Hidden Markov Models

Temporal Models $\quad x_0 \to x_1 \to x_2 \to \cdots \to x_{t-1} \to x_t \to \cdots$

Special case of Bayesian Networks $\quad x_t$ depends on x_{t-1}

X_t : Rain on day t

$P(x_t = 1 \mid x_{t-1} = 1)$

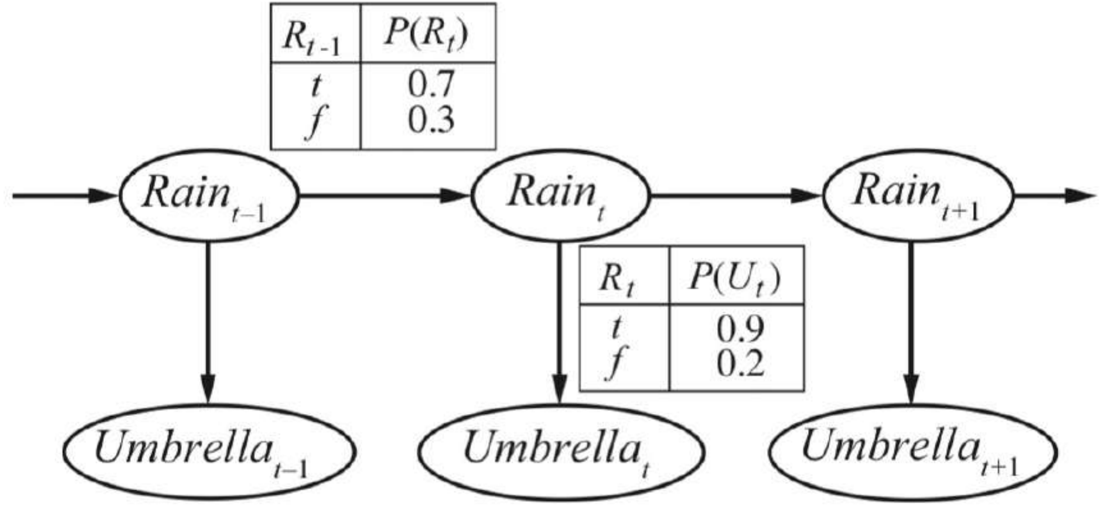
$P(x_t = 1 \mid x_{t-1} = 0)$

x_{t-1}	$P(x_t)$
0	0.1
1	0.8

Observability – indoors, only see umbrella

$$\begin{array}{ccccccc} R_0 \to & R_1 & \to & R_2 & \to \cdots \to & R_t & \to \\ & \downarrow & & \downarrow & & \downarrow & \\ & U_1 & & U_2 & & U_t & \end{array}$$

Hidden Markov Model



R_{t-1}	$P(R_t)$
t	0.7
f	0.3

R_t	$P(U_t)$
t	0.9
f	0.2

$P(U_t \mid R_t)$

$P(R_{t+1} \mid R_t)$

Markov Assumption

Applications: Speech / handwriting recognition

$$
\begin{tikzcd}
W_0 \arrow[r] & W_1 \arrow[r] \arrow[d] & W_2 \arrow[r] \arrow[d] & \cdots \\
& S_1 & S_2 &
\end{tikzcd}
$$

