Modeling time series with hidden Markov models

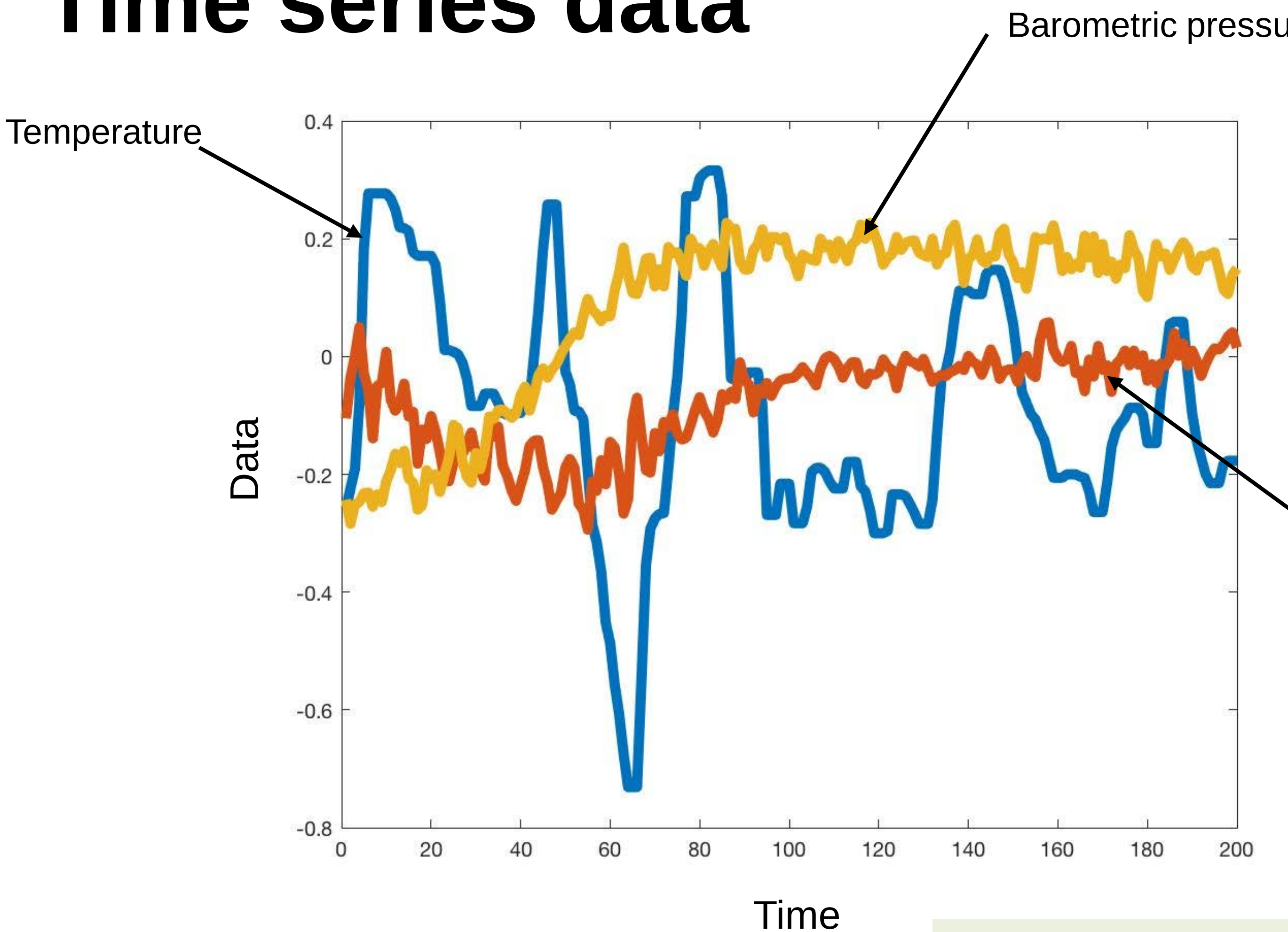
Advanced Machine learning

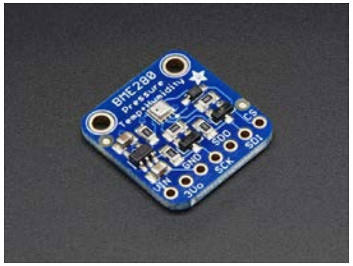
2017

Nadia Figueroa, Jose Medina and Aude Billard

LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Time series data





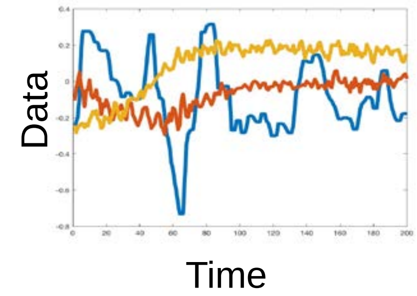
What‘s going on here?

LASA
Learning Algorithms and Systems Laboratory

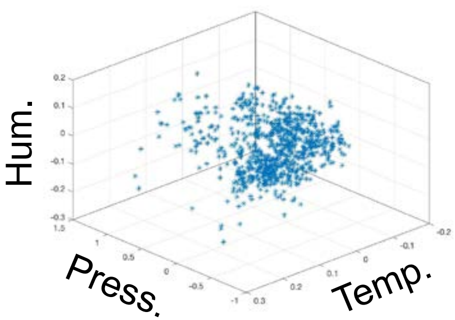
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Time series data

What's the problem setting?

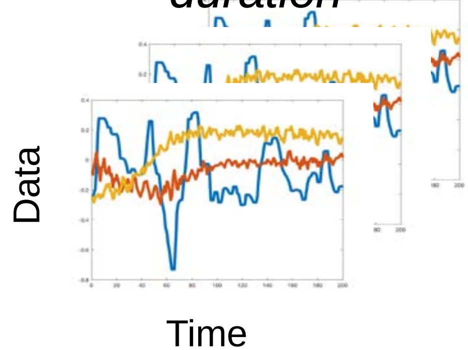


We don't care about time …



$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$$

We have several trajectories with identical duration

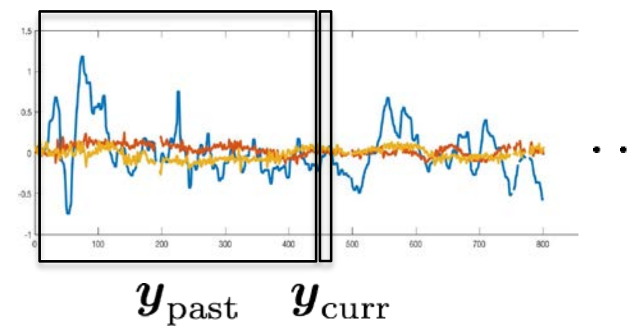


Explicit time dependency

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}, t)$$

$$\boldsymbol{y} = \boldsymbol{f}(t)$$

We have unstructured trajectory(ies)!



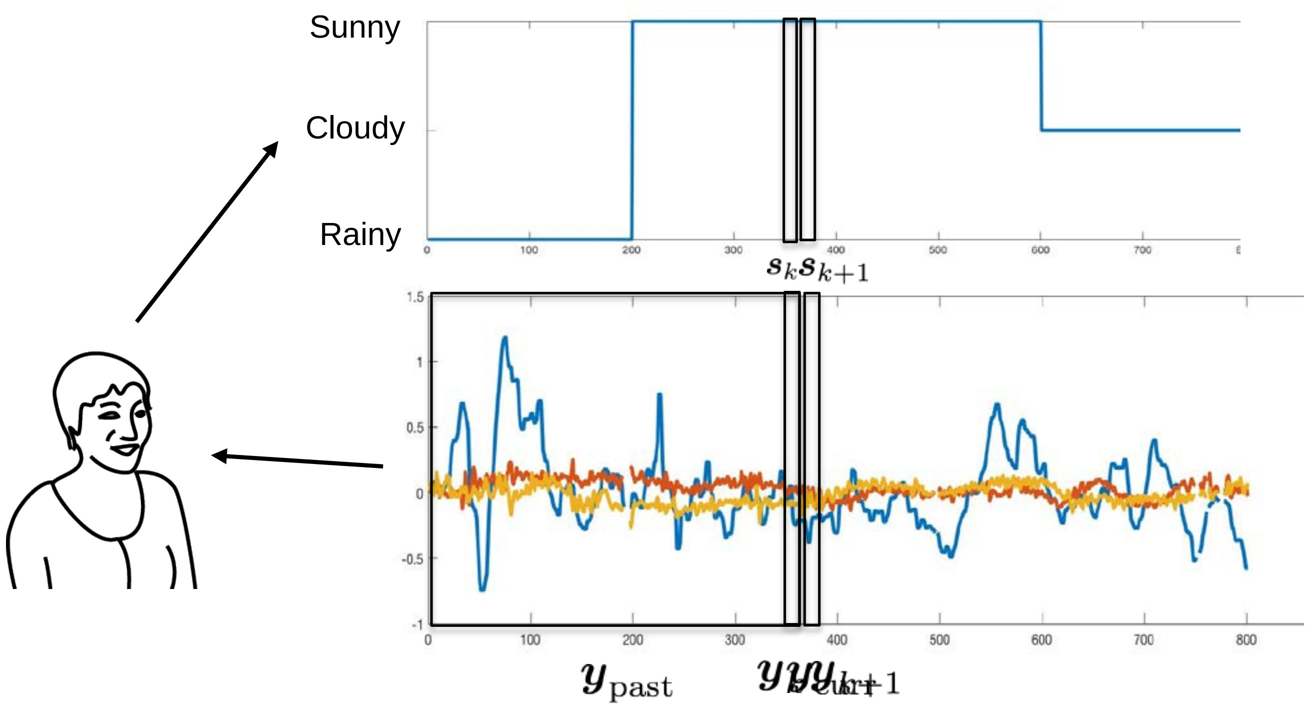
Consider dependency on the past

$$\boldsymbol{y}_{\text{curr}} = \boldsymbol{f}(\boldsymbol{y}_{\text{past}})$$

Too complex!

Unstructured time series data

How to simplify this problem?



. . .

$$s_{k+1} = f_{\text{dyn}}(s_k)$$

. . .

Consider dependency on the past

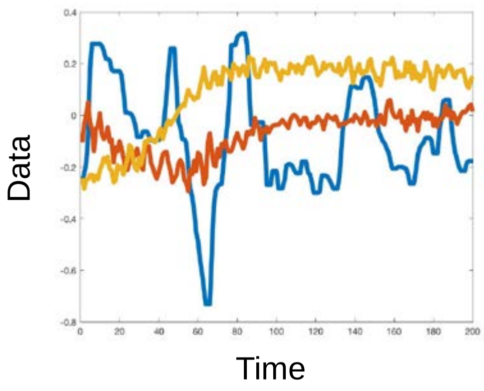
$$y_{k+1} = f(y_{\text{past}})$$

Markov assumption

Outline

First part (10:15 – 11:00):

- Recap on Markov chains
- Hidden Markov Model (HMM)
  - Recognition of time series
  - ML Parameter estimation



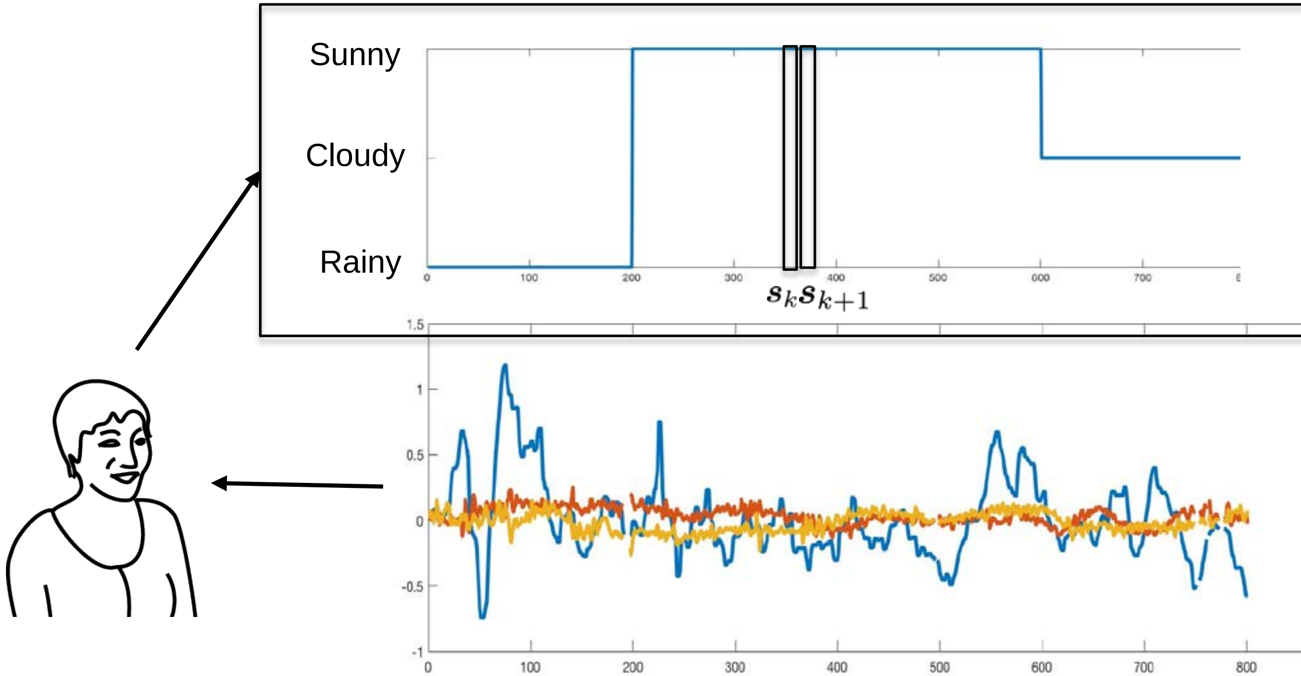
Second part (11:15 – 12:00):

- Time series segmentation
- Bayesian nonparametrics for HMMs

https://github.com/epfl-lasa/ML_toolbox

LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline first part



. . .

