Temporal Inference

16-385 Computer Vision (Kris Kitani)

Carnegie Mellon University

Basic Inference Tasks

Filtering

$$P(\boldsymbol{X}_t | \boldsymbol{e}_{1:t})$$

Posterior probability over the **current** state, given all evidence up to present

Prediction

$$P(\boldsymbol{X}_{t+k} | \boldsymbol{e}_{1:t})$$

Posterior probability over a **future** state, given all evidence up to present

Smoothing

$$P(\boldsymbol{X}_k | \boldsymbol{e}_{1:t})$$

Posterior probability over a **past** state, given all evidence up to present

Best Sequence

$$\arg\max_{\boldsymbol{X}_{1:t}} P(\boldsymbol{X}_{1:t} | \boldsymbol{e}_{1:t})$$

Best state sequence given all evidence up to present

Filtering

$$P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

Posterior probability over the **current** state, given all evidence up to present

Where am I now?

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

posterior — observation model — motion model — prior

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto \underbrace{P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1})}_{\text{observation model}} \sum_{\boldsymbol{X}_t} \underbrace{P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)}_{\text{motion model}} \underline{P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})}$$

What is this?

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

same type of ‘message’

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

same type of 'message'

called a **belief distribution**

sometimes people use this annoying notation instead: $Bel(x_t)$

a belief is a reflection of the systems (robot, tracker) knowledge about the state **X**

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

Where does this equation come from?

(scary math to follow…)

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

just splitting up the notation here

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) = P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{t+1}, \boldsymbol{e}_{1:t})$$

Filtering

Can be computed with recursion (Dynamic Programming)

$$\underline{P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1})} \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) \underline{P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})}$$

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) = P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{t+1}, \boldsymbol{e}_{1:t})$$

Apply Bayes' rule (with evidence)

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) = P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{t+1}, \boldsymbol{e}_{1:t})$$

$$= \frac{P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}, \boldsymbol{e}_{1:t}) P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t})}{P(\boldsymbol{e}_{t+1}|\boldsymbol{e}_{1:t})}$$

Apply Markov assumption on observation model

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

$$\begin{aligned} P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) &= P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{t+1}, \boldsymbol{e}_{1:t}) \\ &= \frac{P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}, \boldsymbol{e}_{1:t}) P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t})}{P(\boldsymbol{e}_{t+1}|\boldsymbol{e}_{1:t})} \\ &= \alpha P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t}) \end{aligned}$$

Condition on the previous state $\mathbf{X}_t$

Filtering

Can be computed with recursion (Dynamic Programming)

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

$$\begin{aligned} P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) &= P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{t+1}, \boldsymbol{e}_{1:t}) \\ &= \frac{P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}, \boldsymbol{e}_{1:t}) P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t})}{P(\boldsymbol{e}_{t+1}|\boldsymbol{e}_{1:t})} \\ &= \alpha P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t}) \\ &= \alpha P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t, \boldsymbol{e}_{1:t}) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t}) \end{aligned}$$

Apply Markov assumption on motion model

Filtering

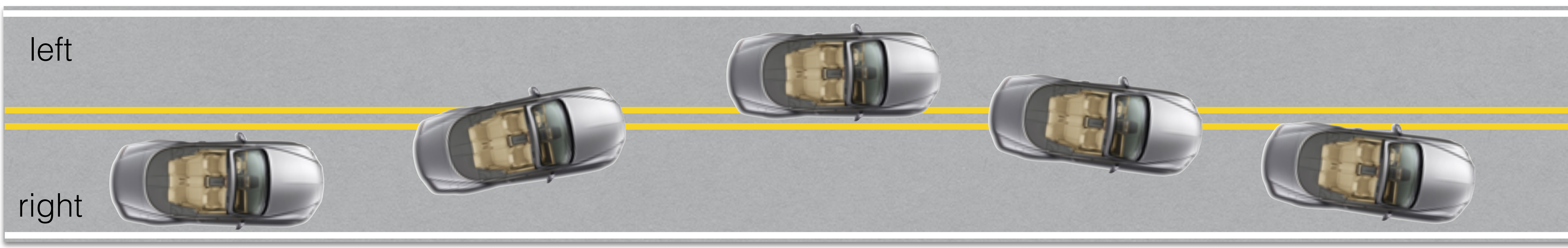
Can be computed with recursion (Dynamic Programming)

$$\underline{P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1})} \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) \underline{P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})}$$