Actual words / letters

Observations

Notation: $x_{i:j} \stackrel{def}{=} x_i, x_{i+1}, \ldots, x_j$ Hidden state

$e_{i:j} \stackrel{def}{=} e_i, e_{i+1}, \ldots, e_j$ Evidence = observations

Typical Inference Problems

$P(x_t \mid e_{1:t})$ Current state from evidence Filtering

$P(x_{t+k} \mid e_{1:t})$ Future state from evidence Prediction

$P(x_k \mid e_{1:t}),\ k < t$ Past state estimation Smoothing

$\operatorname{argmax} P(x_{1:t} \mid e_{1:t})$ Most likely sequence

FILTERING

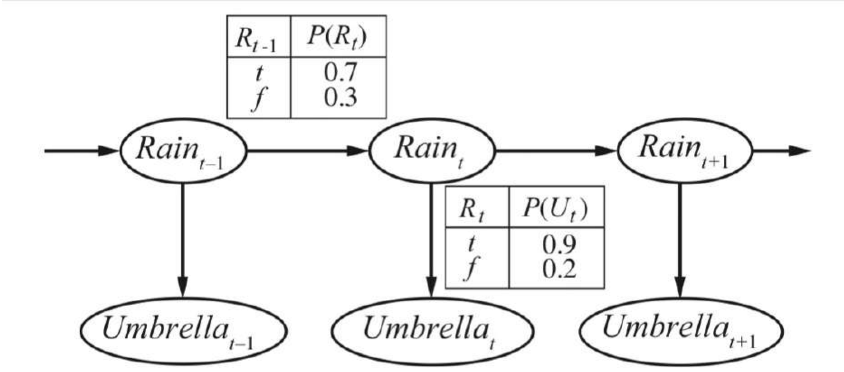
$$\underline{P(X_{t+1} \mid e_{1:t+1})} = P(X_{t+1} \mid e_{1:t}, e_{t+1})$$

$$= P(e_{t+1} \mid X_{t+1}, e_{1:t}) \cdot P(X_{t+1} \mid e_{1:t})$$

$$= P(e_{t+1} \mid X_{t+1}) \cdot P(X_{t+1} \mid e_{1:t})$$

$$= P(e_{t+1} \mid X_{t+1}) \cdot \sum_{x_t} P(X_{t+1} \mid x_t, e_{1:t}) P(x_t \mid e_{1:t})$$

$$= P(e_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) \boxed{P(x_t \mid e_{1:t})}$$



$U_1 = U_2 = t$

$P(R_2) = ?$

Assume $P(R_0) = 0.5$

$$P(R_1) = \sum_{r_0} P(R_1|r_0)\, P(r_0) = \langle 0.5, 0.5 \rangle$$

$$P(R_1|U_1) = P(U_1|R_1)\, P(R_1) = \langle 0.818, 0.182 \rangle$$

$$P(R_2|U_1) = P(R_2|R_1) = \langle 0.627, 0.373 \rangle$$

$$P(R_2|U_1, U_2) = \langle 0.883, 0.117 \rangle$$

PREDICTION

$$
\begin{tikzcd}
x_0 \arrow[r] & x_1 \arrow[r] \arrow[d, dash] & \cdots \arrow[r] & x_t \arrow[r] \arrow[d, dash] & x_{t+1} \arrow[r] & \cdots \arrow[r] & x_{t+k} \\
& e_1 & & e_t & & &
\end{tikzcd}
$$

$$P(x_{t+k} \mid e_{1:t})$$

$$= \underbrace{P(x_t \mid e_{1:t})}_{\text{FILTER}} \cdot P(x_{t+1} \mid x_t) \cdots P(x_{t+k} \mid x_{t+k-1})$$

SMOOTHING

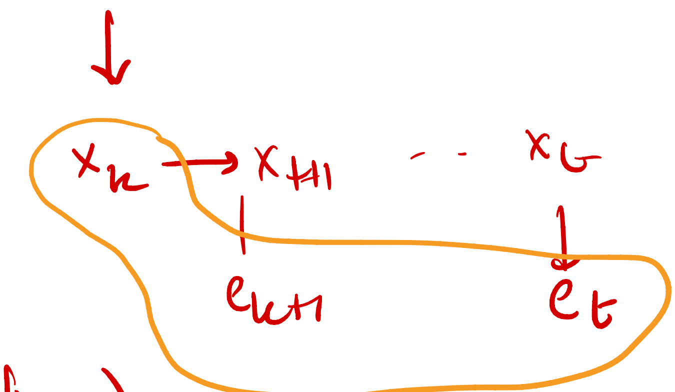
$$P(x_k \mid e_{1:t}), \quad k < t$$

$$
\begin{tikzcd}
x_0 \arrow[r] & x_1 \arrow[r] \arrow[d] & \cdots \arrow[r] & x_k \arrow[r] \arrow[d] & x_{k+1} \arrow[r] \arrow[d] & \cdots \arrow[r] & x_t \arrow[d] \\
& e_1 & & e_k & e_{k+1} & & e_t
\end{tikzcd}
$$

