Lecture 27: Time and Uncertainty (Part 2)

Last Lecture

Filtering and Prediction for Probabilistic Inference over Time

Inference Tasks in Temporal Models

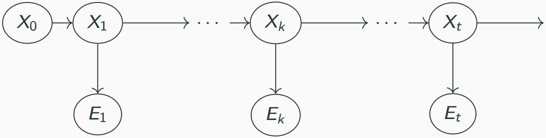
- **Filtering/Monitoring**: $P(X_t|e_{1:t})$
- **Prediction**: $P(X_{t+k}|e_{1:t})$
- **Evidence Likelihood**: $P(e_{1:t})$
- **Smoothing/Hindsight**: $P(X_k|e_{1:t}), \quad 0 \leq k < t$
- **Most likely explanation:** $\arg\max_{x_{1:t}} P(x_{1:t}|e_{1:t})$

Inference Tasks in Temporal Models

- **Filtering/Monitoring**: $P(X_t|e_{1:t})$
- **Prediction**: $P(X_{t+k}|e_{1:t})$
- **Evidence Likelihood**: $P(e_{1:t})$
- **Smoothing/Hindsight**: $P(X_k|e_{1:t}), \quad 0 \leq k < t$
- **Most likely explanation:** $\arg\max_{x_{1:t}} P(x_{1:t}|e_{1:t})$

Smoothing / Hindsight

$$P(X_k|e_{1:t}), \quad 1 \leq k < t$$

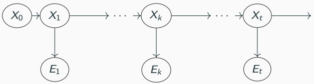


Smoothing/Hindsight

$$P(X_k|e_{1:t}) = P(X_k|e_{1:k}, e_{k+1:t}) \quad \text{(Split Notation)}$$

$$= \alpha P(X_k|e_{1:k})P(e_{k+1:t}|X_k, e_{1:k}) \quad \text{(Bayes)}$$

$$= \alpha \underbrace{P(X_k|e_{1:k})}_{f_{1:k}}\underbrace{P(e_{k+1:t}|X_k)}_{b_{k+1:t}} \quad \text{(Conditional Independence)}$$

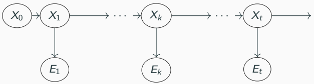


Smoothing/Hindsight

$$P(X_k|e_{1:t}) = P(X_k|e_{1:k}, e_{k+1:t}) \qquad \text{(Split Notation)}$$

$$= \alpha P(X_k|e_{1:k}) P(e_{k+1:t}|X_k, e_{1:k}) \qquad \text{(Bayes)}$$

$$= \alpha \underbrace{P(X_k|e_{1:k})}_{f_{1:k}} \underbrace{P(e_{k+1:t}|X_k)}_{b_{k+1:t}} \qquad \text{(Conditional Independence)}$$

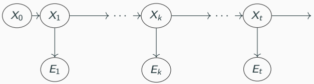


Smoothing/Hindsight

$$P(X_k|e_{1:t}) = P(X_k|e_{1:k}, e_{k+1:t}) \quad \text{(Split Notation)}$$

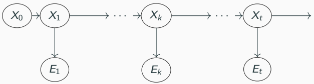
$$= \alpha P(X_k|e_{1:k})P(e_{k+1:t}|X_k, e_{1:k}) \quad \text{(Bayes)}$$

$$= \alpha \underbrace{P(X_k|e_{1:k})}_{f_{1:k}}\underbrace{P(e_{k+1:t}|X_k)}_{b_{k+1:t}} \quad \text{(Conditional Independence)}$$



Smoothing/Hindsight

$$
\begin{aligned}
P(X_k|e_{1:t}) &= P(X_k|e_{1:k}, e_{k+1:t}) && \text{(Split Notation)} \\
&= \alpha P(X_k|e_{1:k})P(e_{k+1:t}|X_k, e_{1:k}) && \text{(Bayes)} \\
&= \alpha \underbrace{P(X_k|e_{1:k})}_{f_{1:k}}\underbrace{P(e_{k+1:t}|X_k)}_{b_{k+1:t}} && \text{(Conditional Independence)}
\end{aligned}
$$



Smoothing: The Backward Message

$$b_{k+1:t} = P(e_{k+1:t}|X_k)$$

Smoothing: The Backward Message

$$b_{k+1:t} = P(e_{k+1:t}|X_k)$$

$$P(e_{k+1}|X_k) = \sum_{x_{k+1}} \underbrace{P(e_{k+1}|x_{k+1})}_{\textit{sensor}} \underbrace{P(e_{k+2:t}|x_{k+1})}_{\textit{recursive}} \underbrace{P(x_{k+1}|X_k)}_{\textit{transition}}$$

