# Lecture 12: Case Studies
# CS486/686 Intro to Artificial Intelligence

2023-6-20

Pascal Poupart
David R. Cheriton School of Computer Science

UNIVERSITY OF WATERLOO

# Outline

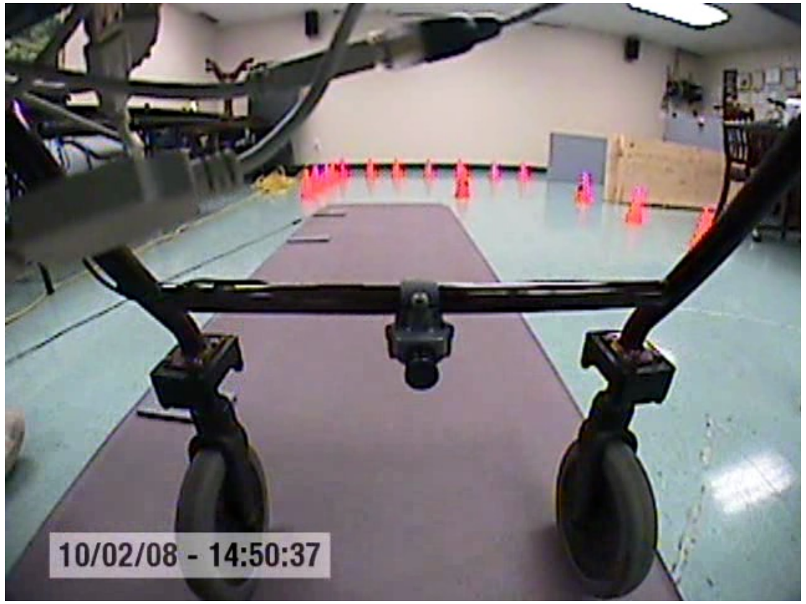
- Case Study 1: Activity Recognition
  - Farheen Omar, Mathieu Sinn, Jakub Truszkowski, Pascal Poupart, James Tung and Allan Caine (2010) **Comparative Analysis of Probabilistic Models for Activity Recognition with an Instrumented Walker**, *UAI*.
- Case Study 2: Sports Analytics
  - Xiangyu Sun, Oliver Schulte, Guiliang Liu, Pascal Poupart (2023) **NTS-NOTEARS: Learning Non-parametric DBNs with Prior Knowledge**, *AISTATS*.

# Case Study 1: Activity Recognition

- Task: infer activities performed by a user of a smart walker
  - Inputs: sensor measurements
  - Output: activity