FILTERING $\longrightarrow$ $\qquad\qquad$ $\longleftarrow$?

$$P(x_k|e_{1:t}) = P(x_k|e_{1:k}, e_{k+1:t})$$

$$= P(x_k|e_{1:k}) \cdot P(e_{k+1:t}|x_k, e_{1:k})$$



$$P(e_{k+1:t}|x_k) = \sum_{x_{k+1}} P(e_{k+1:t}|\cancel{x_k}, x_{k+1}) \, P(x_{k+1}|x_k)$$

$$= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}|x_{k+1}) \, P(x_{k+1}|x_k)$$ MODEL

MODEL $P(e_{k+1}|x_{k+1}) \cdot P(e_{k+2:t}|x_{k+1})$ INDUCTION

Earlier, by filtering

$$P(R_1 \mid U_1) = \langle 0.818, 0.182 \rangle$$

$$P(R_2 \mid U_1, U_2) = \langle 0.883, 0.117 \rangle$$

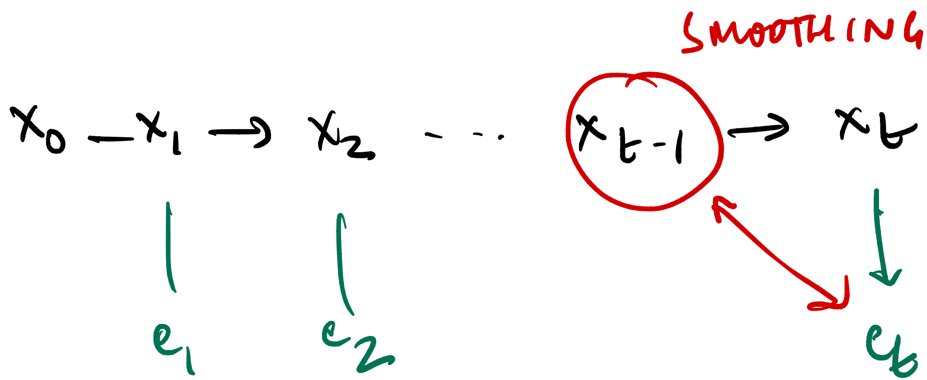
Smoothing

$$P(R_1 \mid U_1, U_2) = \langle 0.883, 0.117 \rangle$$

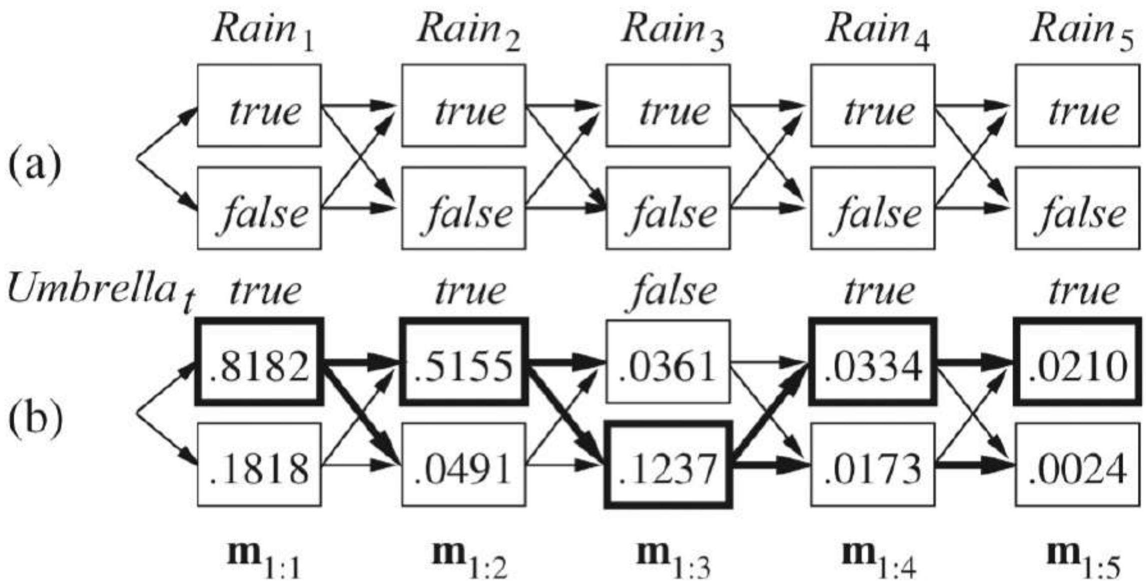
Better estimate!