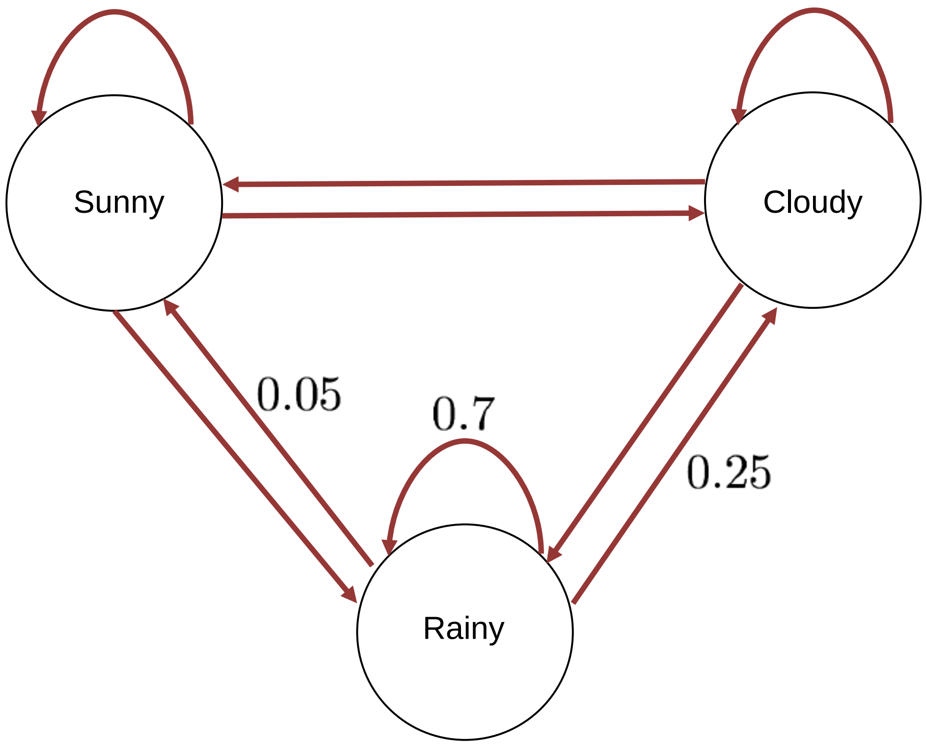
$$s_{k+1} = f_{\text{dyn}}(s_k)$$

. . .

$$y_k = f(s_k)$$

Markov chains

$$\boldsymbol{s}_{k+1} = \boldsymbol{f}_{\text{dyn}}(\boldsymbol{s}_k)$$



$$P(\boldsymbol{s}_{k+1} = \text{Sunny} | \boldsymbol{s}_k = \text{Rainy}) = 0.05$$
$$P(\boldsymbol{s}_{k+1} = \text{Rainy} | \boldsymbol{s}_k = \text{Rainy}) = 0.7$$
$$P(\boldsymbol{s}_{k+1} = \text{Cloudy} | \boldsymbol{s}_k = \text{Rainy}) = 0.25$$

$$\lambda =$$

Transition matrix

$$\begin{array}{cc} & \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{array}{c} \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} & \begin{bmatrix} 0.9 & 0.08 & 0.02 \\ 0.25 & 0.7 & 0.05 \\ 0.05 & 0.25 & 0.7 \end{bmatrix} \end{array}$$

Initial probabilities

$$\begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array}$$
$$\begin{bmatrix} 0.33 & 0.33 & 0.33 \end{bmatrix}$$

Likelihood of a Markov chain

$O =$ Sunny $\rightarrow$ Sunny $\rightarrow$ Cloudy

$\lambda =$

$$P(O|\lambda)?$$

Transition matrix

$$\begin{array}{c c} & \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{array}{c} \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} & \begin{bmatrix} 0.9 & 0.08 & 0.02 \\ 0.25 & 0.7 & 0.05 \\ 0.05 & 0.25 & 0.7 \end{bmatrix} \end{array}$$

$$P(\boldsymbol{s}_0 = \text{Sunny}, \boldsymbol{s}_1 = \text{Sunny}, \boldsymbol{s}_2 = \text{Cloudy}) =$$
$$P(\boldsymbol{s}_0 = \text{Sunny}) P(\boldsymbol{s}_1 = \text{Sunny} | \boldsymbol{s}_0 = \text{Sunny})$$
$$P(\boldsymbol{s}_2 = \text{Cloudy} | \boldsymbol{s}_1 = \text{Sunny}) = 0.33 \cdot 0.9 \cdot 0.08$$

$$P(O|\lambda) = P(\boldsymbol{s}_0) \prod_{k=1}^{T} P(\boldsymbol{s}_k | \boldsymbol{s}_{k-1})$$

Initial probabilities

$$\begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array}$$
$$\begin{bmatrix} 0.33 & 0.33 & 0.33 \end{bmatrix}$$

LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Learning Markov chains

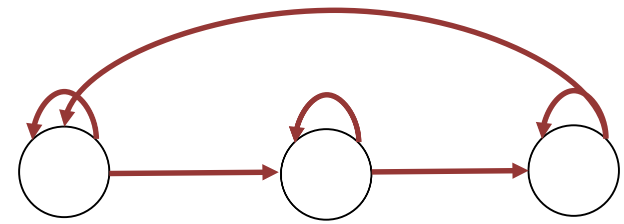
$O =$ Sunny $\longrightarrow$ Sunny $\longrightarrow$ Cloudy

$$\max_{\lambda} \quad \log P(O|\lambda)$$

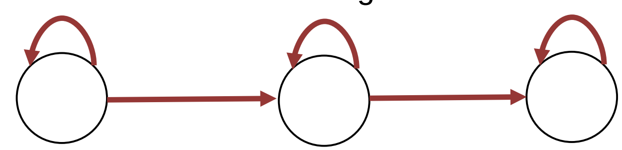
$$\hat{P}(\boldsymbol{s}_{k+1} = \text{Sunny}|\boldsymbol{s}_k = \text{Rainy}) = \frac{\text{number of times we transition Rainy} \rightarrow \text{Sunny}}{\text{number of times we observe Sunny}}$$

Topologies

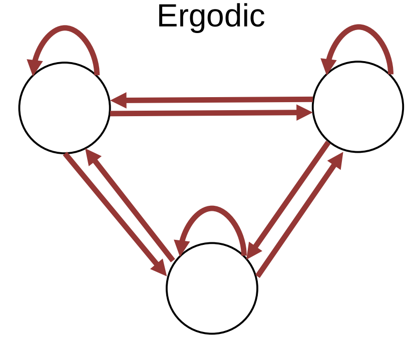
Periodic



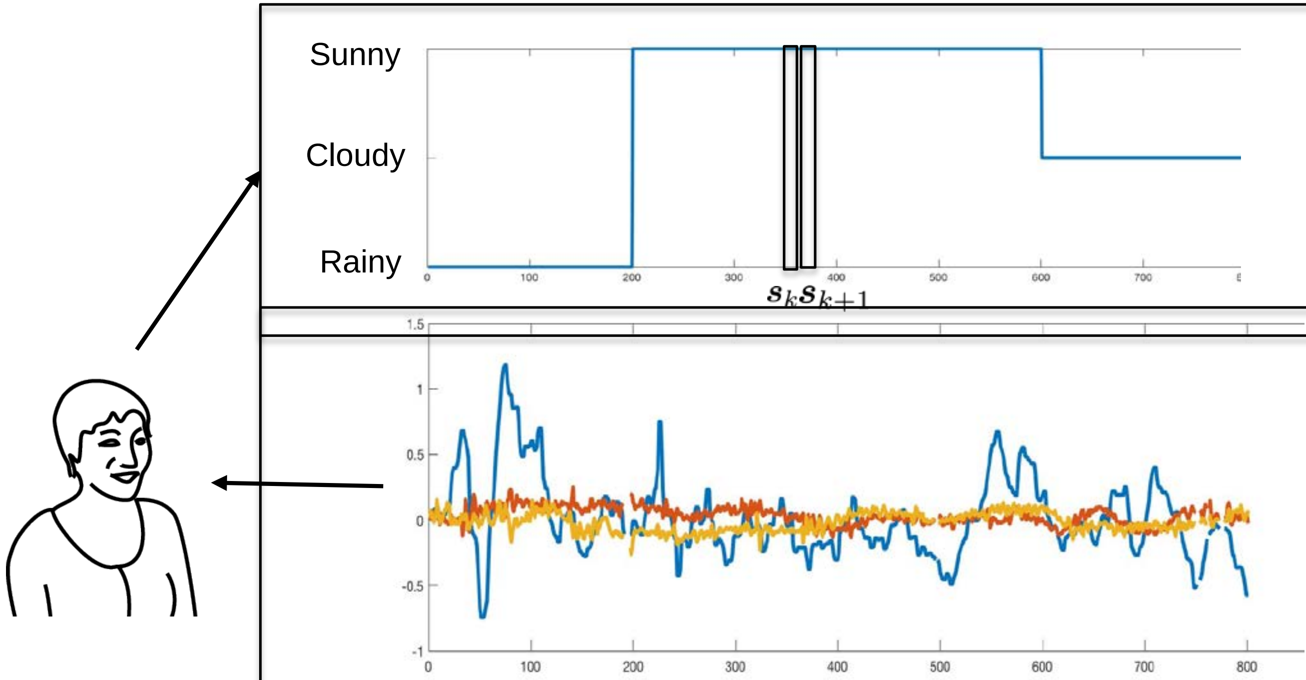
Left-to-right



Ergodic



Outline first part



$$s_{k+1} = f_{\text{dyn}}(s_k)$$

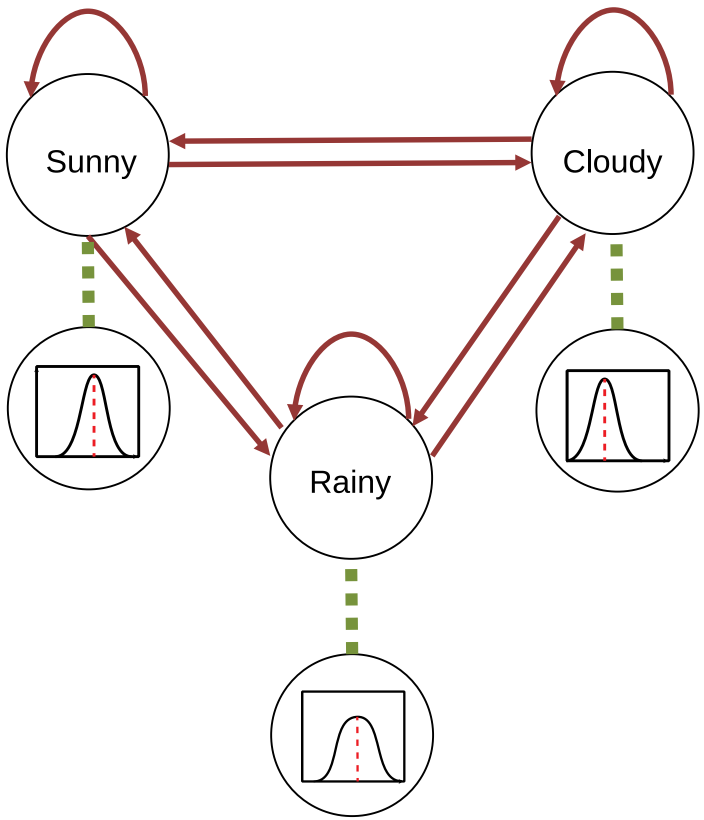
. . .