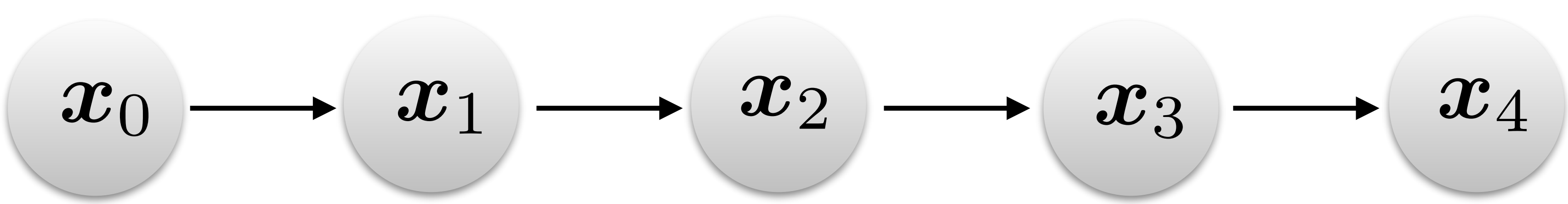
$$\begin{aligned}
P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) &= P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{t+1}, \boldsymbol{e}_{1:t}) \\
&= \frac{P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}, \boldsymbol{e}_{1:t}) P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t})}{P(\boldsymbol{e}_{t+1}|\boldsymbol{e}_{1:t})} \\
&= \alpha P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t}) \\
&= \alpha P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t, \boldsymbol{e}_{1:t}) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t}) \\
&= \alpha P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})
\end{aligned}$$