Most likely explanation

Given $e_{1:t}$, find best $x_{1:t}$



$$\text{Best}(x_{1:t+1} \mid e_{1:t+1}) = \max_{x_{t+1}} P(x_{t+1} \mid x_t, e_{t+1}) \cdot \text{Best}(x_{1:t} \mid e_{1:t})$$



Dynamic Programming

$B(x, t+1)$ - Best prob of $t+1$ length sequence ending in state x

$$= \max_{y} P(x \mid e_{t+1}, y)\, B(y, t)$$

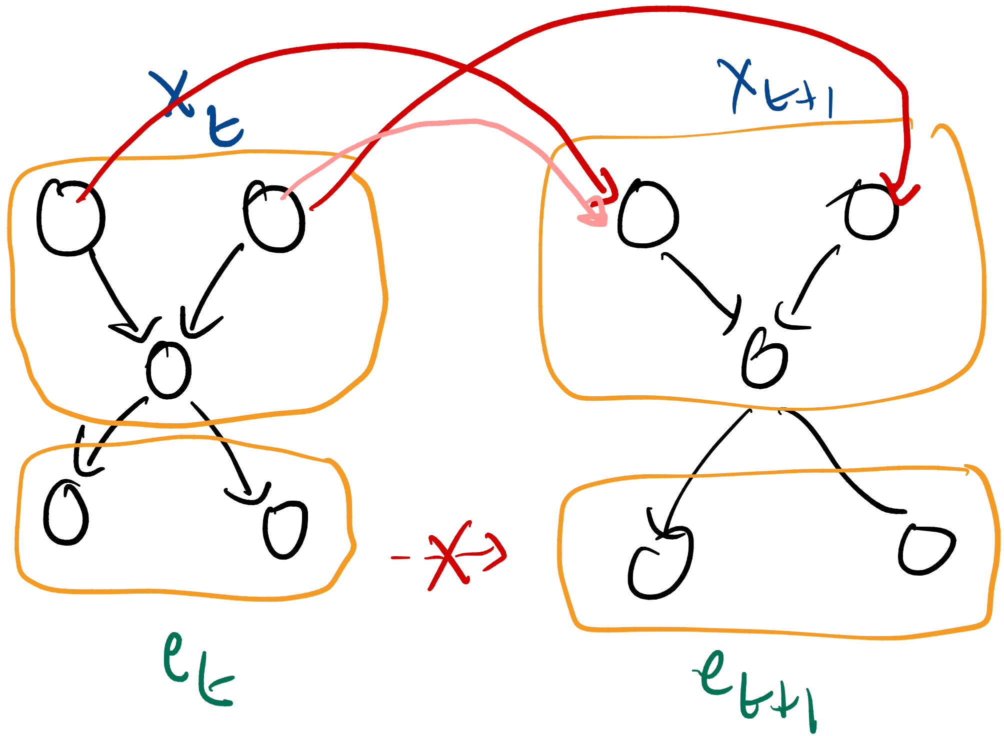
Viterbi

HMM - Linear unrolling

$P(x_{t+1} \mid x_t)$

$P(e_t \mid x_t)$

"Global"



DYNAMIC BAYESIAN NETWORK (DBN)