. . .

$$y_k = f(s_k)$$

Hidden Markov model

$$\boldsymbol{s}_{k+1} = \boldsymbol{f}_{\text{dyn}}(\boldsymbol{s}_k)$$



Transition matrix

$$\begin{array}{c} \\ \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{bmatrix} 0.9 & 0.08 & 0.02 \\ 0.25 & 0.7 & 0.05 \\ 0.05 & 0.25 & 0.7 \end{bmatrix} \end{array}$$

Initial probabilities

$$\begin{array}{c} \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{bmatrix} 0.33 & 0.33 & 0.33 \end{bmatrix} \end{array}$$

$$\boldsymbol{y}_k = \boldsymbol{f}(\boldsymbol{s}_k)$$

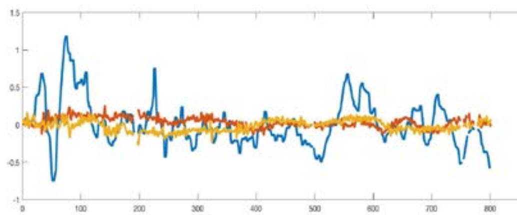
$$P(\boldsymbol{y}_k | \boldsymbol{s}_k = \text{Sunny}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Sunny}}, \boldsymbol{\Sigma}_{\text{Sunny}})$$
$$P(\boldsymbol{y}_k | \boldsymbol{s}_k = \text{Cloudy}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Cloudy}}, \boldsymbol{\Sigma}_{\text{Cloudy}})$$
$$P(\boldsymbol{y}_k | \boldsymbol{s}_k = \text{Rainy}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Rainy}}, \boldsymbol{\Sigma}_{\text{Rainy}})$$

LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Likelihood of an HMM

$$\boldsymbol{s}_{k+1} = \boldsymbol{f}_{\text{dyn}}(\boldsymbol{s}_k)$$

$$O =$$



$$\boldsymbol{y}_1 \cdots \boldsymbol{y}_T$$

Transition matrix

$$\begin{array}{c} \\ \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{bmatrix} 0.9 & 0.08 & 0.02 \\ 0.25 & 0.7 & 0.05 \\ 0.05 & 0.25 & 0.7 \end{bmatrix} \end{array}$$

Initial probabilities

$$\begin{array}{c} \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{bmatrix} 0.33 & 0.33 & 0.33 \end{bmatrix} \end{array}$$

$$P(O|\lambda) = P(\boldsymbol{y}_1 \cdots \boldsymbol{y}_T|\lambda) = \sum_{\boldsymbol{s}_{1\cdots T} \in \mathbb{D}} P(\boldsymbol{y}_1 \cdots \boldsymbol{y}_T, \boldsymbol{s}_1 \cdots \boldsymbol{s}_T|\lambda)$$

$$\mathbb{D} = \{\text{Sunny}, \text{Cloudy}, \text{Rainy}\}$$

Forward variable

$$\alpha(\boldsymbol{s}_k) = P(\boldsymbol{y}_1 \cdots \boldsymbol{y}_k, \boldsymbol{s}_k)$$

$$P(O|\lambda) = \sum_{\boldsymbol{s}_T} \alpha(\boldsymbol{s}_T)$$

$$\boldsymbol{y}_k = \boldsymbol{f}(\boldsymbol{s}_k)$$

$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Sunny}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Sunny}}, \boldsymbol{\Sigma}_{\text{Sunny}})$$
$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Cloudy}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Cloudy}}, \boldsymbol{\Sigma}_{\text{Cloudy}})$$
$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Rainy}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Rainy}}, \boldsymbol{\Sigma}_{\text{Rainy}})$$

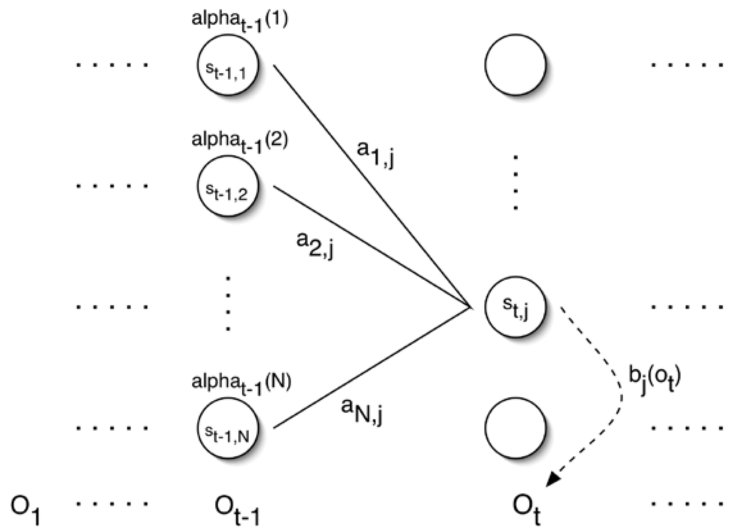
LASA
Learning Algorithms and Systems Laboratory

EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Likelihood of an HMM

$$\boldsymbol{s}_{k+1} = \boldsymbol{f}_{\text{dyn}}(\boldsymbol{s}_k)$$

$$\alpha(\boldsymbol{s}_k) = \sum_{\boldsymbol{s}_{k-1}} \alpha(\boldsymbol{s}_{k-1}) P(\boldsymbol{y}_k|\boldsymbol{s}_k) P(\boldsymbol{s}_k|\boldsymbol{s}_{k-1})$$



Transition matrix

$$\begin{array}{cc} & \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{array}{c} \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} & \begin{bmatrix} 0.9 & 0.08 & 0.02 \\ 0.25 & 0.7 & 0.05 \\ 0.05 & 0.25 & 0.7 \end{bmatrix} \end{array}$$

Initial probabilities

$$\begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array}$$
$$\begin{bmatrix} 0.33 & 0.33 & 0.33 \end{bmatrix}$$

Forward variable

$$\alpha(\boldsymbol{s}_k) = P(\boldsymbol{y}_1 \cdots \boldsymbol{y}_k, \boldsymbol{s}_k)$$

$$P(O|\lambda) = \sum_{\boldsymbol{s}_T} \alpha(\boldsymbol{s}_T)$$

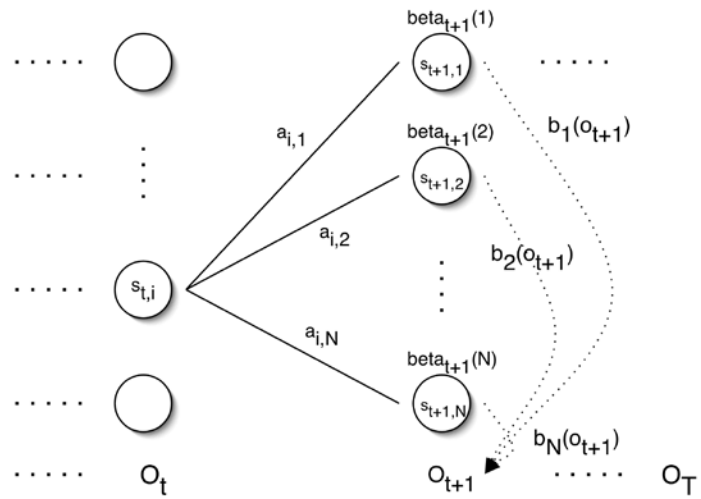
$$\boldsymbol{y}_k = \boldsymbol{f}(\boldsymbol{s}_k)$$

$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Sunny}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Sunny}}, \boldsymbol{\Sigma}_{\text{Sunny}})$$
$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Cloudy}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Cloudy}}, \boldsymbol{\Sigma}_{\text{Cloudy}})$$
$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Rainy}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Rainy}}, \boldsymbol{\Sigma}_{\text{Rainy}})$$

Likelihood of an HMM

$$\boldsymbol{s}_{k+1} = \boldsymbol{f}_{\text{dyn}}(\boldsymbol{s}_k)$$

$$\beta(\boldsymbol{s}_k) = \sum_{\boldsymbol{s}_{k+1}} \beta(\boldsymbol{s}_{k+1}) P(\boldsymbol{y}_{k+1}|\boldsymbol{s}_{k+1}) P(\boldsymbol{s}_{k+1}|\boldsymbol{s}_k)$$



Transition matrix

$$\begin{array}{c|ccc} & \text{Sunny} & \text{Cloudy} & \text{Rainy} \\ \hline \text{Sunny} & 0.9 & 0.08 & 0.02 \\ \text{Cloudy} & 0.25 & 0.7 & 0.05 \\ \text{Rainy} & 0.05 & 0.25 & 0.7 \end{array}$$

Initial probabilities

$$\begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \\ \hline 0.33 & 0.33 & 0.33 \end{array}$$

$$\boldsymbol{y}_k = \boldsymbol{f}(\boldsymbol{s}_k)$$

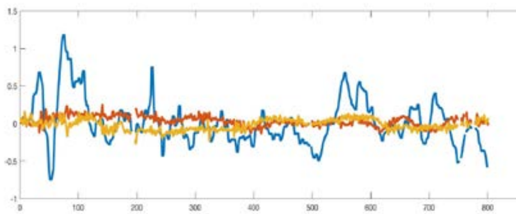
Backward variable

$$\beta(\boldsymbol{s}_k) = P(\boldsymbol{y}_{k+1} \cdots \boldsymbol{y}_T | \boldsymbol{s}_k)$$

$$P(O|\lambda) = \sum_{\boldsymbol{s}_1} \beta(\boldsymbol{s}_1) P(\boldsymbol{s}_1)$$

$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Sunny}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Sunny}}, \boldsymbol{\Sigma}_{\text{Sunny}})$$
$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Cloudy}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Cloudy}}, \boldsymbol{\Sigma}_{\text{Cloudy}})$$
$$P(\boldsymbol{y}_k|\boldsymbol{s}_k = \text{Rainy}) = \mathcal{N}(\boldsymbol{y}_k; \boldsymbol{\mu}_{\text{Rainy}}, \boldsymbol{\Sigma}_{\text{Rainy}})$$

Learning an HMM

$O =$  $\boldsymbol{y}_1 \cdots \boldsymbol{y}_T$

$$\max_{\lambda} \quad \log P(O|\lambda)$$

Baum-Welch algorithm

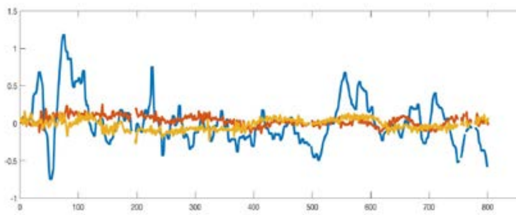
(Expectation-Maximization for HMMs)

- Iterative solution
- Converges to local minimum

Starting from an initial λ find a λ' such that $P(O|\lambda') \geq P(O|\lambda)$

- E-step: Given an observation sequence and a model, find the probabilities of the states to have produced those observations.
- M-step: Given the output of the E-step, update the model parameters to better fit the observations.