Smoothing: The Backward Message

$$b_{k+1:t} = P(e_{k+1:t}|X_k)$$

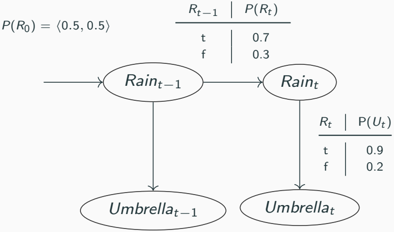
$$P(e_{k+1}|X_k) = \sum_{x_{k+1}} \underbrace{P(e_{k+1}|x_{k+1})}_{\textit{sensor}} \underbrace{P(e_{k+2:t}|x_{k+1})}_{\textit{recursive}} \underbrace{P(x_{k+1}|X_k)}_{\textit{transition}}$$

$$b_{k+1:t} = \text{Backward}(b_{k+2:t}, e_{k+1:t})$$

Example

$$\mathbf{P(R_1|u_1, u_2)}$$

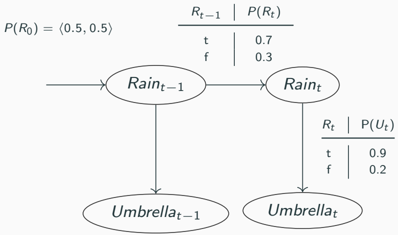
$$P(R_1|u_1, u_2) = \alpha P(R_1|u_1) P(u_2|R_1)$$



Example

$$\mathbf{P(R_1|u_1, u_2)}$$

$$P(R_1|u_1, u_2) = \alpha P(R_1|u_1) P(u_2|R_1)$$

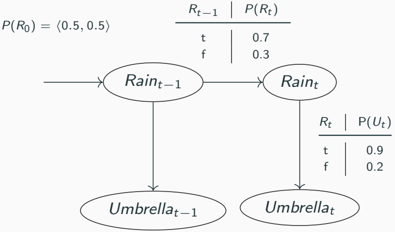


Example

$$\mathbf{P(R_1|u_1, u_2)}$$

$$P(R_1|u_1, u_2) = \alpha P(R_1|u_1) P(u_2|R_1)$$

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



Example

$$\mathbf{P(R_1|u_1, u_2)}$$

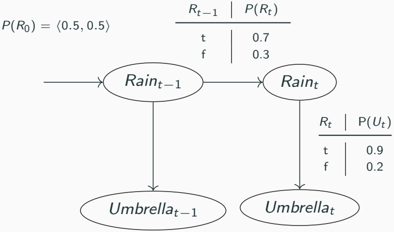
$$P(R_1|u_1, u_2) = \alpha \textcolor{red}{P(R_1|u_1)} P(u_2|R_1)$$

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

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

$$= \langle 0.7, 0.3 \rangle \times 0.5 + \langle 0.3, 0.7 \rangle \times 0.5$$

$$= \langle 0.5, 0.5 \rangle$$



Example

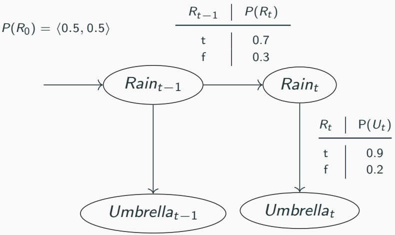
$$\mathbf{P(R_1|u_1, u_2)}$$

$$P(R_1|u_1, u_2) = \alpha P(R_1|u_1) P(u_2|R_1)$$

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

$$P(R_1) = \sum_{r_0} P(R_1|r_0) P(r_0)$$
$$= \langle 0.7, 0.3 \rangle \times 0.5 + \langle 0.3, 0.7 \rangle \times 0.5$$
$$= \langle 0.5, 0.5 \rangle$$

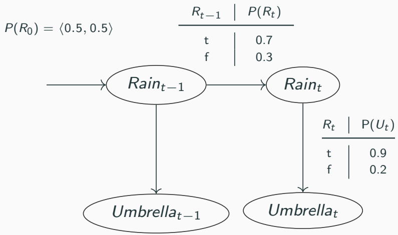
$$P(R_1|u_1) = \alpha P(u_1|R_1) P(R_1)$$
$$= \alpha \langle 0.9, 0.2 \rangle \langle 0.5, 0.5 \rangle$$
$$= \langle 0.818, 0.182 \rangle$$