Backward view


Forward view






UNIVERSITY OF WATERLOO

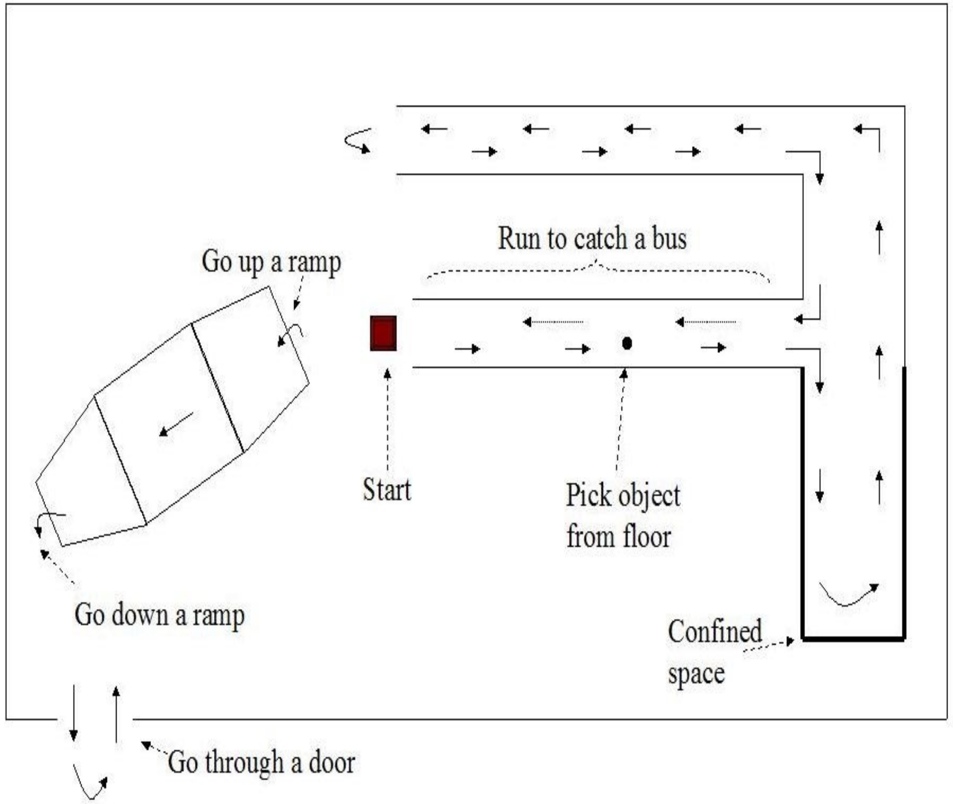
# Inputs: Raw Sensor Data

- 8 channels:
  - Forward acceleration
  - Lateral acceleration
  - Vertical acceleration
  - Load on left rear wheel
  - Load on right rear wheel
  - Load on left front wheel
  - Load on right front wheel
  - Wheel rotation counts (speed)

- Data recorded at 50 Hz and digitized (16 bits)

# Data Collection

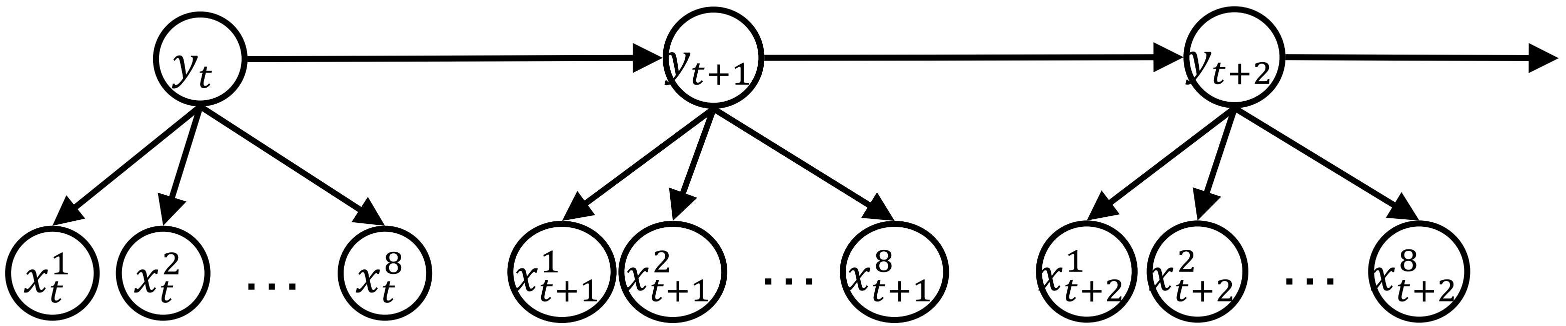
- 8 walker users at Winston Park (84-97 years old)
- 12 older adults (80-89 years old) in the KW area who do not use walkers




Output: Activities

- Not Touching Walker (NTW)
- Standing (ST)
- Walking Forward (WF)
- Turning Left (TL)
- Turning Right (TR)
- Walking Backwards (WB)
- Sitting on the Walker (SW)
- Reaching Tasks (RT)
- Up Ramp/Curb (UR/UC)
- Down Ramp/Curb (DR/DC)