Learning an HMM

$O =$  $\boldsymbol{y}_1 \cdots \boldsymbol{y}_T$

E-step:

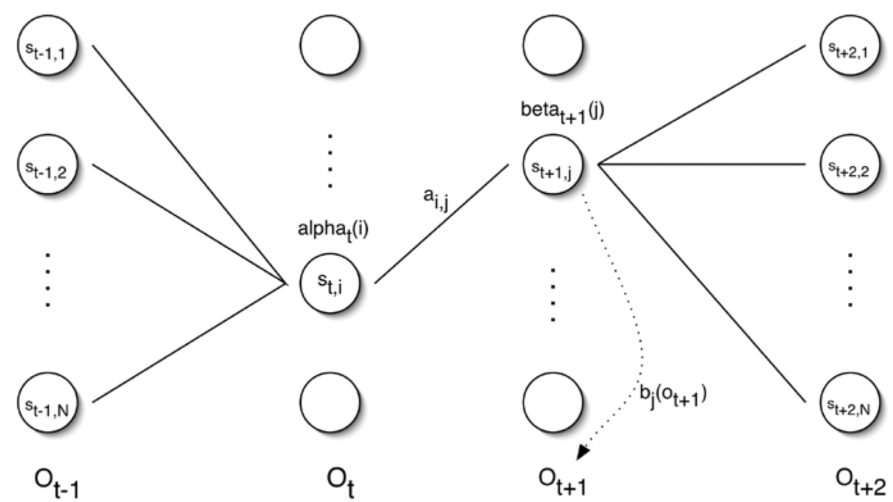
$$\xi_k(i,j) = \frac{\alpha(\boldsymbol{s}_k = i)P(\boldsymbol{s}_k = i, \boldsymbol{s}_{k+1} = j)P(\boldsymbol{y}_k + 1|\boldsymbol{s}_{k+1} = j)\beta(\boldsymbol{s}_{k+1} = j)}{\sum_{\boldsymbol{s}_k}\sum_{\boldsymbol{s}_{k+1}} \alpha(\boldsymbol{s}_k)P(\boldsymbol{s}_k, \boldsymbol{s}_{k+1})P(\boldsymbol{y}_k + 1|\boldsymbol{s}_{k+1})\beta(\boldsymbol{s}_{k+1})}$$

Probability of being in state i at time k and transition to state j

$$\xi_k(i,j) = P(\boldsymbol{s}_k = i, \boldsymbol{s}_k = j | O, \lambda)$$

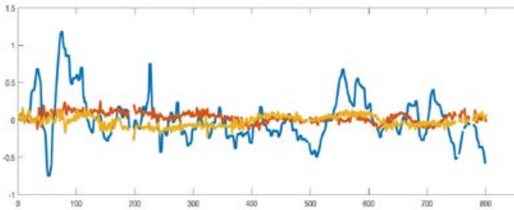
Probability of being in state i at time k

$$\gamma_k(i) = \sum_{j \in \boldsymbol{s}_{k+1}} \xi_k(i,j)$$



LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Learning an HMM

$O =$  $\boldsymbol{y}_1 \cdots \boldsymbol{y}_T$

M-step:

$$\hat{P}(\boldsymbol{s}_{k+1} = \text{Sunny} | \boldsymbol{s}_k = \text{Rainy}) = \frac{\text{number of times we transition Rainy} \to \text{Sunny}}{\text{number of times we observe Sunny}}$$

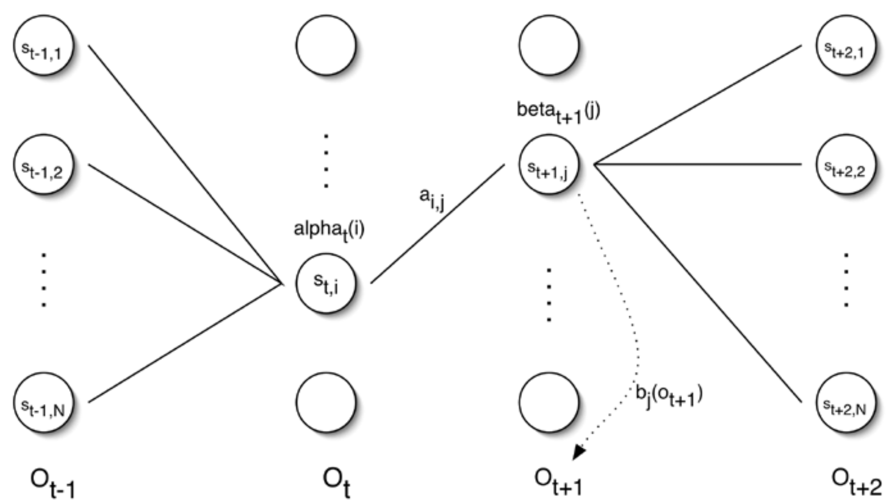
$$P(\boldsymbol{s}_{k+1} = i | \boldsymbol{s}_k = j) = \frac{\sum_{k=1}^{T} \xi_k(i,j)}{\sum_{k=1}^{T} \gamma_k(i)}$$

Probability of being in state i at time k and transition to state j

$$\xi_k(i,j) = P(\boldsymbol{s}_k = i, \boldsymbol{s}_k = j | O, \lambda)$$

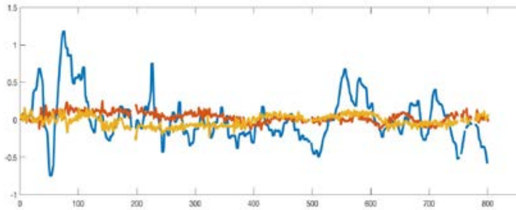
Probability of being in state i at time k

$$\gamma_k(i) = \sum_{j \in \boldsymbol{s}_{k+1}} \xi_k(i,j)$$



LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Learning an HMM

$O =$  $\boldsymbol{y}_1 \cdots \boldsymbol{y}_T$

- HMM is a parametric technique (Fixed number of states, fixed topology)
- → Heuristics to determining the optimal number of states

X : dataset; N : number of datapoints; K: number of free parameters

- Aikaike Information Criterion: $\text{AIC} = -2\ln L + 2K$

- Bayesian Information Criterion: $BIC = -2\ln L + K\ln(N)$

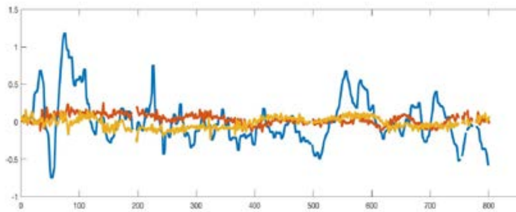
L: maximum likelihood of the model given K parameters

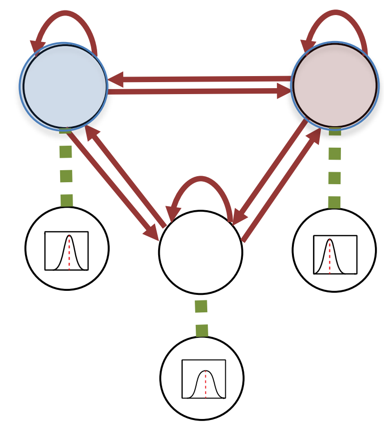
Choosing AIC versus BIC depends on the application:
→Is the purpose of the analysis to make predictions, or to decide which model best represents reality?
AIC may have better predictive ability than BIC, but BIC finds a computationally more efficient solution.

Learning Algorithms and Systems Laboratory

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Applications of HMMs

$O =$ 



State estimation:

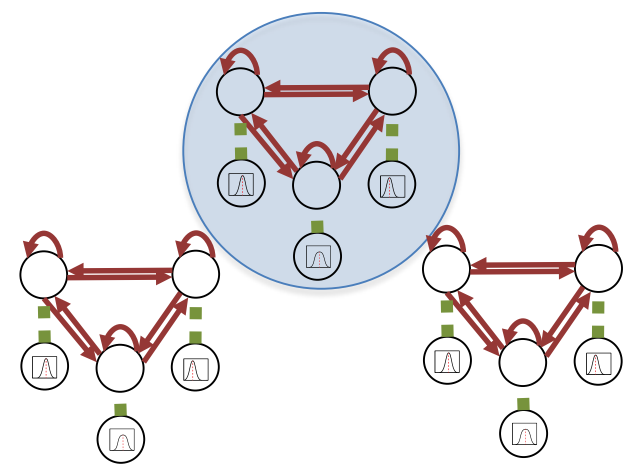
What is the most probable state/state sequence of the system?

Prediction:

What are the most probable next observations/state of the system?

Model selection:

What is the most likely model that represents these observations?

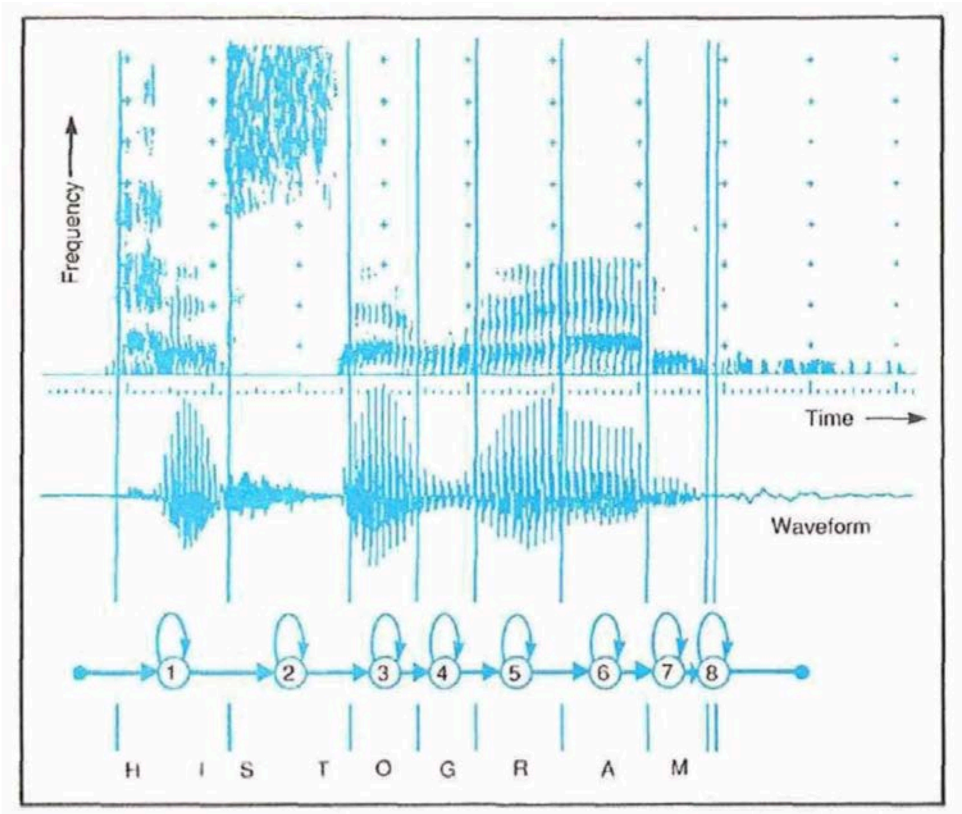


LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Examples

<u>*Speech recognition*</u>:

- Left-to-right model
- States are phonemes
- Observations in frequency domain

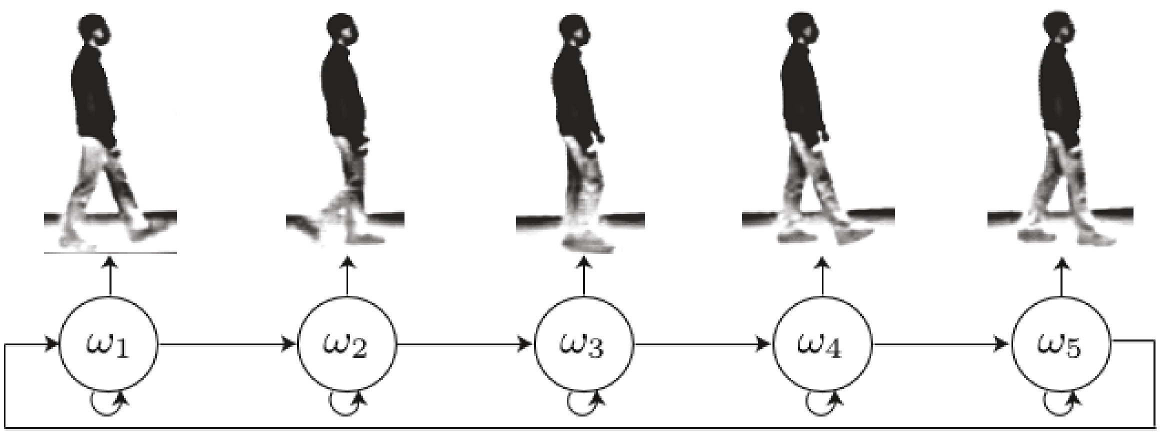


D.B. Paul., Speech Recognition Using Hidden Markov Models, The Lincoln laboratory journal, 1990

Examples

Motion prediction:

- Periodic model
- Observations are observed joints
- Simulate/predict walking patterns



Karg, Michelle, et al. "Human movement analysis: Extension of the f-statistic to time series using hmm." *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*. IEEE, 2013.