Hidden Markov Model example

‘In the trunk of a car of a sleepy driver’ model

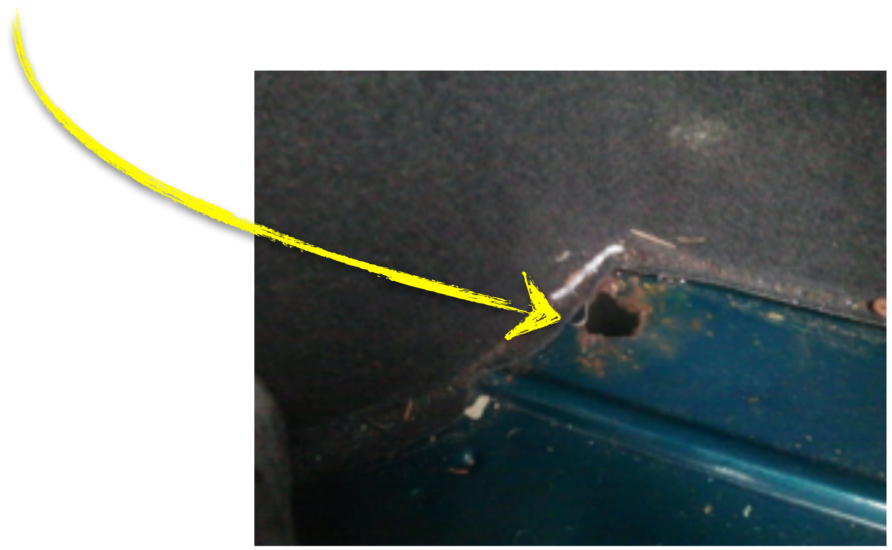


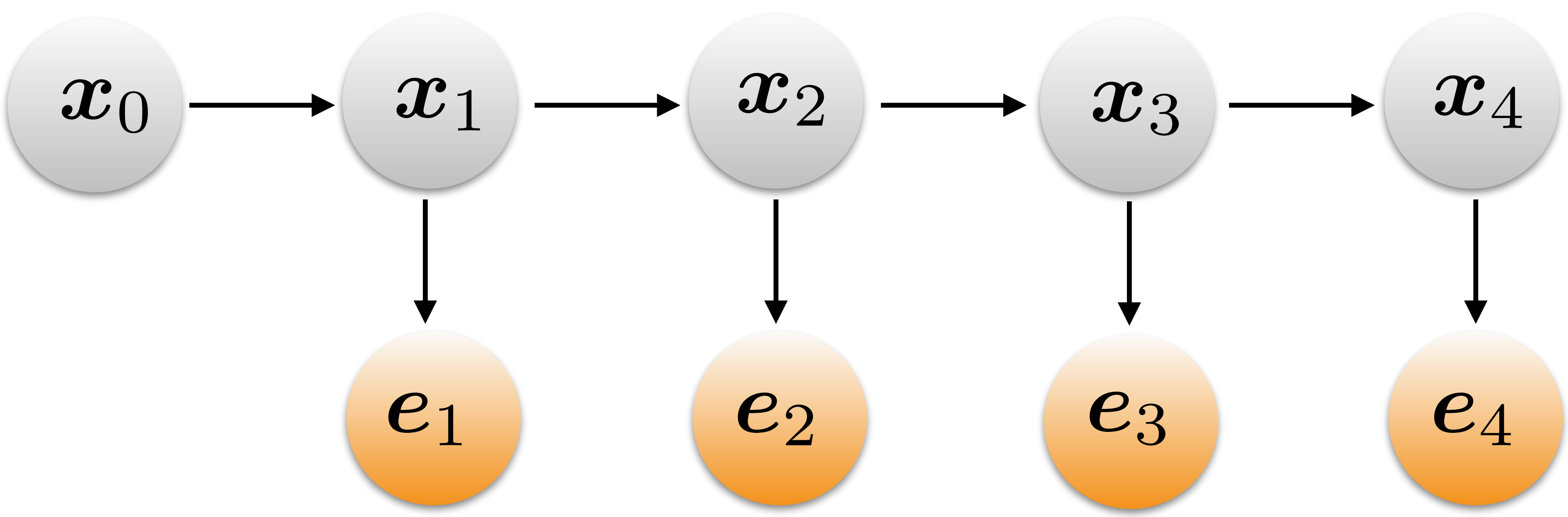
binary random variable (left lane or right lane)



$$\boldsymbol{x} = \{x_{\text{left}}, x_{\text{right}}\}$$

From a hole in the car you can see the ground





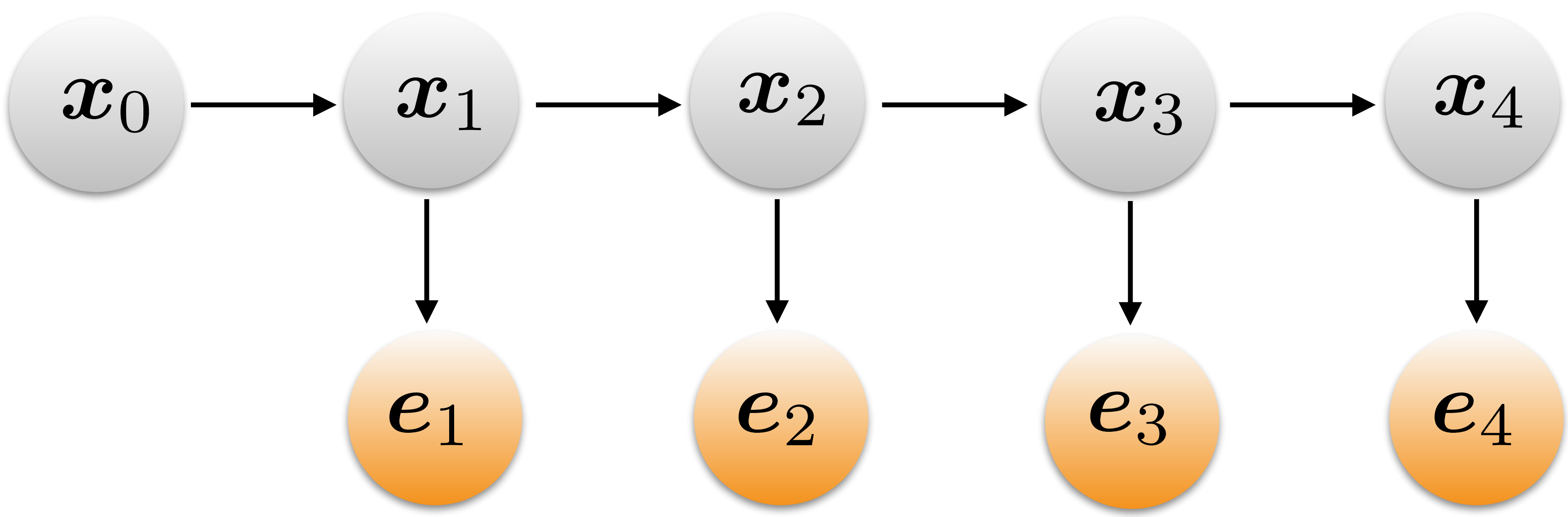
binary random variable (center lane is yellow or road is gray)

$$\boldsymbol{e} = \{e_{\text{gray}}, e_{\text{yellow}}\}$$

	x_{left}	x_{right}
$P(\boldsymbol{x}_0)$	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

What needs to sum to one?



