squares roughly 15 feet away, since 2 miles an hour is about 1.5 feet per second.

   (c) Assume that we get dropped off somewhere on the campus blindfolded but we dont know where. Whats a good starting, prior distribution $P(X_1)$?

Uniform, i.e. each location is equally likely.

(d) Suppose that every ten seconds we can stop and take our blindfold off, and look for Smith Tower (on Pomona's campus). If we can see it, we measure approximately how far away it is by measuring its apparent height. We then report the approximate distance in 100 foot increments. What should each evidence variable E_t represent? What is the domain of each evidence variable? How big is this domain?

Each E_t should be the distance from Smith tower. The domain is $[0, 100, 200, ..., 7500]$ (the length of the diagonal of the spacewed never be this far away, but we certainly cant be farther). The size of the domain is 75.

(e) Formalize the emission model $P(E_t|X_t)$

Again, since we dont know which direction we are from the tower, we would have a ring of equal probability at all the positions that are E_t away from the tower, with some smoothing for measurement error. If we assume we can tell the direction (e.g., we have a compass or something) then E_t gives us X_t, with some smoothing around that position to account for noise.

(f) Suppose we walk "straight" for 1 minute (60 seconds), stopping every 10 seconds to measure our distance to Smith Tower. 1) We want to know where we are. What HMM question is this? 2) We want to know where we walked for that 60 seconds. What HMM questions is this?

In the first case, this is the filtering question. Given the 6 observations of the tower, we want to figure out the hidden state variables for those time steps. This is the most likely sequence and can be found using the Viterbi algorithm.