Examples

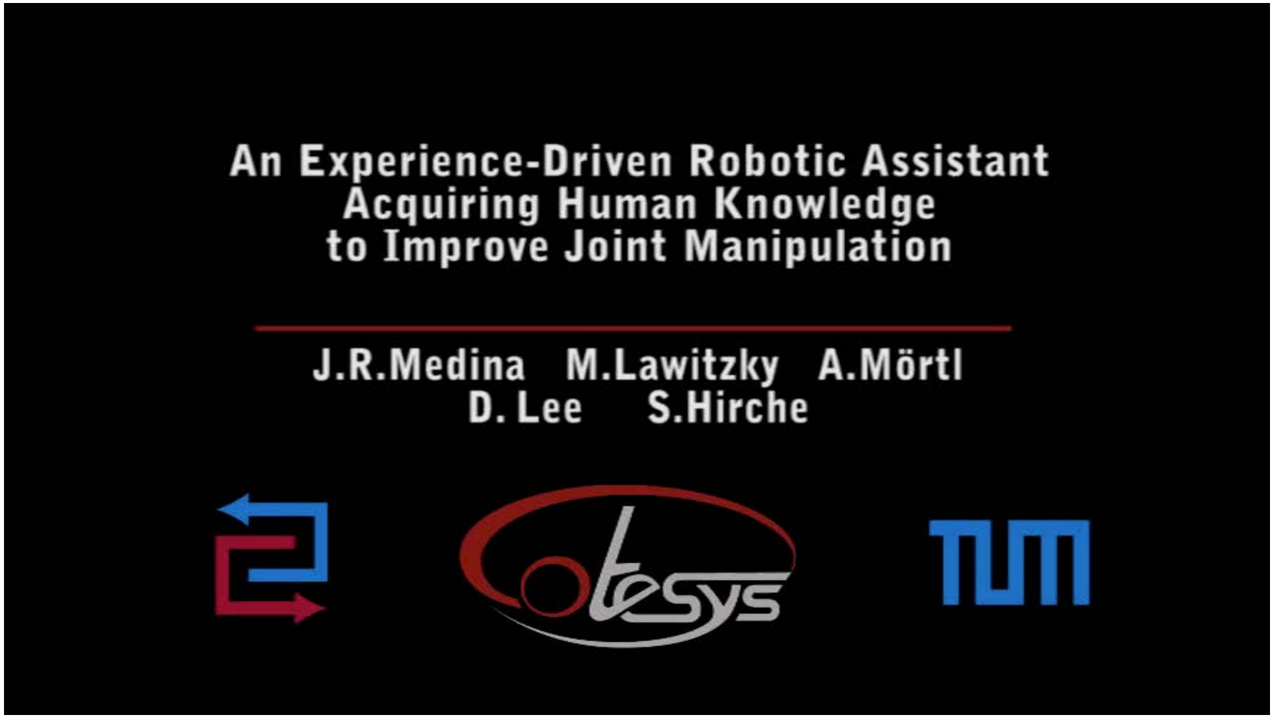
<u>Motion prediction</u>:

- Left-to-right models
- Autonomous segmentation
- Recognition + prediction

Examples

Motion prediction:

- Left-to-right models
- Autonomous segmentation
- Recognition + prediction

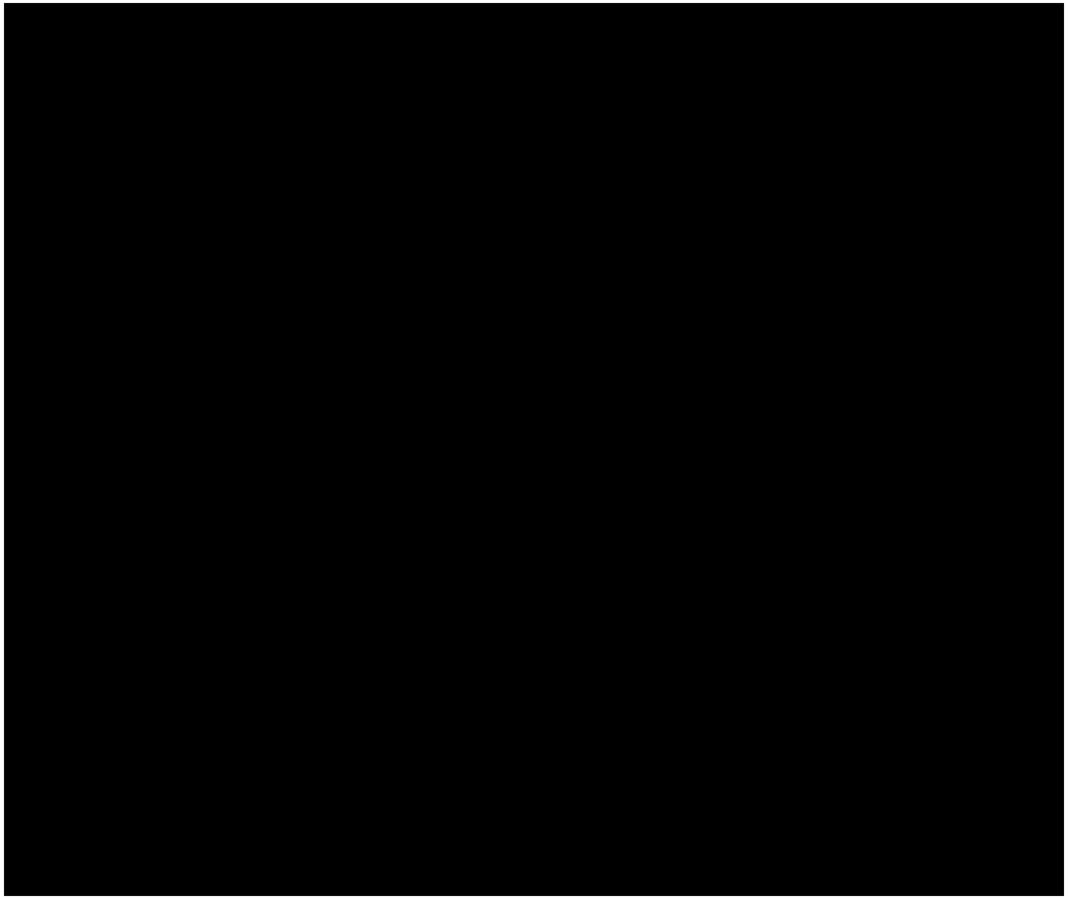


Examples

<u>*Motion prediction*</u>:

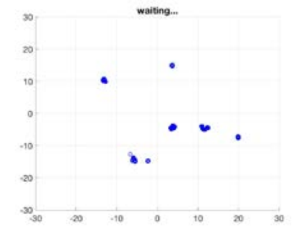
- Left-to-right model

- Each state is a dynamical system



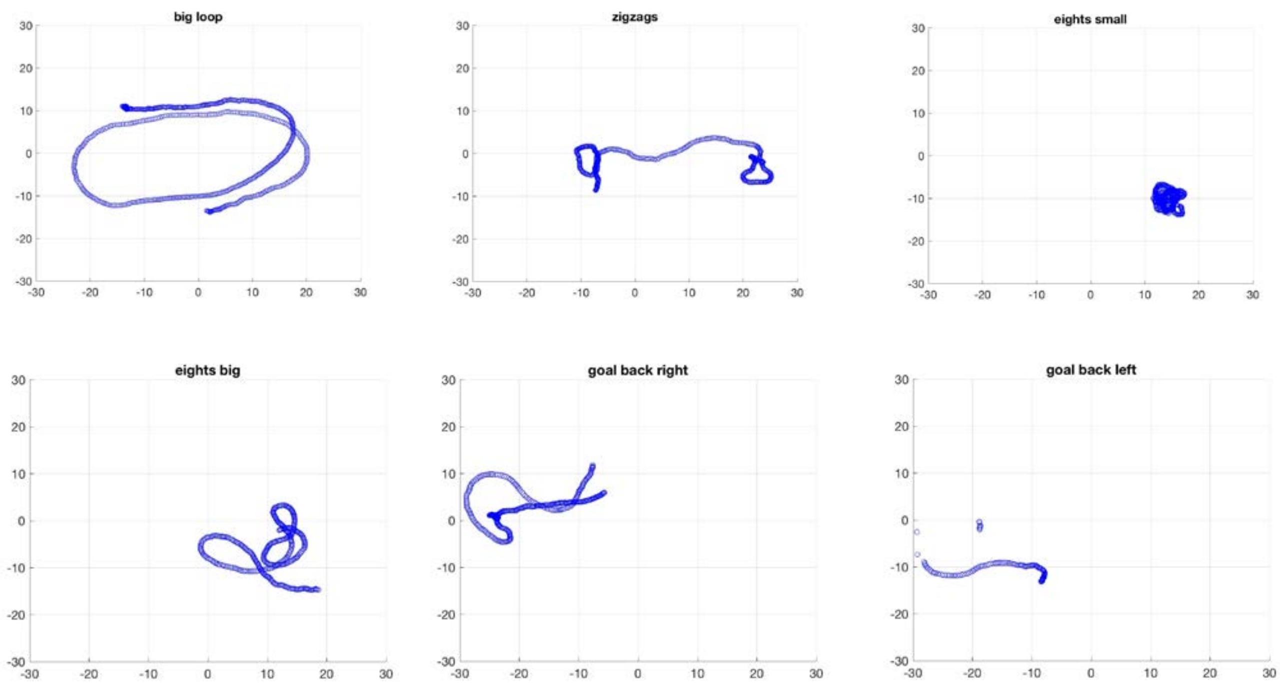
Examples

<u>Toy training set</u>
- 1 player
- 7 actions
- 1 Hidden Markov model per action



<u>Motion recognition</u>:

- Recognition of most likely motion and prediction of next step.

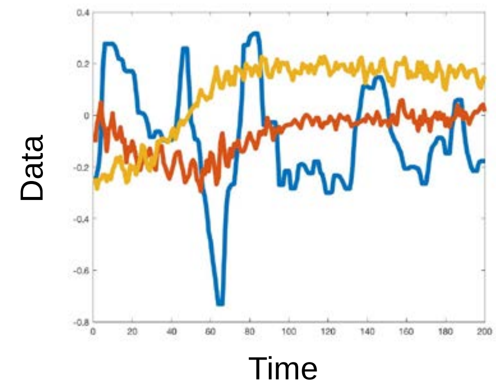


MATLAB demo

Outline

<u>*First part*</u> (10:15 – 11:00):

- Recap on Markov chains
- Hidden Markov Model (HMM)
    - Recognition of time series
    - ML Parameter estimation



<u>*Second part*</u> (11:15 – 12:00):