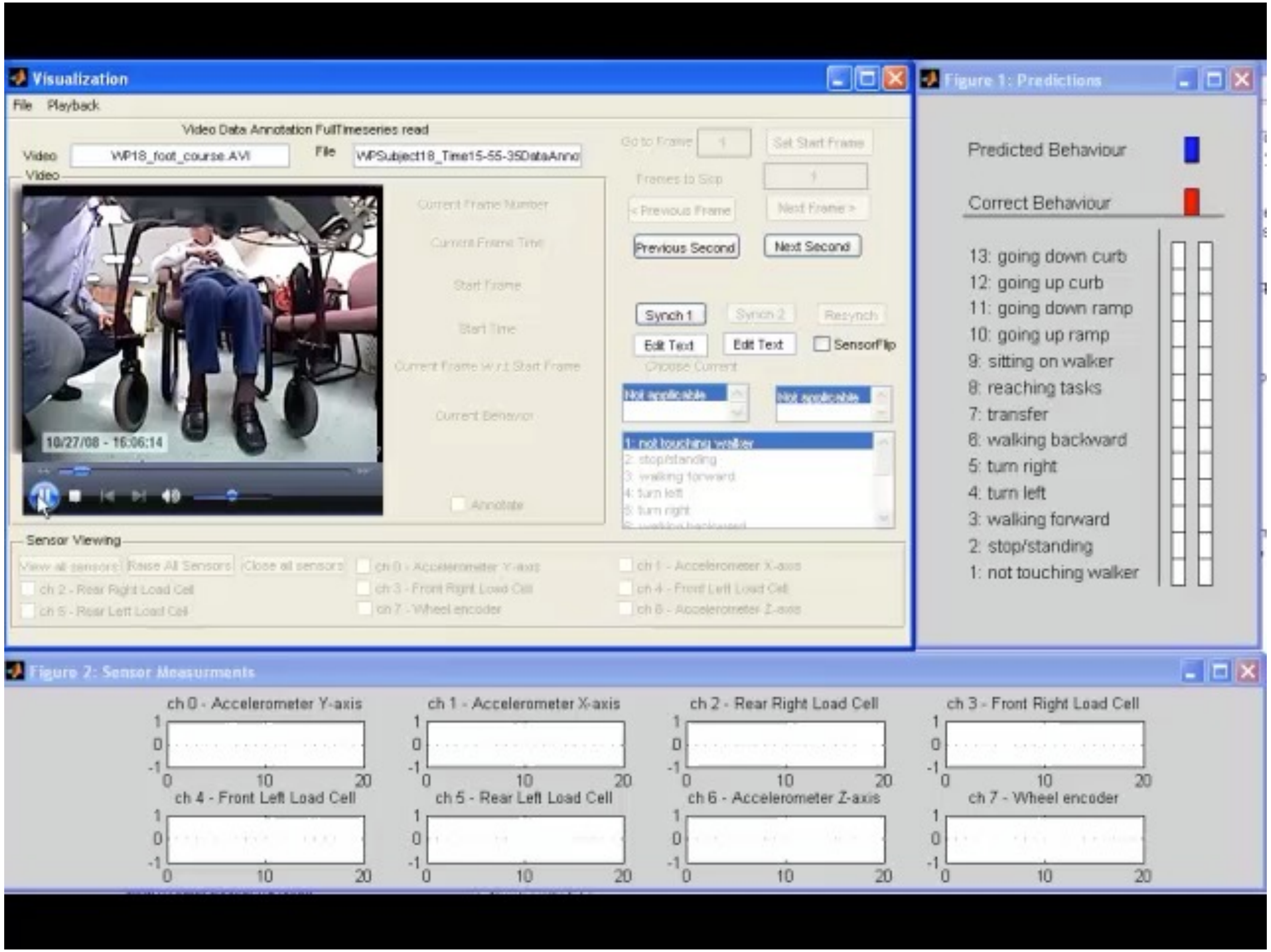
# Hidden Markov Model (HMM)




- Parameters
  - Initial state distribution: $\pi_{class} = \Pr(y_1 = class)$
  - Transition probabilities: $\theta_{class'|class} = \Pr(y_{t+1} = class'|y_t = class)$
  - Emission probabilities: $\phi^i_{val|class} = \Pr(x_t^i = val|y_t = class)$

    or $N(val|\mu^i_{class}, \sigma^i_{class}) = \Pr(x_t^i = val|y_t = class)$
- Maximum likelihood:
  - Supervised: $\pi^*, \theta^*, \phi^* = argmax_{\pi,\theta,\phi} \Pr(y_{1:T}, x_{1:T}|\pi, \theta, \phi)$
  - Unsupervised: $\pi^*, \theta^*, \phi^* = argmax_{\pi,\theta,\phi} \Pr(x_{1:T}|\pi, \theta, \phi)$