$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

This is filtering!

What's the probability of being in the left lane at t=4?

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filtering: $$P(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1} \mid \boldsymbol{X}_t) P(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$$

*What is the belief distribution if I see **yellow** at t=1* $p(\boldsymbol{x}_1 \mid \boldsymbol{e}_1 = e_{\text{yellow}}) = ?$

Prediction step: $$p(\boldsymbol{x}_1) = \sum_{\boldsymbol{x}_0} p(\boldsymbol{x}_1 \mid \boldsymbol{x}_0) p(\boldsymbol{x}_0)$$

Update step: $$p(\boldsymbol{x}_1 \mid \boldsymbol{e}_1) = \alpha \; p(\boldsymbol{e}_1 \mid \boldsymbol{x}_1) p(\boldsymbol{x}_1)$$

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filtering: $$P(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1} \mid \boldsymbol{X}_t) P(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$$

What is the belief distribution if I see ***yellow*** *at t=1* $p(\boldsymbol{x}_1 \mid \boldsymbol{e}_1 = e_{\text{yellow}}) = ?$

Prediction step:

$$p(\boldsymbol{x}_1) = \sum_{\boldsymbol{x}_0} p(\boldsymbol{x}_1 \mid \boldsymbol{x}_0) p(\boldsymbol{x}_0)$$

$$= [0.7 \;\; 0.3](0.5) + [0.3 \;\; 0.7](0.5)$$

$$= \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \vert \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \vert \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filtering: $$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

*What is the belief distribution if I see **yellow** at t=1* $p(\boldsymbol{x}_1|\boldsymbol{e}_1 = e_{\text{yellow}}) = ?$

Update step: $p(\boldsymbol{x}_1|\boldsymbol{e}_1) = \alpha \; p(\boldsymbol{e}_1|\boldsymbol{x}_1) p(\boldsymbol{x}_1)$

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filtering: $$P(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1} \mid \boldsymbol{X}_t) P(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$$

What is the belief distribution if I see ***yellow*** *at t=1* $p(\boldsymbol{x}_1 \mid \boldsymbol{e}_1 = e_{\text{yellow}}) = ?$

Update step: $$p(\boldsymbol{x}_1 \mid \boldsymbol{e}_1) = \alpha\ p(\boldsymbol{e}_1 \mid \boldsymbol{x}_1) p(\boldsymbol{x}_1)$$

$$= \alpha\ (0.9 \quad 0.2) .* (0.5 \quad 0.5)$$ observed yellow

$$= \alpha \begin{bmatrix} 0.9 & 0.0 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.1 \end{bmatrix}$$

$$\approx \begin{bmatrix} 0.818 \\ 0.182 \end{bmatrix}$$

more likely to be in which lane?

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filtering: $$P(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1} \mid \boldsymbol{X}_t) P(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$$

*What is the belief distribution if I see **yellow** at t=1* $p(\boldsymbol{x}_1 \mid \boldsymbol{e}_1 = e_{\text{yellow}}) = ?$

Summary

Prediction step:

$$p(\boldsymbol{x}_1) = \sum_{\boldsymbol{x}_0} p(\boldsymbol{x}_1 \mid \boldsymbol{x}_0) p(\boldsymbol{x}_0)$$

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

Update step:

$$p(\boldsymbol{x}_1 \mid \boldsymbol{e}_1) = \alpha \; p(\boldsymbol{e}_1 \mid \boldsymbol{x}_1) p(\boldsymbol{x}_1)$$

$$\approx \begin{bmatrix} 0.818 \\ 0.182 \end{bmatrix}$$

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filtering: $$P(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1} \mid \boldsymbol{X}_t) P(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$$

*What if you see **yellow** again at **t=2*** $p(\boldsymbol{x}_2 \mid e_1, e_2) = ?$

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filtering: $$P(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1} \mid \boldsymbol{X}_t) P(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$$

*What if you see **yellow** again at **t=2*** $p(\boldsymbol{x}_2 \mid e_1, e_2) = ?$

Prediction step: $$p(\boldsymbol{x}_2 \mid e_1) = \sum_{\boldsymbol{x}_1} p(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) p(\boldsymbol{x}_1 \mid e_1)$$