- Time series segmentation
- Bayesian non-parametrics for HMMs

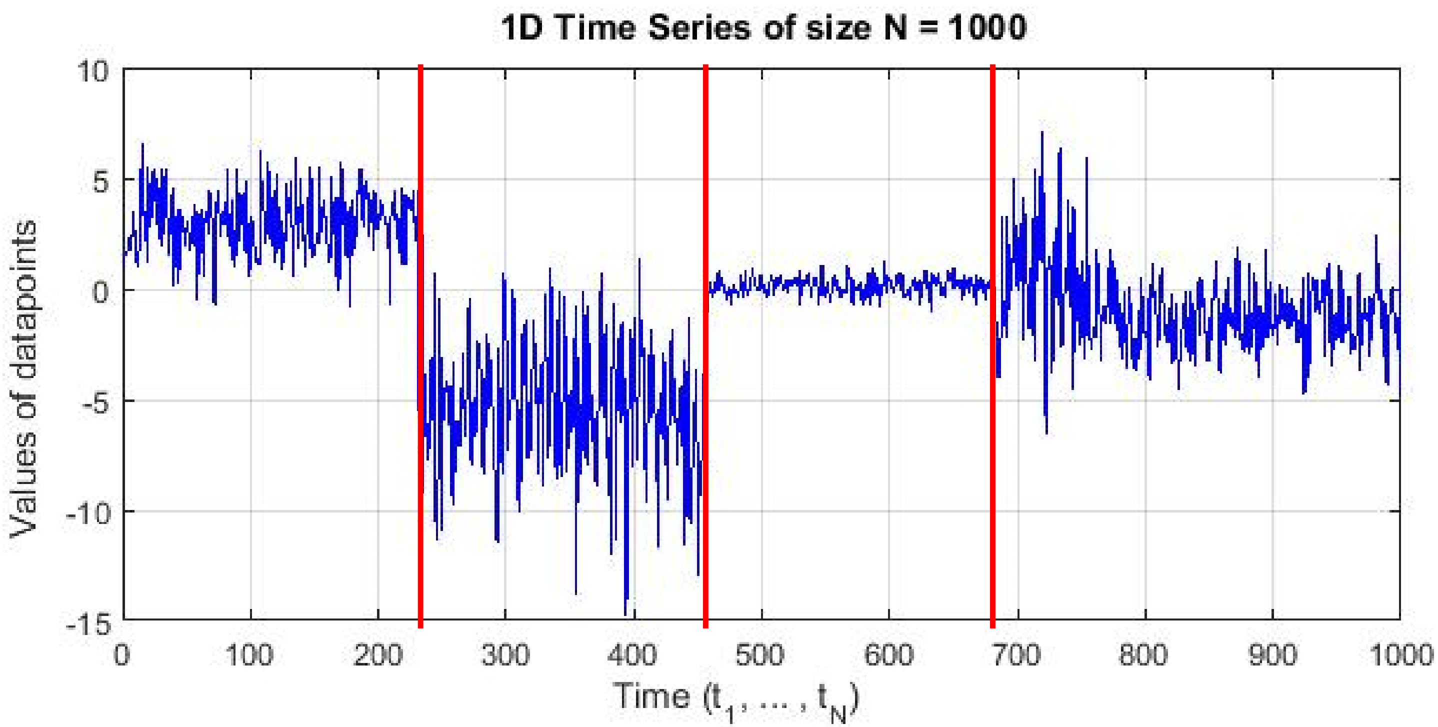
```
https://github.com/epfl-lasa/ML_toolbox
```

LASA
Learning Algorithms and Systems Laboratory

Time series Segmentation

Times-series = Sequence of discrete segments

$$O = $$



$$y_1 \cdots y_T$$

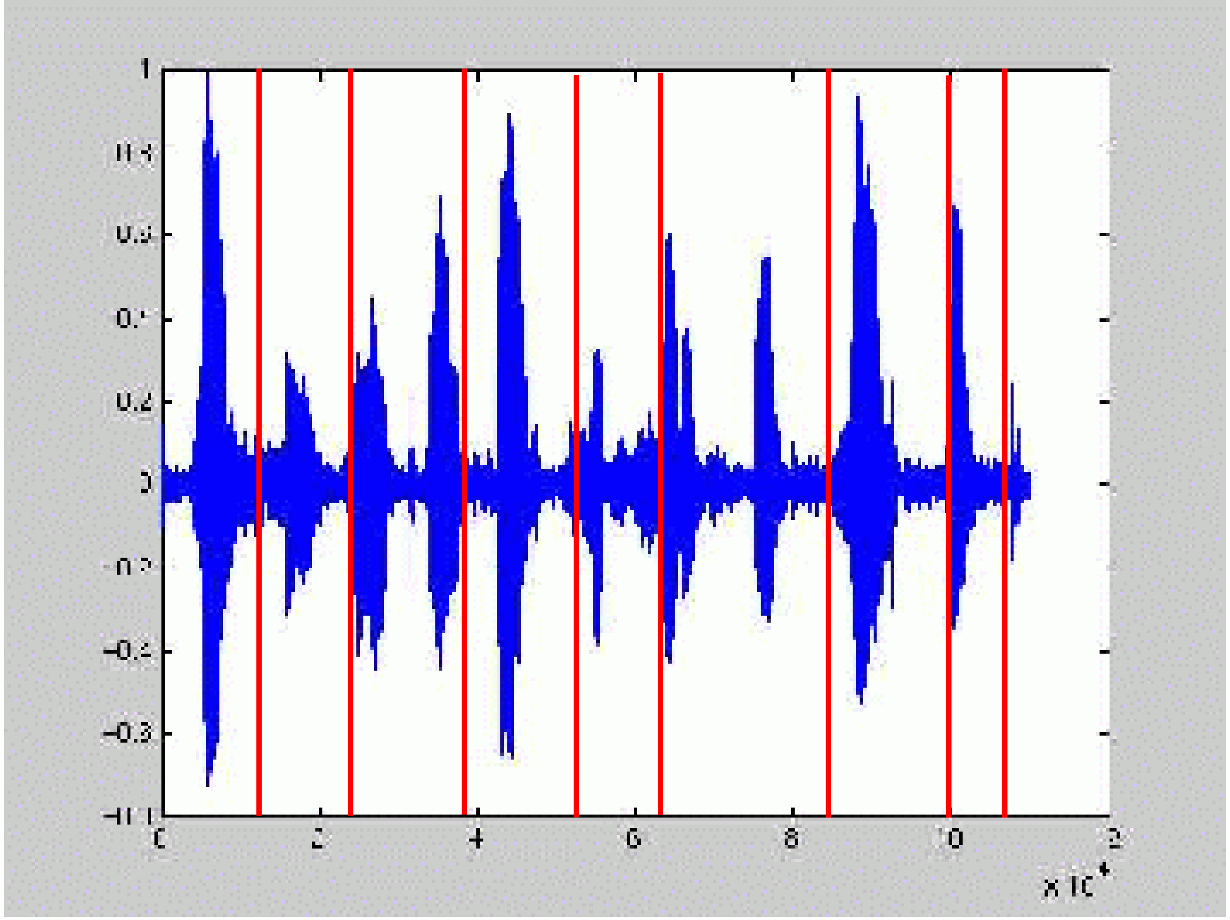
Why is this an important problem?

LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Segmentation of Speech Signals

Segmenting a continuous speech signal into sets of distinct words.

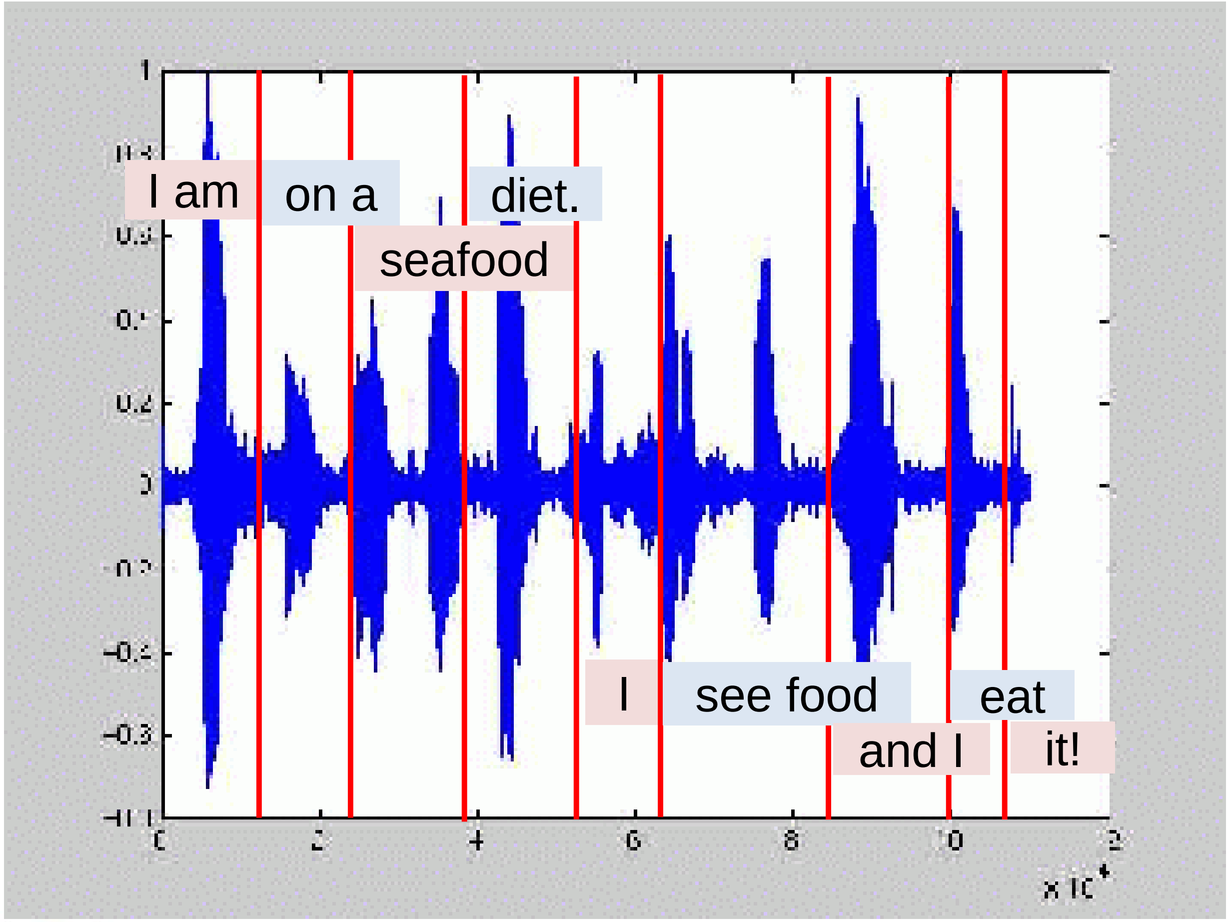
$$O = \qquad y_1 \cdots y_T$$



Segmentation of Speech Signals

Segmenting a continuous speech signal into sets of distinct words.

$$O = \quad y_1 \cdots y_T$$



Segmentation of Human Motion Data

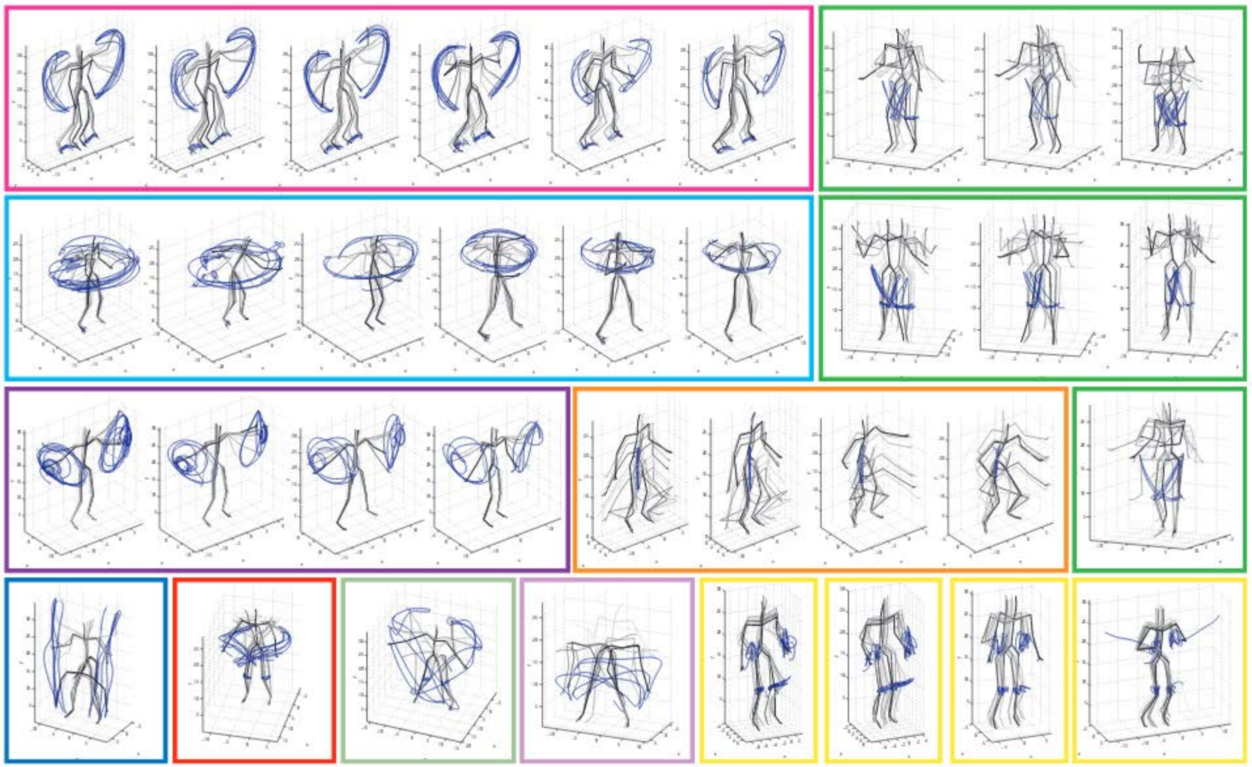
Segmention of Continuous Motion Capture data from exercise routines into motion categories

