About this class

Temporal Models in General

Hidden Markov Models

The Kalman Filter

Basic Framework

[Most of this lecture from Russell & Norvig, Chap 15]

The world evolves over time. We describe it with certain state variables. These variables exist at each time period. Some are observable and some unobservable. For simplicity, we assume the same variables remain (un)observable at each time period.

Let X_t denote the state variables

Let E_t denote the evidence variables

How do we define a state? Everything that's important to the world?

Stationary processes: Laws governing state change and observations do not change over time (different from static, where the state itself does not change)

Markov assumption: Current state depends only on a finite history of past states. Standard $n = 1$ – then

$$\Pr(X_t|X_{0:t-1}) = \Pr(X_t|X_{t-1})$$

This is the *transition model*

Another standard Markovian assumption:

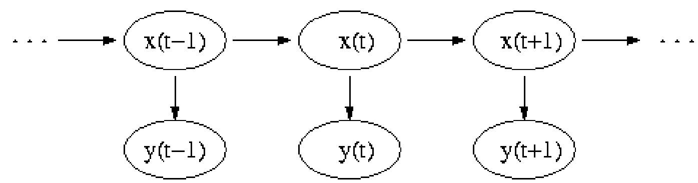
$$\Pr(E_t|X_{0:t}, E_{0:t-1}) = \Pr(E_t|X_t)$$

This is the *observation model*

Also need to specify a prior $\Pr(X_0)$

An Example

[Image from Wikipedia]



Suppose the x's are the underlying financial status of a company and the y's are the S&P bond ratings (evidence visible to the world). Then the bond rating relies only on the current financial status of the company, and it's all you get to see.

To fully specify the model, let's say bond ratings can only be A or B, and financial status can only be Good or Bad. Then suppose $\Pr(y = A|\text{Good}) = 0.8$ and $\Pr(y = A|\text{Bad}) =$

0.05 with the probabilities of B ratings obviously one minus these values. Still need to specify the prior and the transition model. Example of transition model:

$$\Pr(\text{Good}_{t+1}|\text{Good}_t) = 0.9$$

$$\Pr(\text{Good}_{t+1}|\text{Bad}_t) = 0.05$$

To increase expressivity, could incorporate more information by adding state variables (most recent profits, monthly sales, sector condition etc.), or could increase the amount of history you condition on ($\Pr(\text{Good}_3|\text{Good}_2, \text{Good}_1) > \Pr(\text{Good}_3|\text{Good}_2, \text{Bad}_1)$).

Inference in Markov Models

1. Filtering: $\Pr(X_t|e_{1:t})$ – update belief state

2. Prediction (a subproblem of filtering, almost?): $\Pr(X_{t+k}|e_{1:t})$

3. Smoothing: $\Pr(X_k|e_{1:t}), k < t$ – update estimates of what states you were previously in, based on evidence that would have been in the future then...

4. Finding most probably state sequence: Speech recognition!

$$\arg\max_{x_{1:t}} \Pr(x_{1:t}|e_{1:t})$$

Filtering

We will show that you can perform online recursive estimation

$$\Pr(X_{t+1}|e_{1:t+1}) = \Pr(X_{t+1}|e_{1:t}, e_{t+1})$$

$$= \alpha \Pr(e_{t+1}|X_{t+1}, e_{1:t}) \Pr(X_{t+1}|e_{1:t})$$

By the Markov property of the observations:

$$= \alpha \Pr(e_{t+1}|X_{t+1}) \Pr(X_{t+1}|e_{1:t})$$

$$= \alpha \Pr(e_{t+1}|X_{t+1}) \sum_{x_t} \Pr(X_{t+1}|x_t, e_{1:t}) \Pr(x_t|e_{1:t})$$

$$= \alpha \Pr(e_{t+1}|X_{t+1}) \sum_{x_t} \Pr(X_{t+1}|x_t) \Pr(x_t|e_{1:t})$$

The first two terms follow from the observation model and the transition model, respectively. The third term is precisely what we are trying to estimate for time $t+1$, so we've come up with a recursive formulation that solves the problem.