Example

$$\mathbf{P}(\mathbf{R_1}|\mathbf{u_1}, \mathbf{u_2})$$

$P(R_1|u_1, u_2) = \alpha P(R_1|u_1)P(u_2|R_1)$

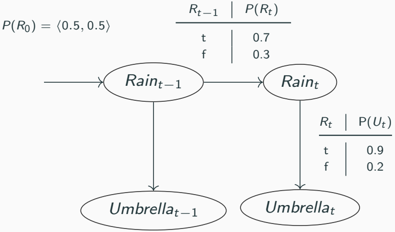


Example

$$\mathbf{P(R_1|u_1, u_2)}$$

$P(R_1|u_1, u_2) = \alpha P(R_1|u_1)P(u_2|R_1)$

$P(u_2|R_1)$



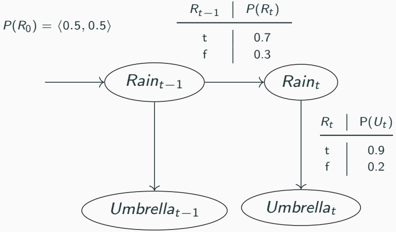
Example

$$\mathbf{P(R_1|u_1, u_2)}$$

$P(R_1|u_1, u_2) = \alpha P(R_1|u_1)P(u_2|R_1)$

$P(u_2|R_1)$

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

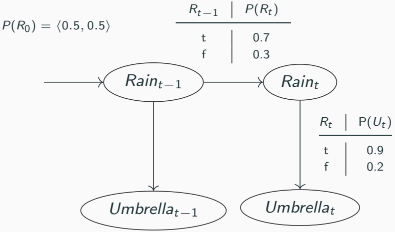


Example

$$\mathbf{P}(\mathbf{R_1}|\mathbf{u_1}, \mathbf{u_2})$$

$$P(R_1|u_1, u_2) = \alpha P(R_1|u_1)P(u_2|R_1)$$

$$P(u_2|R_1)$$
$$= \sum_{r_2} P(u_2|r_2)P(|r_2)P(r_2|R_1)$$
$$= (0.9 \times 1 \times \langle 0.7, 0.3 \rangle) + (0.2 \times 1 \times \langle 0.3, 0.7 \rangle)$$
$$= \langle 0.69, 0.41 \rangle$$



Example

$$\mathbf{P(R_1|u_1, u_2)}$$

$$P(R_1|u_1, u_2) = \alpha P(R_1|u_1)P(u_2|R_1)$$
$$P(R_1|u_1) = \langle 0.818, 0.182\rangle$$
$$P(u_2|R_1) = \langle 0.69, 0.41\rangle$$

$$P(R_1|u_1, u_2) = \alpha\langle 0.818, 0.182\rangle \times \langle 0.69, 0.41\rangle$$
$$= \langle 0.883, 0.117\rangle$$

Example

$$\mathbf{P(R_1|u_1, u_2)}$$

$$P(R_1|u_1, u_2) = \alpha P(R_1|u_1)P(u_2|R_1)$$
$$P(R_1|u_1) = \langle 0.818, 0.182\rangle$$
$$P(u_2|R_1) = \langle 0.69, 0.41\rangle$$

$$P(R_1|u_1, u_2) = \alpha\langle 0.818, 0.182\rangle \times \langle 0.69, 0.41\rangle$$
$$= \langle 0.883, 0.117\rangle$$

Deriving the Backwards Message

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

$$\begin{aligned}
P(e_{k+1:t}|X_k) &= \sum_{x_{k+1}} P(e_{k+1:t}, x_{k+1}|X_k) && \text{marginalisation} \\
&= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}, x_{k+1}|X_k) && \text{(split notation)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|e_{k+2:t}, x_{k+1}, X_k)P(e_{k+2:t}, x_{k+1}|X_k) && \text{(Bayes)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}, x_{k+1}|X_k) && \text{(independence)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1}, X_k)P(x_{k+1}|X_k) && \text{(Bayes)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k) && \text{(independence)}
\end{aligned}$$

Deriving the Backwards Message

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

$$P(e_{k+1:t}|X_k) = \sum_{x_{k+1}} P(e_{k+1:t}, x_{k+1}|X_k) \quad \text{marginalisation}$$

$$= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}, x_{k+1}|X_k) \quad \text{(split notation)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|e_{k+2:t}, x_{k+1}, X_k)P(e_{k+2:t}, x_{k+1}|X_k) \quad \text{(Bayes)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}, x_{k+1}|X_k) \quad \text{(independence)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1}, X_k)P(x_{k+1}|X_k) \quad \text{(Bayes)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k) \quad \text{(independence)}$$