# Demo


Visualization
File Playback
Video Data Annotation FullTimeseries read
Video
WP18_foot_course.AVI
File
Previous Second
Next Second
Synch 1
Edit Text
Edit Text
SensorFlip
10/27/08 - 16:06:14
Sensor Viewing
Figure 1: Predictions
Predicted Behaviour
Correct Behaviour
13: going down curb
12: going up curb
11: going down ramp
10: going up ramp
9: sitting on walker
8: reaching tasks
7: transfer
6: walking backward
5: turn right
4: turn left
3: walking forward
2: stop/standing
1: not touching walker
Figure 2: Sensor Measurments
ch 0 - Accelerometer Y-axis
ch 1 - Accelerometer X-axis
ch 2 - Rear Right Load Cell
ch 3 - Front Right Load Cell
ch 4 - Front Left Load Cell
ch 5 - Rear Left Load Cell
ch 6 - Accelerometer Z-axis
ch 7 - Wheel encoder

# Maximum Likelihood

- Supervised Learning: $y$'s are known
- Objective: $argmax_{\pi,\theta,\phi} \Pr(y_{1..t}, x_{1..t}|\pi, \theta, \phi)$

- Derivation:
  - Set derivative to 0
  - Isolate parameters $\pi, \theta, \phi$

- Consider a single input $x$ per time step
- Let $y \in \{c_1, c_2\}$ and $x \in \{v_1, v_2\}$

# Multinomial Emissions

- Let $\#c_i^{start}$ be # times of that process **starts** in class $c_i$
- Let $\#c_i$ be # of times that process is in class $c_i$
- Let $\#(c_i, c_j)$ be # of times that $c_i$ follows $c_j$
- Let $\#(v_i, c_j)$ be # of times that $v_i$ occurs with $c_j$
- $\Pr(y_{0..t}, x_{1..t})$

$$= \Pr(y_0) \prod_{i=1}^{t} \Pr(y_i | y_{i-1}) \Pr(x_i | y_i)$$

$$= (\pi_{c_1})^{\#c_1^{start}} (1 - \pi_{c_1})^{\#c_2^{start}} (\theta_{c_1|c_1})^{\#(c_1,c_1)} (1 - \theta_{c_1|c_1})^{\#(c_2,c_1)} (\theta_{c_1|c_2})^{\#(c_1,c_2)} (1 - \theta_{c_1|c_2})^{\#(c_2,c_2)}$$

$$(\phi_{v_1|c_1})^{\#(v_1,c_1)} (1 - \phi_{v_1|c_1})^{\#(v_2,c_1)} (\phi_{v_1|c_2})^{\#(v_1,c_2)} (1 - \phi_{v_1|c_2})^{\#(v_2,c_2)}$$

# Multinomial Emissions

- $argmax_{\pi,\theta,\phi} \Pr(y_{1..t}, x_{1..t} | \pi, \theta, \phi)$

$$\Rightarrow \begin{cases} argmax_{\pi_{c_1}} (\pi_{c_1})^{\#c_1^{start}} (1 - \pi_{c_1})^{\#c_2^{start}} \\ argmax_{\theta_{c_1|c_1}} (\theta_{c_1|c_1})^{\#(c_1,c_1)} (1 - \theta_{c_1|c_1})^{\#(c_2,c_1)} \\ argmax_{\theta_{c_1|c_2}} (\theta_{c_1|c_2})^{\#(c_1,c_2)} (1 - \theta_{c_1|c_2})^{\#(c_2,c_2)} \\ argmax_{\phi_{v_1|c_1}} (\phi_{v_1|c_1})^{\#(v_1,c_1)} (1 - \phi_{v_1|c_1})^{\#(v_2,c_1)} \\ argmax_{\phi_{v_1|c_2}} (\phi_{v_1|c_2})^{\#(v_1,c_2)} (1 - \phi_{v_1|c_2})^{\#(v_2,c_2)} \end{cases}$$

# Multinomial Emissions

- Optimization problem:

$$\underset{\pi_{c_1}}{\operatorname{argmax}} \left(\pi_{c_1}\right)^{\#c_1^{start}} \left(1-\pi_{c_1}\right)^{\#c_2^{start}}$$

$$= \underset{\pi_{c_1}}{\operatorname{argmax}} \ (\#c_1^{start})\log(\pi_{c_1}) + (\#c_2^{start})\log(1-\pi_{c_1})$$