The filtered estimate $\Pr(x_t|e_{1:t})$ is sometimes called the forward message $f_{1:t}$

Prediction:

$$\Pr(X_{t+k+1}|e_{1:t}) = \sum_{x_{t+k}} \Pr(X_{t+k+1}|x_{t+k}) \Pr(x_{t+k}|e_{1:t})$$

Eventually converges to the stationary distribution of the Markov process and essentially loses all informational value not already contained in the model

Smoothing

$$\Pr(X_k|e_{1:t}) = \Pr(X_k|e_{1:k}, e_{k+1:t})$$

$$= \alpha \Pr(e_{k+1:t}|X_k, e_{1:k}) \Pr(X_k|e_{1:k})$$

$$= \alpha \Pr(X_k|e_{1:k}) \Pr(e_{k+1:t}|X_k)$$

$$= \alpha f_{1:k} b_{k+1:t}$$

Where the "backward message" is:

$$\Pr(e_{k+1:t}|X_k) = \sum_{x_{k+1}} \Pr(e_{k+1:t}|X_k, x_{k+1}) \Pr(x_{k+1}|X_k)$$

$$= \sum_{x_{k+1}} \Pr(e_{k+1}, e_{k+2:t}|x_{k+1}) \Pr(x_{k+1}|X_k)$$

$$= \sum_{x_{k+1}} \Pr(e_{k+1}|x_{k+1}) \Pr(e_{k+2:t}|x_{k+1}) \Pr(x_{k+1}|X_k)$$

Points to note: first and third terms come from the model. Second term is the recursive step. Initialize the backward phase with:

$$b_{t+1:t} = 1$$

Smoothing and Filtering Example

Umbrella World: Rainy days and director's umbrellas. The security guard never leaves the building.

Abusing notation ever-so-slightly

$$\Pr(R_t|R_{t-1}) = 0.7$$

$$\Pr(R_t|\neg R_{t-1}) = 0.3$$

$$\Pr(U_t|R_t) = 0.9$$

$$\Pr(U_t|\neg R_t) = 0.2$$

Umbrella appears on both days

Day 0 belief about rain: $(0.5, 0.5)$

Day 1 action:

$$\Pr(R_1) = \sum_{r_0} P(R_1|r_0)P(r_0)$$

$$= (0.7, 0.3) \times 0.5 + (0.3, 0.7) \times 0.5 = (0.5, 0.5)$$

Observation update:

$$\Pr(R_1|u_1) = \alpha \Pr(u_1|R_1) \Pr(R_1)$$

$$= \alpha(0.9, 0.2) \times (0.5, 0.5) \simeq (0.818, 0.182)$$

Now for Day 2:

$$\sum_{r_1} \Pr(R_2|r_1) \Pr(r_1|u_1)$$

$$= (0.7, 0.3) \times 0.818 + (0.3, 0.7) \times 0.182$$

$$\simeq (0.627, 0.373)$$

Observation update:

$$\alpha(0.9, 0.2) \times (0.627, 0.373) \simeq (0.883, 0.117)$$

Now suppose we want to compute the smoothed estimate of rain on Day 1. We already have the forward message $(0.818, 0.182)$

$$\text{Pr}(u_2|R_1) = \sum_{r_2} \text{Pr}(u_2|r_2) 1 P(r_2|R_1)$$

$$= (0.9 \times 1 \times (0.7, 0.3)) + (0.2 \times 1 \times (0.3, 0.7))$$

$$= (0.69, 0.41)$$

Then

$$\text{Pr}(R_1|u_1, u_2) = \alpha(0.818, 0.182) \times (0.69, 0.41)$$

$$\simeq (0.883, 0.117)$$

Note that the smoothed estimate is higher than the filtered estimate, because the evidence of an umbrella at step 2 propagates backward to make it seem more likely that it rained on the previous day!

Most Likely Sequence

Why not just compute the smoothed estimates and then pick the most likely ones? Not the same as most likely path! Example – suppose there was an impossible transition?

