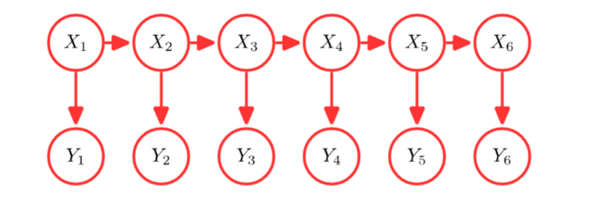
Review Session - PAI 2018 - Temporal Models

January 24, 2019

Learning & Adaptive Systems group
Institute of Machine Learning
ETH Zürich

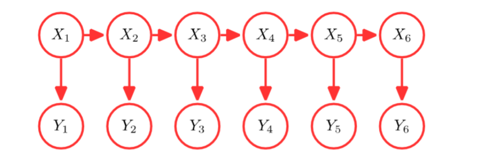
Hidden Markov models

- Instead of assuming all variables are observed (Y_t), we assume we have variables X_t, which we cannot observe.
- Assume: A graphical model for *Hidden Markov Model* (HMM)



- $\{X_t\}_t$ are called states and $\{Y_t\}_t$ are called observations.
- In this course: distributions assumed to be $P(X_t|X_{t+1}), P(Y_t|X_t)$ categorical or Gaussian mostly, however can be arbitrary.

What Inference Questions can we ask?



- **Bayesian Filtering**

$$P(X_t|Y_{1:t})$$

- **Prediction**

$$P(X_{t+\tau}|Y_{1:t}) \text{ where } \tau \geq 1$$

- **Smoothing** (similar to filtering but on past data)

$$P(X_\tau|Y_{1:t}) \text{ where } \tau < t$$

- **Most Probable Explanation (MPE)**

$$X_{1:t} = \arg\max P(\hat{X}_{1:t}|Y_{1:t})$$

How to do it?

- Prior $P(X_1)$ needed.
- Bayesian Filtering - usually done recursively. Assume you have $P(X_t|Y_{1:t-1})$, get $P(X_t|Y_{1:t})$ using Markov property and observation probabilities.
- Prediction (for one step; can be generalized)

$$P(X_{t+1}|Y_{1:t}) = \sum_x P(X_{t+1}, X_t = x|Y_{1:t}) = \sum_x P(X_{t+1}|X_t = x)P(X_t|Y_{1:t})$$

- Smoothing - can be recast as calculating a marginal! HMM being a polytree, we know that Belief propagation is applicable and converges fast with forward-backward passes!
- MPE - Sum product algorithm (search largest product).

Kalman Filter I - Basics

- A special case of HMM is Kalman filter, when X_t and Y_t are Gaussian with Linear dynamics and observation model.
- **Motion model = System dynamics** $P(X_{t+1}|X_t)$ and is modeled using,

$$X_{t+1} = \mathbf{F}X_t + \epsilon_t$$

  where, $\epsilon_t \sim \mathcal{N}(0, \Sigma_x)$. *N.B.: Sum of Normal RV is a Normal RV!*
- **Sensor model = Observations** $P(Y_t|X_t)$,

$$Y_t = \mathbf{H}X_t + \nu_t$$

  where, $\nu_t \sim \mathcal{N}(0, \Sigma_x)$. *N.B.: ϵ_t and ν_t are multivariate normals!*

Kalman Filter II - How to?

- In general, this is the same as previously; but now the domain is different. Previously categorical distributions - now continuous.
- Filtering:

$$P(X_t|Y_t = y_{1:t}) = \frac{1}{Z} P(X_t|Y_{t-1} = y_{1:t-1}) P(Y_t = y_t|X_t)$$

- Prediction:

$$P(X_{t+1}|Y_{1:t} = y_{1:t}) = \int P(X_{t+1}|X_t = x) P(X_t = x|Y_{1:t} = y_{1:t}) dx$$

- All these are Gaussian integrals; that can be solved exactly.
- We can calculate the Kalman filter uncertainty offline. Independent of $y_{1:t}$.
    - Review my multivariate Gaussians!

$$\mu_{A|B} = \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(\mathbf{x}_B - \mu_B)$$
$$\Sigma_{A|B} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}$$