Jumping Jacks

Knee Raises

Arm Circles

Squats



Emily Fox et al., Sharing Features among Dynamical Systems with Beta Processes, NIPS, 2009

LASA
Learning Algorithms and Systems Laboratory

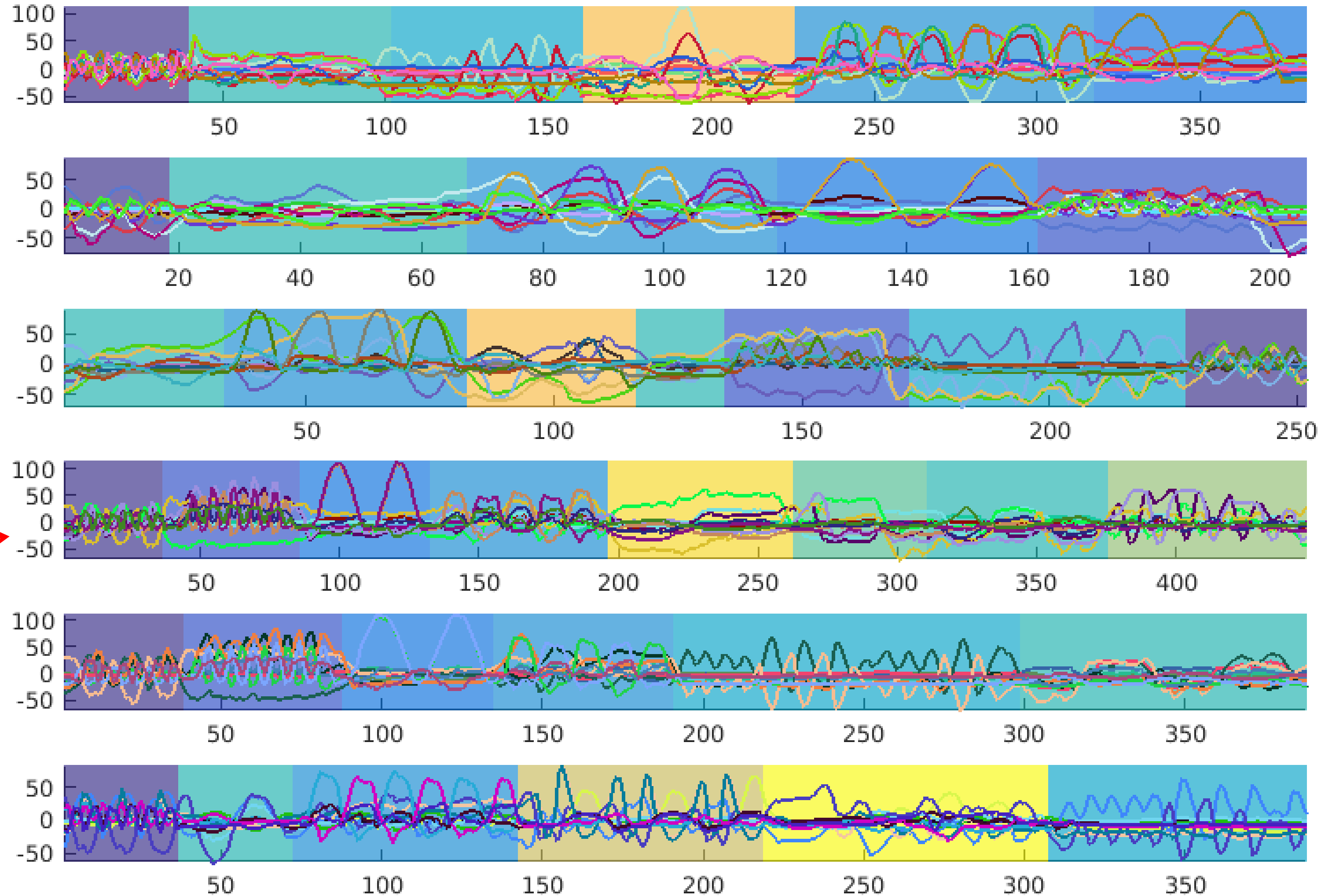
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Segmentation in Human Motion Data

12 Variables

- Torso position
- Waist Angles (2)
- Neck Angle
- Shoulder Angles
- ..

$$\mathbf{y}_i \in \mathbb{R}^{12}$$



Emily Fox et al., Sharing Features among Dynamical Systems with Beta Processes, NIPS, 2009

LASA Learning Algorithms and Systems Laboratory
EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

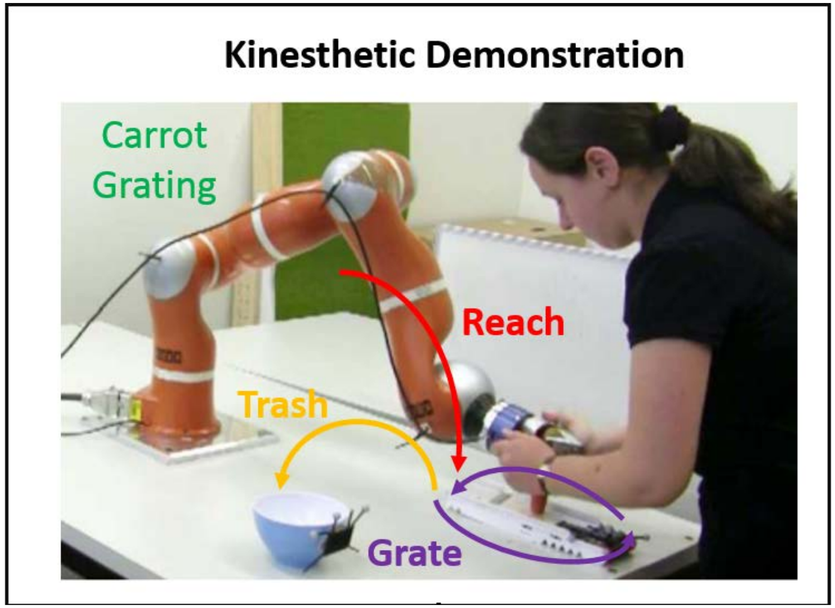
Segmentation in Robotics

Learning Complex Sequential Tasks from Demonstration

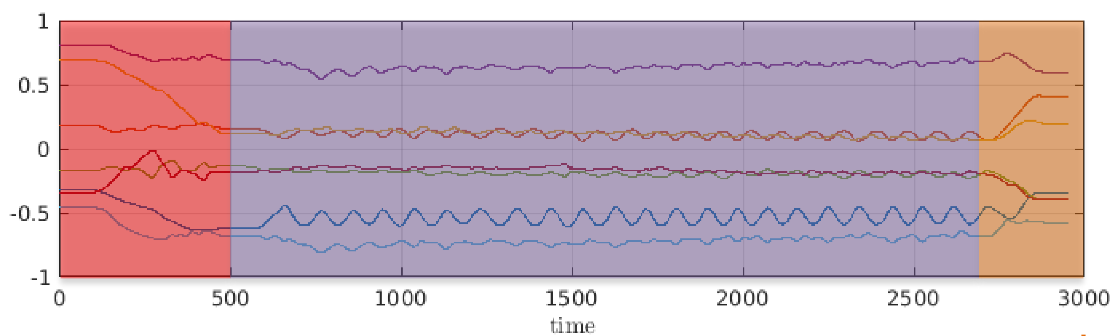
<u>7 Variables</u>

- Position
  x, y, z
- Orientation
  q_i, q_j, q_k, q_w

$\boxed{\mathbf{y}_i \in \mathbb{R}^7}$



$O =$



$\boldsymbol{y}_1 \cdots \boldsymbol{y}_T$

Reach Grate Trash

LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

HMM for Time series Segmentation

Assumptions:

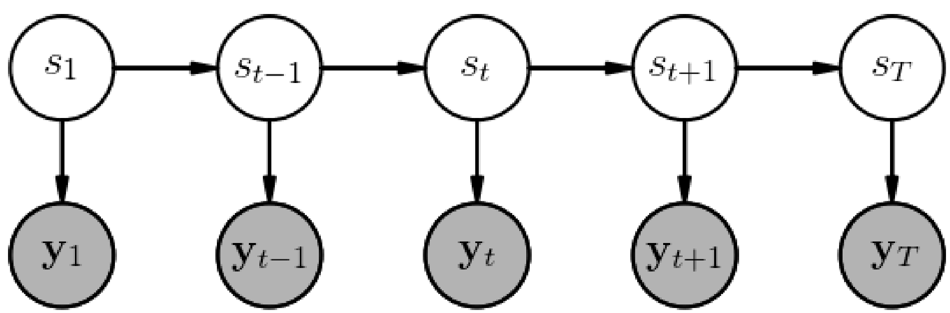
- The time-series has been generated by a system that transitions between a set of hidden states:

$$s \in \{1, \ldots, K\}$$

- At each time step, a sample is drawn from an emission model associated to the current hidden state:

$$\mathbf{y}_t | s_t = k \qquad \sim \mathcal{N}(\theta_{s_t})$$

$$\text{with} \quad \theta_{s_t} \quad = \{\mu_k, \Sigma_k\}$$

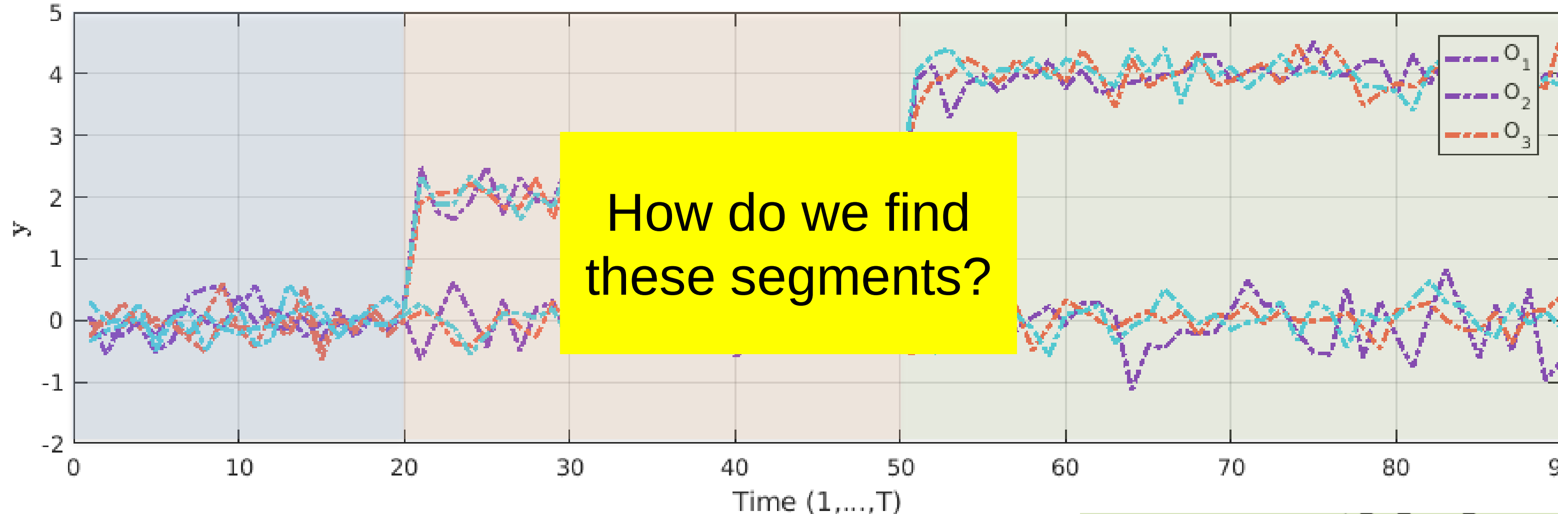


LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

HMM for Time series Segmentation

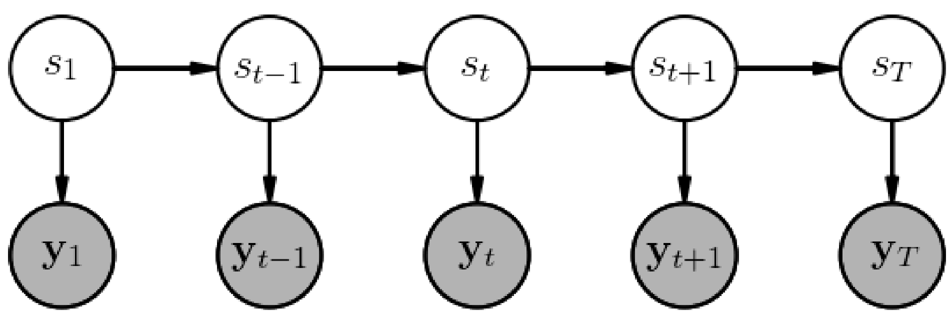
$s_{1:20} = 1$ $\quad s_{21:50} = 2$ $\quad s_{51:90} = 3$