Example: reconstructing speech through uttered phonemes

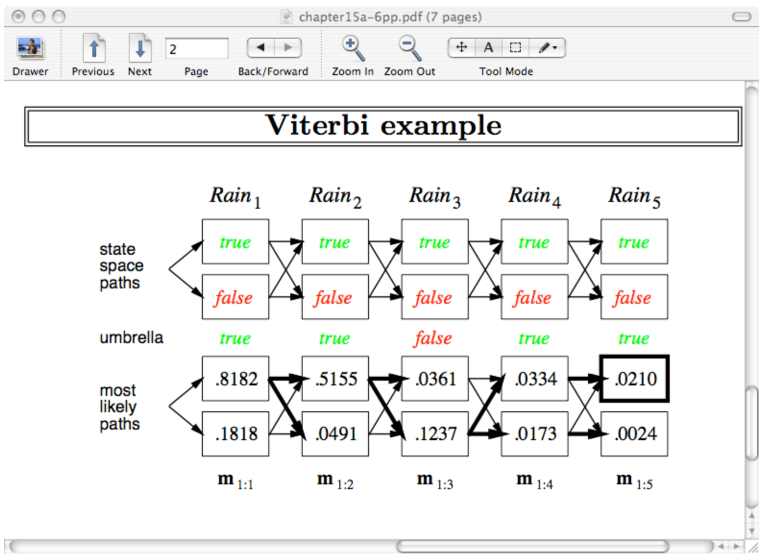
Viterbi Algorithm

Key insight: Consider all possibilities of the state at time $t+1$. For each one (for example, Rain = True), consider all paths that reach it. The most likely path to that state must consist of a most likely path to *some state* at time t followed by a transition to that state.

Probabilities of paths:

$$\max_{x_1,\ldots,x_t} \text{Pr}(x_1, \ldots, x_t, X_{t+1}|e_{1:t+1})$$

$$= \alpha \Pr(e_{t+1}|X_{t+1}) \max_{x_t} \Pr(X_{t+1}|x_t)$$
$$\max_{x_1,\ldots x_{t-1}} \Pr(x_1,\ldots,x_t|e_{1:t})$$



Hidden Markov Models

Exactly what we've been talking about: *single* discrete unobserved random variable

Allows for nice matrix representations of many things

Domain of X_t is $\{1,\ldots,S\}$

Transition matrix: $T_{ij} = P(X_t = j|X_{t-1} = i)$, e.g., $\begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$

Observation matrix: O_t for each time step, diagonal elements $P(e_t|X_t = i)$
e.g., with $U_1 = true$, $O_1 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.2 \end{pmatrix}$

Forward and backward messages as column vectors:

$$\begin{aligned} f_{1:t+1} &= \alpha O_{t+1} T^T f_{1:t} \\ b_{k+1:t} &= T O_{k+1} b_{k+2:t} \end{aligned}$$

The Kalman Filter

Gaussian prior, linear Gaussian transition model and observation model

Transition model:

$$p(x_{t+1}|x_t) = N(Fx_t, \Sigma_x)$$

Observation model:

$$p(z_t|x_t) = N(Hx_t, \Sigma_z)$$

Nice thing about Gaussians, yet again: posterior representation is Gaussian, allows us to do things

Importantly,

1. If current distribution is Gaussian and transition model is linear Gaussian, then the one-step predicted distribution is Gaussian:

$$p(X_{t+1}|e_{1:t}) = \int_{x_t} p(X_{t+1}|x_t)p(x_t|e_{1:t})dx_t$$

2. If the predicted distribution is Gaussian and the sensor model is linear Gaussian, then the updated distribution after conditioning on the new observation is Gaussian:

$$p(X_{t+1}|e_{1:t+1}) = \alpha p(e_{t+1}|X_{t+1})p(X_{t+1}|e_{1:t})$$

Example: State of a Random Walk

Prior: $N(x_0, \sigma_0^2)$

Transition: $x_{t+1} \leftarrow N(x_t, \sigma_x^2)$

Observation: $z_t \leftarrow N(x_t, \sigma_z^2)$

Update equations become:

$$\mu_{t+1} = \frac{(\sigma_t^2 + \sigma_x^2) z_{t+1} + \sigma_z^2 \mu_t}{\sigma_t^2 + \sigma_x^2 + \sigma_z^2}$$

$$\sigma_{t+1}^2 = \frac{(\sigma_t^2 + \sigma_x^2)\sigma_z^2}{\sigma_t^2 + \sigma_x^2 + \sigma_z^2}$$