Update step: $$p(\boldsymbol{x}_1 \mid e_1, e_2) = \alpha \; p(\boldsymbol{e}_1 \mid \boldsymbol{x}_1) p(\boldsymbol{x}_1)$$

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filering: $$P(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1} \mid \boldsymbol{X}_t) P(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$$

*What if you see **yellow** again at **t=2*** $p(\boldsymbol{x}_2 \mid e_1, e_2) = ?$

Prediction step: $$p(\boldsymbol{x}_2 \mid e_1) = \sum_{\boldsymbol{x}_1} p(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) p(\boldsymbol{x}_1 \mid e_1)$$

$$= \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.818 \\ 0.182 \end{bmatrix} = \begin{bmatrix} 0.627 \\ 0.373 \end{bmatrix}$$

Why does the probability of being in the left lane go down?

$P(\boldsymbol{x}_0)$	x_{left}	x_{right}
	0.5	0.5

$P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$	x_{left}	x_{right}
x_{left}	0.7	0.3
x_{right}	0.3	0.7

$P(\boldsymbol{e}_t \mid \boldsymbol{x}_t)$	x_{left}	x_{right}
e_{yellow}	0.9	0.2
e_{gray}	0.1	0.8

Filtering: $P(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1} \mid \boldsymbol{X}_t) P(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$

*What if you see **yellow** again at **t=2*** $p(\boldsymbol{x}_2 \mid e_1, e_2) = ?$

Update step: $$p(\boldsymbol{x}_2 \mid e_1, e_2) = \alpha \, p(e_2 \mid \boldsymbol{x}_2) p(\boldsymbol{x}_2 \mid e_1)$$

$$= \alpha \begin{bmatrix} 0.9 & 0.0 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.627 \\ 0.373 \end{bmatrix}$$

$$\approx \begin{bmatrix} 0.883 \\ 0.117 \end{bmatrix}$$

Why does the probability of being in the left lane go up?

Basic Inference Tasks

Filtering

$$P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})$$

Posterior probability over the **current** state, given all evidence up to present

Prediction

$$P(\boldsymbol{X}_{t+k}|\boldsymbol{e}_{1:t})$$

Posterior probability over a **future** state, given all evidence up to present

Smoothing

$$P(\boldsymbol{X}_k|\boldsymbol{e}_{1:t})$$

Posterior probability over a **past** state, given all evidence up to present

Best Sequence

$$\arg\max_{\boldsymbol{X}_{1:t}} P(\boldsymbol{X}_{1:t}|\boldsymbol{e}_{1:t})$$

Best state sequence given all evidence up to present

Prediction

$$P(\mathbf{X}_{t+k}|\mathbf{e}_{1:t})$$

Posterior probability over a **future** state, given all evidence up to present

Where am I going?

Prediction

same recursive form as filtering but…

$$P(\boldsymbol{X}_{t+k+1}|\boldsymbol{e}_{1:t}) = \sum_{\boldsymbol{x}_{t+k}} P(\boldsymbol{X}_{t+k+1}|\boldsymbol{x}_{t+k})P(\boldsymbol{x}_{t+k}|\boldsymbol{e}_{1:t})$$

no new evidence!

What happens as you try to predict further into the future?

Prediction

$$P(\boldsymbol{X}_{t+k+1}|\boldsymbol{e}_{1:t}) = \sum_{\boldsymbol{x}_{t+k}} P(\boldsymbol{X}_{t+k+1}|\boldsymbol{x}_{t+k})P(\boldsymbol{x}_{t+k}|\boldsymbol{e}_{1:t})$$

no new evidence

What happens as you try to predict further into the future?

Approaches its 'stationary distribution'

Basic Inference Tasks

Filtering

$$P(\boldsymbol{X}_t | \boldsymbol{e}_{1:t})$$

Posterior probability over the **current** state, given all evidence up to present

Prediction

$$P(\boldsymbol{X}_{t+k} | \boldsymbol{e}_{1:t})$$

Posterior probability over a **future** state, given all evidence up to present

Smoothing

$$P(\boldsymbol{X}_k | \boldsymbol{e}_{1:t})$$

Posterior probability over a **past** state, given all evidence up to present

Best Sequence

$$\arg\max_{\boldsymbol{X}_{1:t}} P(\boldsymbol{X}_{1:t} | \boldsymbol{e}_{1:t})$$

Best state sequence given all evidence up to present

Smoothing

$$P(\mathbf{X}_k | \mathbf{e}_{1:t})$$

Posterior probability over a **past** state, given all evidence up to present

Wait, what did I do yesterday?