- Set derivative to 0:

$$0 = \frac{\#c_1^{start}}{\pi_{c_1}} - \frac{\#c_2^{start}}{1-\pi_{c_1}}$$

$$\Longrightarrow (1-\pi_{c_1})(\#c_1^{start}) = \left(\pi_{c_1}\right)(\#c_2^{start})$$

$$\Longrightarrow \pi_{c_1} = \frac{\#c_1^{start}}{\#c_1^{start}+\#c_2^{start}}$$

# Relative Frequency Counts

- Maximum likelihood solution

$$\pi_{c_1^{start}} = \#c_1^{start}/(\#c_1^{start} + \#c_2^{start})$$

$$\theta_{c_1|c_1} = \#(c_1, c_1)/(\#(c_1, c_1) + \#(c_2, c_1))$$

$$\theta_{c_1|c_2} = \#(c_1, c_2)/(\#(c_1, c_2) + \#(c_2, c_2))$$

$$\phi_{v_1|c_1} = \#(v_1, c_1)/(\#(v_1, c_1) + \#(v_2, c_1))$$

$$\phi_{v_1|c_2} = \#(v_1, c_2)/(\#(v_1, c_2) + \#(v_2, c_2))$$

# Gaussian Emissions

- Maximum likelihood solution

$$\pi_{c_1^{start}} = \#c_1^{start} / (\#c_1^{start} + \#c_2^{start})$$

$$\theta_{c_1|c_1} = \#(c_1, c_1)/(\#(c_1, c_1) + \#(c_2, c_1))$$

$$\theta_{c_1|c_2} = \#(c_1, c_2)/(\#(c_1, c_2) + \#(c_2, c_2))$$

$$\mu_{c_1} = \frac{1}{\#c_1} \sum_{\{t|y_t = c_1\}} x_t, \qquad \sigma_{c_1}^2 = \frac{1}{\#c_1} \sum_{\{t|y_t = c_1\}} (x_t - \mu_{c_1})^2$$

$$\mu_{c_2} = \frac{1}{\#c_2} \sum_{\{t|y_t = c_2\}} x_t, \qquad \sigma_{c_2}^2 = \frac{1}{\#c_2} \sum_{\{t|y_t = c_2\}} (x_t - \mu_{c_2})^2$$

# Case Study 2: Sports Analytics

- Xiangyu Sun, Oliver Schulte, Guiliang Liu, Pascal Poupart (2023) **NTS-NOTEARS: Learning Non-parametric DBNs with Prior Knowledge**, *AISTATS*.

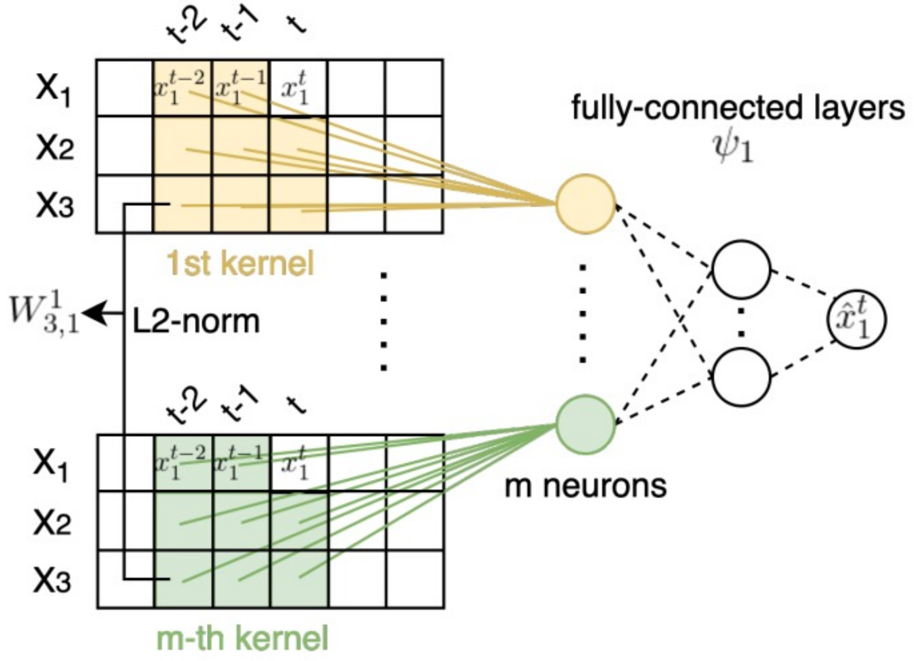
- New technique to learn the structure of a Dynamic Bayesian Network

- Application: Professional Hockey (NHL 2018-2019)
  - What are the factors contributing to a goal?