Deriving the Backwards Message

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

$$P(e_{k+1:t}|X_k) = \sum_{x_{k+1}} P(e_{k+1:t}, x_{k+1}|X_k) \quad \text{marginalisation}$$

$$= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}, x_{k+1}|X_k) \quad \text{(split notation)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|e_{k+2:t}, x_{k+1}, X_k) P(e_{k+2:t}, x_{k+1}|X_k) \quad \text{(Bayes)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1}) P(e_{k+2:t}, x_{k+1}|X_k) \quad \text{(independence)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1}) P(e_{k+2:t}|x_{k+1}, X_k) P(x_{k+1}|X_k) \quad \text{(Bayes)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1}) P(e_{k+2:t}|x_{k+1}) P(x_{k+1}|X_k) \quad \text{(independence)}$$

Deriving the Backwards Message

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

$$P(e_{k+1:t}|X_k) = \sum_{x_{k+1}} P(e_{k+1:t}, x_{k+1}|X_k) \quad \text{marginalisation}$$

$$= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}, x_{k+1}|X_k) \quad \text{(split notation)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|e_{k+2:t}, x_{k+1}, X_k)P(e_{k+2:t}, x_{k+1}|X_k) \quad \text{(Bayes)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}, x_{k+1}|X_k) \quad \text{(independence)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1}, X_k)P(x_{k+1}|X_k) \quad \text{(Bayes)}$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k) \quad \text{(independence)}$$

Deriving the Backwards Message

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

$$
\begin{aligned}
P(e_{k+1:t}|X_k) &= \sum_{x_{k+1}} P(e_{k+1:t}, x_{k+1}|X_k) && \text{marginalisation} \\
&= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}, x_{k+1}|X_k) && \text{(split notation)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|e_{k+2:t}, x_{k+1}, X_k)P(e_{k+2:t}, x_{k+1}|X_k) && \text{(Bayes)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}, x_{k+1}|X_k) && \text{(independence)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1}, X_k)P(x_{k+1}|X_k) && \text{(Bayes)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k) && \text{(independence)}
\end{aligned}
$$

Deriving the Backwards Message

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

$$\begin{aligned}
P(e_{k+1:t}|X_k) &= \sum_{x_{k+1}} P(e_{k+1:t}, x_{k+1}|X_k) && \text{marginalisation} \\
&= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}, x_{k+1}|X_k) && \text{(split notation)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|e_{k+2:t}, x_{k+1}, X_k)P(e_{k+2:t}, x_{k+1}|X_k) && \text{(Bayes)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}, x_{k+1}|X_k) && \text{(independence)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1}, X_k)P(x_{k+1}|X_k) && \text{(Bayes)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k) && \text{(independence)}
\end{aligned}$$

Deriving the Backwards Message

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

$$\begin{aligned}
P(e_{k+1:t}|X_k) &= \sum_{x_{k+1}} P(e_{k+1:t}, x_{k+1}|X_k) && \text{marginalisation} \\
&= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}, x_{k+1}|X_k) && \text{(split notation)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|e_{k+2:t}, x_{k+1}, X_k)P(e_{k+2:t}, x_{k+1}|X_k) && \text{(Bayes)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}, x_{k+1}|X_k) && \text{(independence)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1}, X_k)P(x_{k+1}|X_k) && \text{(Bayes)} \\
&= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k) && \text{(independence)}
\end{aligned}$$

Inference Tasks in Temporal Models

- **Filtering/Monitoring**: $P(X_t|e_{1:t})$
- **Prediction**: $P(X_{t+k}|e_{1:t})$
- **Evidence Likelihood**: $P(e_{1:t})$
- **Smoothing/Hindsight**: $P(X_k|e_{1:t}), \quad 0 \leq k < t$
- **Most likely explanation:** $\arg\max_{x_{1:t}} P(x_{1:t}|e_{1:t})$

Inference Tasks in Temporal Models

- **Filtering/Monitoring**: $P(X_t|e_{1:t})$
- **Prediction**: $P(X_{t+k}|e_{1:t})$
- **Evidence Likelihood**: $P(e_{1:t})$
- **Smoothing/Hindsight**: $P(X_k|e_{1:t}), \quad 0 \leq k < t$
- **Most likely explanation:** $\arg\max_{x_{1:t}} P(x_{1:t}|e_{1:t})$

Most likely sequence

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

Say $U_{[1:5]} = [true, true, false, true, true]$

Most likely weather sequence that caused it?

