1 Reasoning over time (22/11/2018)

1.1 Objectives

At the end of this repetition you should be able to:

- Define a Markov Model
- Define what is Filtering/Prediction/Smoothing/Most Likely Explanation in general and how it is done in the context of Markov Model
- Explain the simplified matrix representation of HMM (Hidden Markov Model).
- Define what is a Kalman filter and be able to use it in the context of a dynamic random process.

1.2 Exercises

a The coin (≈ 30 minutes)

You are in a room containing a table, on this table are placed 3 very precious biased coins (named A, B and C). Suddenly another person enters the room and takes the coins. He decides to throw 4 times a coin and then asks you 4 questions, before that he provides you the following information. First he tells you that he has selected the first coin uniformly at random. Then, to chose the next coins he has kept the same coin with a probability 2/3 or he has replaced it by another coin with equal probabilities, afterward he has thrown the selected coin. When you entered the room you inspected the coins and noticed that the coin A has a head probability of 80%, the B 50% and the C 20%. The result of the throws are head, head, tail and head. If you answer right to his questions he will give you the coins. The questions are the following:

1. Give the hidden Markov model of this statement.
    - The hidden state at time t is $X_t \in D_{X_t} = \{A, B, C\} = \{1, 2, 3\}$, it represents the coin chosen.
    - The evidence at time t is $E_t \in D_{E_t} = \{Head, Tail\} = \{1, 2\}$, it represents the result of the throw.
    - The prior matrix
    $$f_0 = P(X_0) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}^T$$
    .
    - The transition matrix
    $$T = P(X_t|X_{t-1}) = \begin{bmatrix} \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix} \quad \text{where} \quad T_{ij} = P(X_t = j|X_{t-1} = i)$$
    .
    - The sensor matrix
    $$B = P(E_t|X_t) = \begin{bmatrix} 0.8 & 0.2 \\ 0.5 & 0.5 \\ 0.2 & 0.8 \end{bmatrix} \quad \text{where} \quad B_{ij} = P(E_t = j|X_t = i)$$
    .

2. What are the probabilities of the last coin given the sequence of evidence ?
   Let's first define the observation matrices:
   $$O_1 = O_2 = O_4 = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \quad O_3 = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.8 \end{bmatrix}$$

We apply a forward pass to filter and to derive the following equalities:

$$\begin{aligned}
f_1 &= P(X_1|e_1) = \alpha_1 P(X_1)P(e_1|X_1) \\
&= \alpha_1 \sum_{X_0} P(X_0)P(X_1|X_0)P(e_1|X_1) = \alpha_1 O_1 T^T f_0 && (1)\\
f_2 &= P(X_2|e_{1:2}) = \alpha_2 \sum_{X_1} P(X_{1:2}, e_2|e_1) \\
&= \alpha_2 \sum_{X_1} P(e_2|X_2, X_1, e_1)P(X_2|X_1, e_1)P(X_1|e_1) = \alpha_2 O_2 T^T f_1 && (2)\\
f_3 &= P(X_3|e_{1:3}) = \alpha_3 O_3 T^T f_2 && (3)\\
f_4 &= P(X_4|e_{1:4}) = \alpha_4 O_4 T^T f_3 \approx \begin{bmatrix} 0.472 & 0.374 & 0.154 \end{bmatrix}^T && (4)
\end{aligned}$$

.

3. What are the probabilities of the first coin chosen given the sequence evidence ? And of the first coin thrown ?
   We apply a backward pass to smooth and to derive the following equalities:

$$\begin{aligned}
b_4 &= P(e_4|X_3) = \sum_{X_4} P(e_4, X_4|X_3) \\
&= \sum_{X_4} P(e_4|X_3, X_4)P(X_4|X_3) = \sum_{X_4} P(e_4|X_4)P(X_4|X_3) \\
&= TO_4 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} && (5)\\
b_3 &= P(e_{3:4}|X_2) = \sum_{X_3} P(e_{3:4}, X_3|X_2) \\
&= \sum_{X_3} P(e_{3:4}|X_2, X_3)P(X_3|X_2) = \sum_{X_3} P(e_4|X_3)P(e_3|X_3, e_4)P(X_3|X_2) \\
&= TO_3 b_4 && (6)\\
b_2 &= P(e_{2:4}|X_1) = \sum_{X_2} P(e_{2:4}, X_2|X_1) \\
&= \sum_{X_2} P(e_{2:4}|X_1, X_2)P(X_2|X_1) = \sum_{X_2} P(e_2|X_2, e_{3:4})P(e_{3:4}|X_2)P(X_2|X_1) \\
&= TO_2 b_3 && (7)\\
b_1 &= P(e_{1:4}|X_0) = TO_1 b_2 && (8)
\end{aligned}$$

The probability of the first coin given the evidence selected is:

$$P(X_0|e_{1:4}) = \alpha P(X_0, e_{1:4}) = \alpha f_0 \times b_1 \approx \begin{bmatrix} 0.457 & 0.331 & 0.212 \end{bmatrix}^T \quad (9)$$

The probability of the first coin thrown given the evidence selected is:

$$\begin{aligned}
P(X_1|e_{1:4}) &= \alpha P(X_1, e_{2:4}|e_1) = \alpha P(X_1|e_1)P(e_{2:4}|X_1, e_1) \\
&= \alpha f_1 \times b_2 \approx \begin{bmatrix} 0.580 & 0.329 & 0.091 \end{bmatrix}^T && (10)
\end{aligned}$$