Particle Filters I

- What if the "motion" model is nonlinear? - For example Particle filters.
- Idea: Pick samples $\{x_i\}$ that are propagated through the dynamics and then produce a histogram! Samples: $\{x_{t,i}\}_{i=1}^N$ from distribution $P(X_t|y_{1:t}) \approx \frac{1}{N}\sum_{i=1}^N \delta_{x,x_{t,i}}$.
- Propagate particles $x_i' \sim P(X_{t+1}|X_t, x_{t,i})$, calculate weights $w_i = \frac{1}{Z}P(y_{t+1}|x_i')$ (transition model known).
- Next step $x_{i,t+1} \sim \frac{1}{N}\sum_{i=1}^N w_i\delta_{x,x_{t,i}}$
- Why this re-weighting? If we do not do it; we have a mode collapse (out of scope of the course).

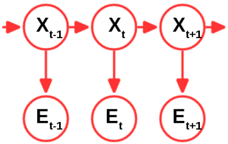
Exam 2013 Problem 5 - I

5 [12 points] Temporal Models: Exploratory Rover

An exploratory rover is navigating through planet Vulcan, a deserted planet scattered with volcanic vents and radioactive valleys. The rover is equipped with a thermometer that registers only two levels, *hot* and *cold*. The rover sends back thermal responses $E =$ `hot` when it is at a **V**olcanic vent (V), or a **R**adioactive valley (R), and $E =$ `cold` when it is at **N**ormal area (N). There is no chance of a mistaken reading.

The rover can only stay in one area on any given day. It travels around according to the following transition probabilities:

	$P(X_t \mid X_{t-1} = N)$	$P(X_t \mid X_{t-1} = V)$	$P(X_t \mid X_{t-1} = R)$
$X_t = N$	0.7	0.6	0.2
$X_t = V$	0.2	0.3	0.2
$X_t = R$	0.1	0.1	0.6



Exam 2013 Problem 5 - II

1. Define $\mathbf{X}_t = (P(X_t = N), P(X_t = V), P(X_t = R))$

   $$\mathbf{T} = \begin{pmatrix} 0.7 & 0.6 & 0.2 \\ 0.2 & 0.3 & 0.2 \\ 0.1 & 0.1 & 0.6 \end{pmatrix}, \text{ then } \mathbf{X}_t\mathbf{T} = \mathbf{X}_{t+1}.$$

2. Further, we know that observation model for Y as well. $P(E_t = \text{hot}|X_t = V) = P(E_t = \text{hot}|X_t = R) = 1$ and $P(E_t = \text{cold}|X_t = N) = 1$. So very simple model.
3. a) Observe {cold, hot hot}, what is $P(X_{1:3} = (NNN)|E_{1:3} = (CHH))$?

Using d-separation rules

$$P(X_{1:3}|E_{1:3}) = \underbrace{P(X_3|X_2, E_3)}_{P(E_3|X_3)P(X_3|X_2)/P(X_2,E_3)} P(X_2|X_1, E_{2:3})P(X_1|E_{1:3})$$

$P(E_3 = hot|X_3 = N) = 0$, so the overall probability is zero.

Exam 2013 Problem 5 - III

- b) Given $E\{\text{cold,hot,hot,cold}\}$ what is the most likely $\{X_{1:4}\}$. *Hint: Not many calculations*.

This is a MPE query, in other words,

$$X_{1:4} = \arg\max_{\hat{X}_{1:4}} P(\hat{X}_{1:4}|Y_{1:4})$$

Given that in cold state can be only observed when $X_i = N$, means that $X_1 = N$, $X_4 = N$. Now, we need to look which state is more probable for the middle points: VV,VR,RV,RR. We can calculate these probabilities using the transition matrix $\mathbf{T}$ i.e. $P(X_2 = V, X_3 = R, X_1 = N, X_4 = N) = P(X_2 = V|X_1 = N)P(X_3 = R|X_2 = V)P(X_4 = N|X_3 = R)$

Exam 2013 Problem 5 - IV

- c) We use particle filter with $N = 8$ particles.
  - 5 N ($i \in 1, \dots 5$)
  - 2 V ($i \in 6, 7$)
  - 1 R ($i = 8$)

  The rover sends back $E =$ hot. What are the weights for the next sampling?

From the previous slides, we know

$$w_i \propto P(E = \text{hot} | X = x_i)$$

For example for $w_1 \propto P(E = \text{hot} | X = N) = 0$. Similar for others. Weights need to sum to 1 so we can normalize them.