# Random Variables and Dependencies

| Variables | Type | Range |
|---|---|---|
| time remaining in seconds | continuous | [0, 3600] |
| adjusted x coordinate of puck | continuous | [-100, 100] |
| adjusted y coordinate of puck | continuous | [-42.5, 42.5] |
| score differential | categorical | $(-\infty, +\infty)$ |
| manpower situation | categorical | {short handed, even strength, power play} |
| x velocity of puck | continuous | $(-\infty, +\infty)$ |
| y velocity of puck | continuous | $(-\infty, +\infty)$ |
| event duration | continuous | $[0, +\infty)$ |
| angle between puck and net | continuous | $[-\pi, +\pi]$ |
| home team taking possession | binary | {true, false} |
| shot | binary | {true, false} |
| goal | binary | {true, false} |

Dependencies modeled by convolutional neural networks





UNIVERSITY OF WATERLOO

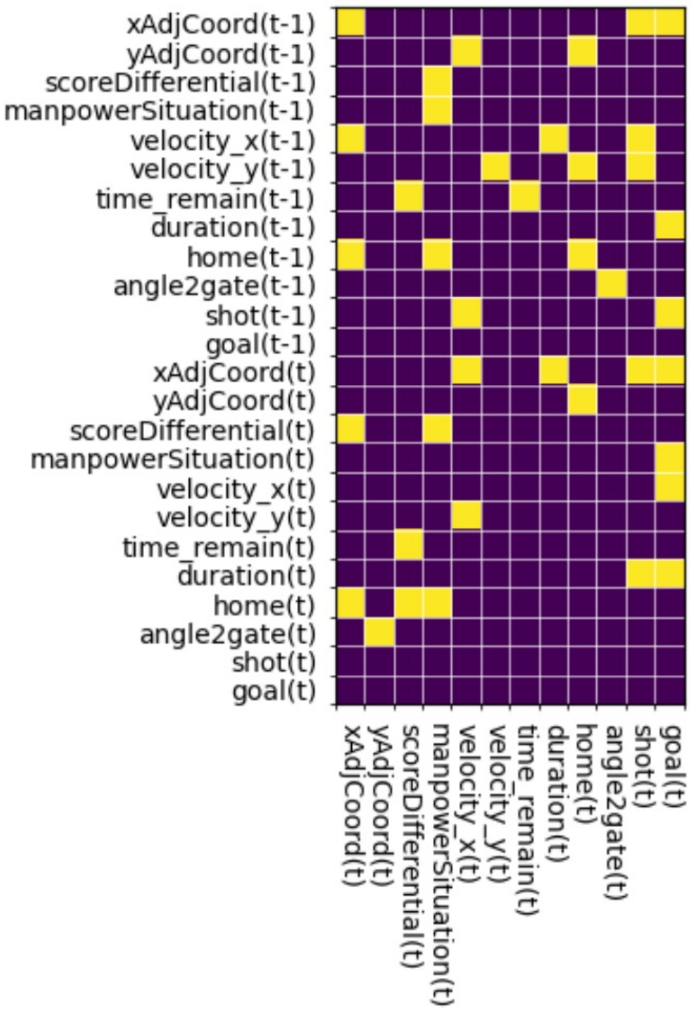
# Objective and Acyclicity Constraint

$$\min_{\theta} F(\theta)$$
$$\textit{subject to } h(W^{K+1}) = 0$$

where

$$F(\theta) = \frac{1}{T-K} \cdot$$
$$\sum_{t=K+1}^{T} \sum_{j=1}^{d} \mathcal{L}(X_j^t, CNN_{\theta_j}(\{\boldsymbol{X}^{t-k} : 1 \leq k \leq K\}, \boldsymbol{X}_{-j}^t))$$
$$+ \sum_{k=1}^{K+1} \lambda_1^k \cdot ||\phi_j^k||_{L^1} + \frac{1}{2}\lambda_2 \cdot ||\theta_j||_{L^2}$$

$$h(W^{K+1}) = tr(e^{W^{K+1} \circ W^{K+1}}) - d = 0$$






UNIVERSITY OF WATERLOO