Smoothing

some time in the past

$$P(\boldsymbol{X}_k|\boldsymbol{e}_{1:t}) \qquad 1 \leq k < t$$

$$\begin{aligned} P(\boldsymbol{X}_k|\boldsymbol{e}_{1:t}) &= P(\boldsymbol{X}_k|\boldsymbol{e}_{1:k}, \boldsymbol{e}_{k+1:t}) \\ &= \alpha P(\boldsymbol{X}_k|\boldsymbol{e}_{1:k}) P(\boldsymbol{e}_{k+1:t}|\boldsymbol{X}_k, \boldsymbol{e}_{1:k}) \\ &= \alpha P(\boldsymbol{X}_k|\boldsymbol{e}_{1:k}) P(\boldsymbol{e}_{k+1:t}|\boldsymbol{X}_k) \end{aligned}$$

'forward' message

'backward' message

this is just filtering

this is backwards filtering
Let me explain…

Backward message

$$P(\boldsymbol{e}_{k+1:t}|\boldsymbol{X}_k) = \sum_{\boldsymbol{x}_{k+1}} P(\boldsymbol{e}_{k+1:t}|\boldsymbol{X}_k, \boldsymbol{x}_{k+1})P(\boldsymbol{x}_{k+1}|\boldsymbol{X}_k)$$

conditioning

copied from last slide

$$= \sum_{\boldsymbol{x}_{k+1}} P(\boldsymbol{e}_{k+1:t}|\boldsymbol{x}_{k+1})P(\boldsymbol{x}_{k+1}|\boldsymbol{X}_k)$$

Markov Assumption

$$= \sum_{\boldsymbol{x}_{k+1}} P(\boldsymbol{e}_{k+1}, \boldsymbol{e}_{k+2:t}|\boldsymbol{x}_{k+1})P(\boldsymbol{x}_{k+1}|\boldsymbol{X}_k)$$

split

$$= \sum_{\boldsymbol{x}_{k+1}} P(\boldsymbol{e}_{k+1}|\boldsymbol{x}_{k+1})P(\boldsymbol{e}_{k+2:t}|\boldsymbol{x}_{k+1})P(\boldsymbol{x}_{k+1}|\boldsymbol{X}_k)$$

observation model — recursive message — motion model

This is just a 'backwards' version of filtering where

initial message $P(\boldsymbol{e}_{t-1:t}|\boldsymbol{X}_t) = \mathbf{1}$

Basic Inference Tasks

Filtering

$$P(\boldsymbol{X}_t|\mathbf{e}_{1:t})$$

Posterior probability over the **current** state, given all evidence up to present

Prediction

$$P(\boldsymbol{X}_{t+k}|\mathbf{e}_{1:t})$$

Posterior probability over a **future** state, given all evidence up to present

Smoothing

$$P(\boldsymbol{X}_k|\mathbf{e}_{1:t})$$

Posterior probability over a **past** state, given all evidence up to present

Best Sequence

$$\arg\max_{\boldsymbol{X}_{1:t}} P(\boldsymbol{X}_{1:t}|\mathbf{e}_{1:t})$$

Best state sequence given all evidence up to present

Best Sequence

$$\arg\max_{\boldsymbol{X}_{1:t}} P(\boldsymbol{X}_{1:t} | \boldsymbol{e}_{1:t})$$

Best state sequence given all evidence up to present

I must have done something right, right?

Best Sequence

'Viterbi Algorithm'

$$\max_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t} P(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t,\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1})$$

$$= \alpha P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \max_{\boldsymbol{x}_t} \left[P(\boldsymbol{X}_{t+1}|\boldsymbol{x}_t) \underbrace{\max_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{t-1}} P(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{t-1},\boldsymbol{X}_t|\boldsymbol{e}_{1:t})}_{\text{recursive message}} \right]$$

Identical to filtering but with a max operator

Recall: Filtering equation

$$P(\boldsymbol{X}_{t+1}|\boldsymbol{e}_{1:t+1}) \propto P(\boldsymbol{e}_{t+1}|\boldsymbol{X}_{t+1}) \sum_{\boldsymbol{X}_t} P(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) \underbrace{P(\boldsymbol{X}_t|\boldsymbol{e}_{1:t})}_{\text{recursive message}}$$

Now you know how to answer all the important questions in life:

Where am I now?

Where am I going?

Wait, what did I do yesterday?

I must have done something right, right?