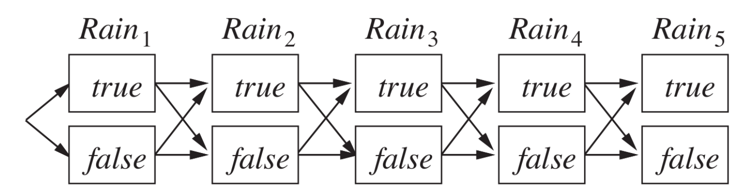
Most likely sequence

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

Naive (Wrong) Approach: Use smoothing again?

- $P(X_1|e_{1:t})$
- $P(X_2|e_{1:t})$
- $P(X_3|e_{1:t})$
- ...

Most Likely Sequence: Viterbi Algorithm



Most Likely Sequence: The Viterbi Algorithm

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

Most Likely Sequence: The Viterbi Algorithm

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

Most Likely Sequence: The Viterbi Algorithm

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

$$= \alpha \underbrace{P(e_{t+1}|X_{t+1})}_{\textit{sensor}} \max_{x_t} \left(\underbrace{P(X_{t+1}|x_t)}_{\textit{transition}} \underbrace{\max_{x_1,\ldots,x_{t-1}} P(x_1,\ldots,x_{t-1}|e_{1:t})}_{\text{recursive}} \right)$$

Most Likely Sequence: The Viterbi Algorithm

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

$$= \alpha \underbrace{P(e_{t+1}|X_{t+1})}_{\textit{sensor}} \max_{x_t}\left(\underbrace{P(X_{t+1}|x_t)}_{\textit{transition}} \underbrace{\max_{x_1,\ldots,x_{t-1}} P(x_1,\ldots,x_{t-1}|e_{1:t})}_{\text{recursive}} \right)$$

$$m_{1:t} = \max_{x_1,\ldots,x_{t-1}} P(x_1,\ldots,x_{t-1},X_t|e_{1:t})$$

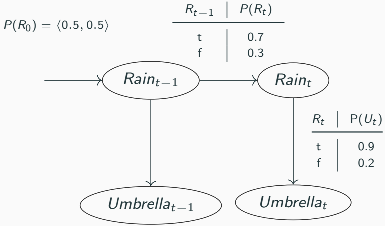
Hidden Markov Model

Hidden Markov Model: Temporal probabilistic model in which state of the process is described by a single variable

$$O = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

$$P_o = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$T = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$



(Bonus) Stationary Distribution of Prediction

$$P_o = \begin{bmatrix} 0.99 \\ 0.01 \end{bmatrix}$$

$$T = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$P(X_0) = P_0 = \langle 0.99, 0.01 \rangle$$

(Bonus) Stationary Distribution of Prediction

$$P_o = \begin{bmatrix} 0.99 \\ 0.01 \end{bmatrix}$$

$$T = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$P(X_0) = P_0 = \langle 0.99, 0.01 \rangle$$

$$P(X_1) = P_0^T T = \langle 0.892, 0.108 \rangle$$

(Bonus) Stationary Distribution of Prediction

$$P_o = \begin{bmatrix} 0.99 \\ 0.01 \end{bmatrix}$$

$$T = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$P(X_0) = P_0 = \langle 0.99, 0.01 \rangle$$

$$P(X_1) = P_0^T T = \langle 0.892, 0.108 \rangle$$

$$P(X_2) = P_0^T TT = \langle 0.813, 0.186 \rangle$$

(Bonus) Stationary Distribution of Prediction

$$P_o = \begin{bmatrix} 0.99 \\ 0.01 \end{bmatrix}$$

$$T = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$P(X_0) = P_0 = \langle 0.99, 0.01 \rangle$$
$$P(X_1) = P_0^T T = \langle 0.892, 0.108 \rangle$$
$$P(X_2) = P_0^T TT = \langle 0.813, 0.186 \rangle$$
$$P(X_5) = P_0^T T^5 = \langle 0.660, 0.339 \rangle$$

(Bonus) Stationary Distribution of Prediction

$$P_o = \begin{bmatrix} 0.99 \\ 0.01 \end{bmatrix}$$

$$T = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$P(X_0) = P_0 = \langle 0.99, 0.01 \rangle$$

$$P(X_1) = P_0^T T = \langle 0.892, 0.108 \rangle$$

$$P(X_2) = P_0^T TT = \langle 0.813, 0.186 \rangle$$

$$P(X_5) = P_0^T T^5 = \langle 0.660, 0.339 \rangle$$

$$\ldots$$

$$P(X_{20}) = P_0^T T^{20} = \langle 0.506, 0.494 \rangle$$

Summary

- Smoothing/Hindsight: The Forward-Backward algorithm
- Most likely Sequence: Viterbi Algorithm
- Hidden Markov Models