4. What is the most likely sequence of coins thrown ?
   The most likely sequence given the evidence is:

$$x_{1:4}^{ML} = \arg\max_{x_{1:4}} P(x_{1:4}|e_{1:4}) \quad (11)$$

$$(12)$$

. It can be computed efficiently by using the Viterbi algorithm. Let define the vector $m_i \in R_+^3$ such that $m_i(j)$ gives the probability of the most likely path to the i^{th} state with value j. It can be computed recursively with the following equalities:

$$
\begin{aligned}
m_1 &= P(X_1|e_1) = f_1 && (13)\\
m_t &= \max_{x_{1:t-1}} P(X_t, x_{1:t-1}|e_{1:t}) && (14)\\
&= \max_{x_{1:t-1}} \alpha P(X_t, x_{1:t-1}, e_t|e_{1:t-1}) && (15)\\
&= \max_{x_{1:t-1}} \alpha P(e_t|X_t, x_{1:t-1}, e_{1:t-1})P(X_t|x_{t-1}, x_{1:t-2}, e_{1:t-1})P(x_{1:t-1}|e_{1:t-1}) && (16)\\
&= \alpha P(e_t|X_t) \max_{x_{t-1}} P(X_t|x_{t-1}) \max_{x_{1:t-2}} P(x_{1:t-1}|e_{1:t-1}) && (17)\\
&= \alpha P(e_t|X_t) \max_{x_{t-1}} P(X_t|x_{t-1}) m_{t-1}(x_{t-1}) && (18)
\end{aligned}
$$

b Hyperloop(≈ 25 minutes)

This year ULiège has decided to get into the hyperloop competition[1]. Briefly, what you should do to win this competition is to build the fastest and most reliable autonomous pod.

One of the most important engineering problem to build the pod is to be able to compute a robust estimation of the state of the pod (position, speed) given many noisy sensors.

This morning you received a mail asking you what would be your solution to this estimation problem. The mail contains information about the sensors they plan to put in the pod. They say that there will use 3 unbiased speed sensors with a 99.7% accuracy[2] of 0.1m/s and a GPS sensor (also unbiased) which provides the pod's position (in one dimension) with a 99.7% precision of 1 meter. After some research on the web you found out that you should use a Kalman filter to solve this task. Thus you have to define the key elements of the Kalman filter in the context of the state estimation of the ULiège's pod. Please define below these elements, you can assume that the acceleration a is distributed normally around $\mu_a m/s^2$ with a variance equal to σ_a^2.
To define the kalman filter we have to define the following 3 elements:

1. The prior (which is assumed to be Gaussian):

$$p(x_0) = \mathcal{N}(x_0|\mu_{x_0}, \Sigma_{x_0})$$

.

2. The transition model (which is assumed to be Gaussian):

$$p(x_t|x_{t-1}) = \mathcal{N}(x_t|Ax_{t-1} + b, \Sigma_x)$$

.

3. The measurement model:

$$p(e_t|x_t) = \mathcal{N}(e_t|Cx_t + d, \Sigma_e)$$

.

In this case we can assume the initial position and speed of the pod known with the same accuracy as the sensors gives, so we have the following definition for μ_{x_0} and Σ_{x_0}:

$$\mu_{x_0} = \begin{bmatrix} x_0 \\ \dot{x_0} \end{bmatrix}$$

and

$$\Sigma_{x_0} = \begin{bmatrix} (\frac{1}{3})^2 & 0 \\ 0 & (\frac{0.1}{3\sqrt{3}})^2 \end{bmatrix}$$

. Where x_0 and $\dot{x_0}$ respectively denote the initial position and speed of the pod. The covariance matrix values come from the fact that the 99.7% precision is equal to 3σ and that the variance of the mean of n random

[1] https://www.spacex.com/hyperloop

[2] It means that 99.7% of the value measured will get a smaller error. e.g. (https://math.stackexchange.com/questions/1412683/3-sigma-approximation)

variables is equal to the sum of each variance divided by n^2. Then to derive the transition model we use the laws of physics and we easily get:

$$A = \begin{bmatrix} 1 & \Delta_t \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} \frac{\mu_a}{2}\Delta_t^2 \\ \mu_a \Delta_t \end{bmatrix}$$

where Δ_t denotes the time elapsed between two measurements. The covariance matrix of the transition model can be computed by the afine transformation formula of multivariate Gaussian which tells that if $y \sim \mathcal{N}(\mu_y, \Sigma_y)$ then $x = Qy + m \sim \mathcal{N}(\mu_x, \Sigma_x)$ where $\mu_x = Q\mu_y + m$ and $\Sigma_x = Q\Sigma_y Q^T$. Here we have $a \sim \mathcal{N}\mu_a, \sigma_a)$ which yields to :

$$\Sigma_x = \begin{bmatrix} \frac{1}{2}\Delta_t^2 \\ \Delta_t \end{bmatrix} \sigma_a^2 \begin{bmatrix} \frac{1}{2}\Delta_t^2 & \Delta_t \end{bmatrix} = \sigma_a^2 \begin{bmatrix} \frac{1}{4}\Delta_t^4 & \frac{1}{2}\Delta_t^3 \\ \frac{1}{2}\Delta_t^3 & \Delta_t^2 \end{bmatrix}$$

. Finally we can easily get that the mean of the sensors is given by:

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad d = 0$$

. The covariance matrix is similarly computed and is equal to:

$$\Sigma_e = \begin{bmatrix} (\frac{1}{3})^2 & 0 & 0 & 0 \\ 0 & (\frac{0.1}{3})^2 & 0 & 0 \\ 0 & 0 & (\frac{0.1}{3})^2 & 0 \\ 0 & 0 & 0 & (\frac{0.1}{3})^2 \end{bmatrix}$$

1.3 Supplementary material

`http://ai.berkeley.edu/sections/section_6.pdf`