$\mathbf{y} \in \mathbb{R}^2$



How do we find these segments?

$$\mathbf{y}_{1:20} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.5 & 0.2\\0.2 & 0.3\end{bmatrix}\right) \quad \mathbf{y}_{21:50} \sim \mathcal{N}\left(\begin{bmatrix}0\\2\end{bmatrix}, \begin{bmatrix}0.3 & -0.2\\-0.2 & 0.5\end{bmatrix}\right) \quad \mathbf{y}_{51:90} \sim \mathcal{N}\left(\begin{bmatrix}0\\4\end{bmatrix}, \begin{bmatrix}0.5 & 0\\0 & 0.3\end{bmatrix}\right)$$



LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

HMM for Time series Segmentation

Steps for Segmentation with HMM:

1. Learn the HMM parameters through Maximum Likelihood Estimate (MLE):

$$\lambda = \{\pi, A, \Theta\}$$

Initial State Probabilities | Transition Matrix | Emission Model Parameters

HMM Likelihood

$$P(O|\lambda) = P(\boldsymbol{y}_1 \cdots \boldsymbol{y}_T|\lambda) = \sum_{\boldsymbol{s}_{1\ldots T} \in \mathbb{D}} P(\boldsymbol{y}_1 \cdots \boldsymbol{y}_T, \boldsymbol{s}_1 \cdots \boldsymbol{s}_T|\lambda)$$

$$\max_{\lambda} \quad \log P(O|\lambda)$$

Baum-Welch algorithm

(Expectation-Maximization for HMMs)

- Iterative solution
- Converges to local minimum

Hyper-parameter:
Number of states possible K

HMM for Time series Segmentation

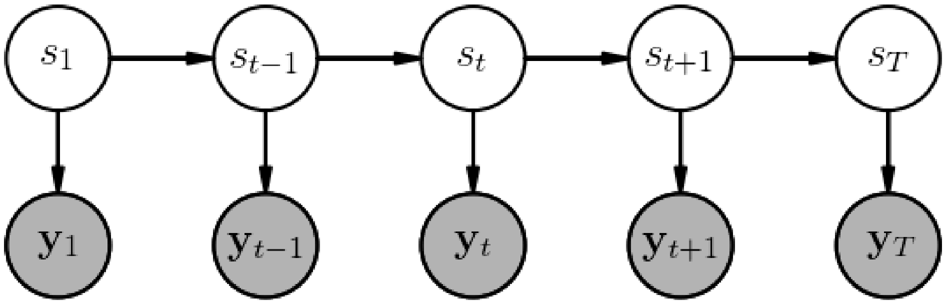
<u>*Steps for Segmentation with HMM*</u>:

2. Find the most probable sequence of states generating the observations through the **Viterbi algorithm**:

$$S^* = \arg\max_{S} \quad p(Y, S|\Theta)$$

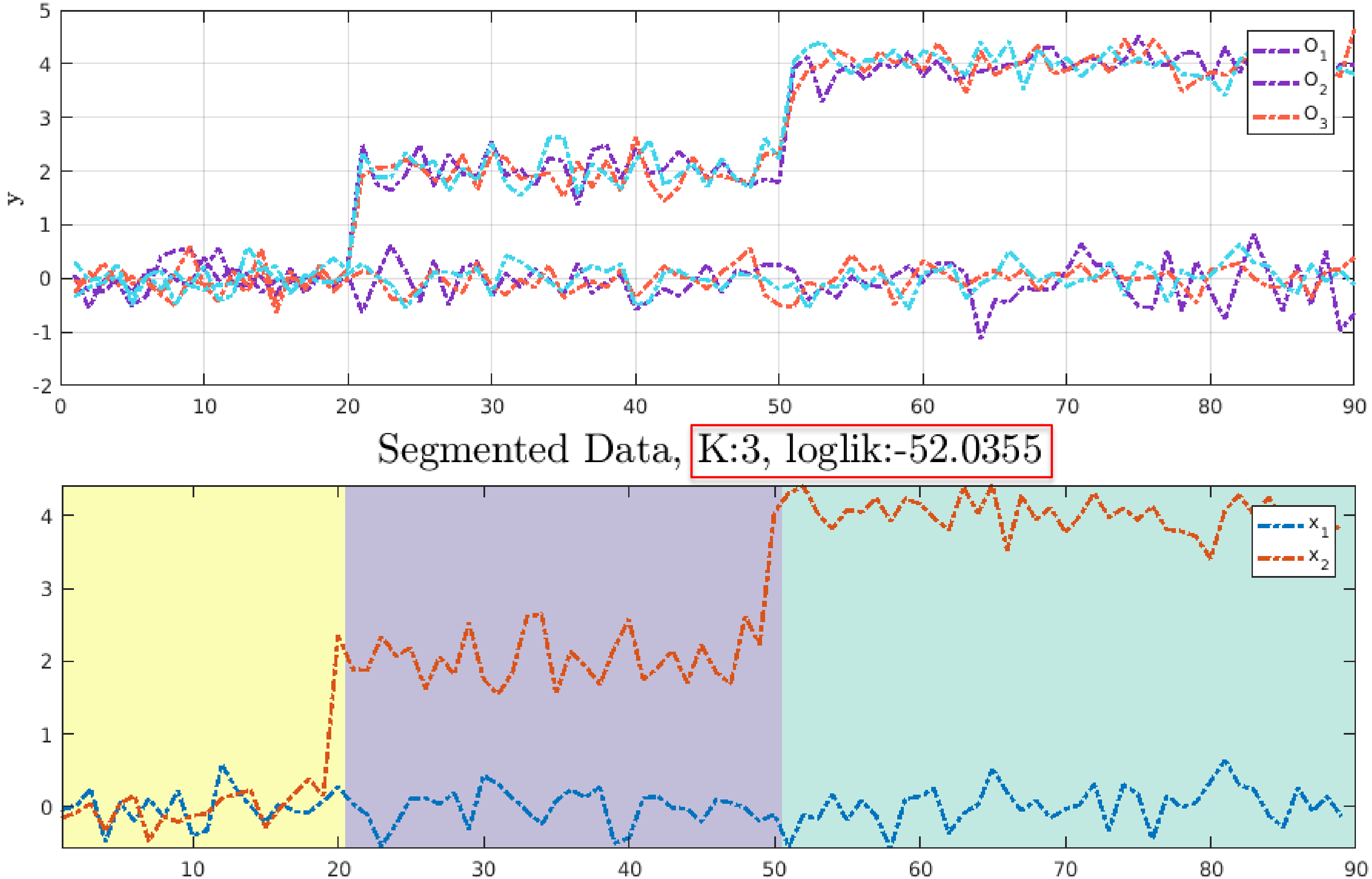
HMM Joint Probability Distribution

$$p(Y, S|\Theta) = p(s_1|\pi) \left[\prod_{t=2}^{T} p(s_t|s_{t-1}, A) \right] \prod_{t=1}^{T} p(\mathbf{y}_t|s_t, \Theta)$$



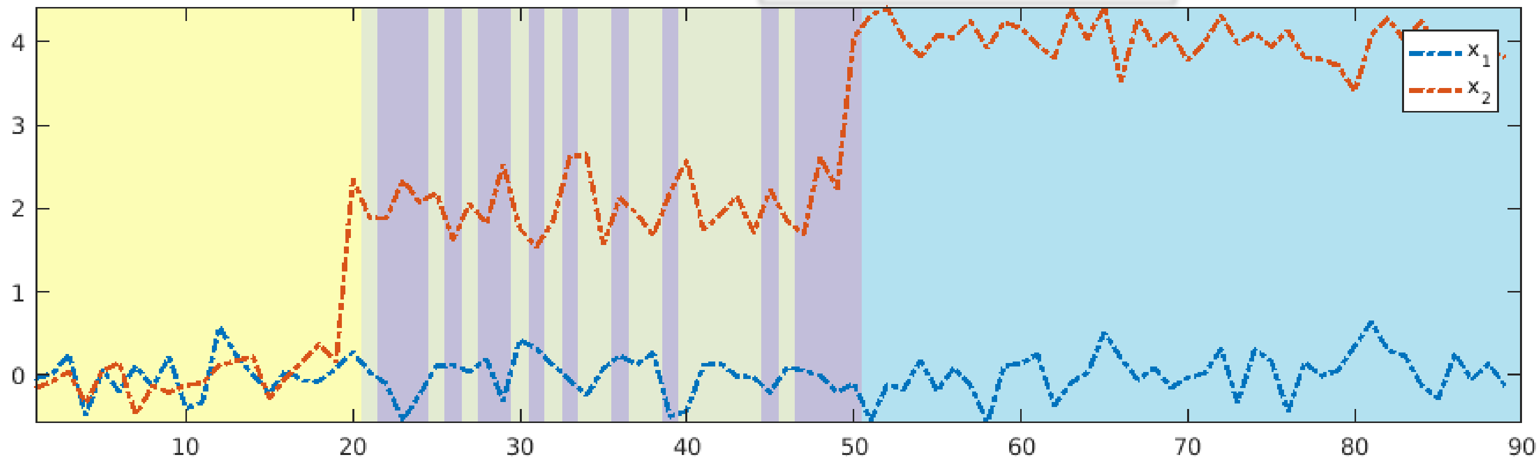
HMM for Time series Segmentation

$\mathbf{y} \in \mathbb{R}^2$

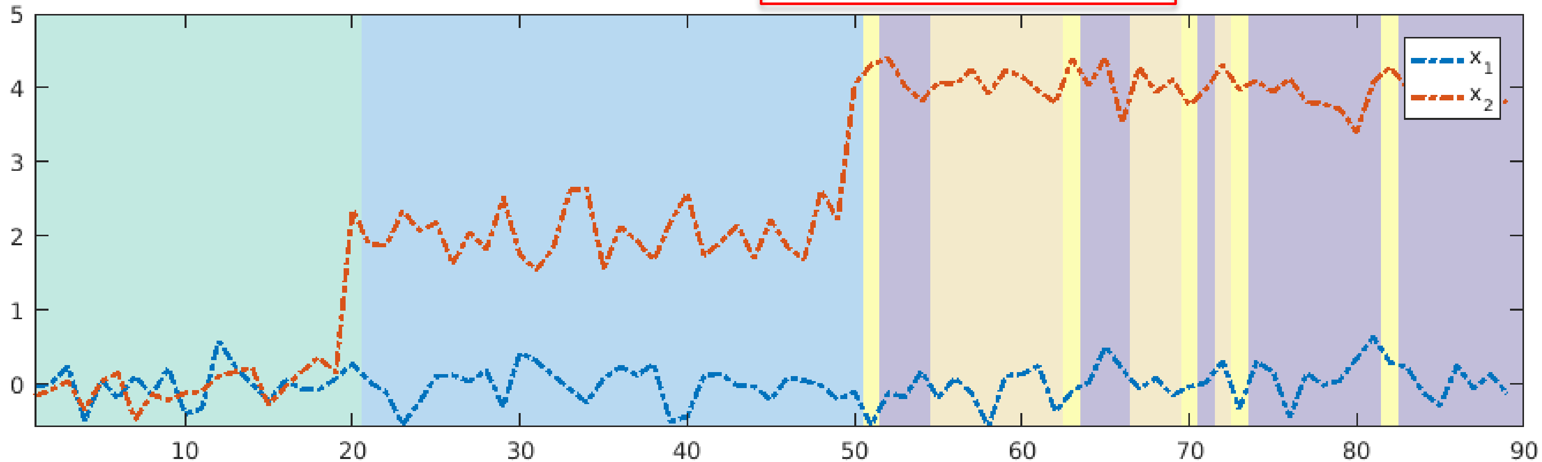


HMM for Time series Segmentation

Segmented Data, K:4, loglik:-36.7299

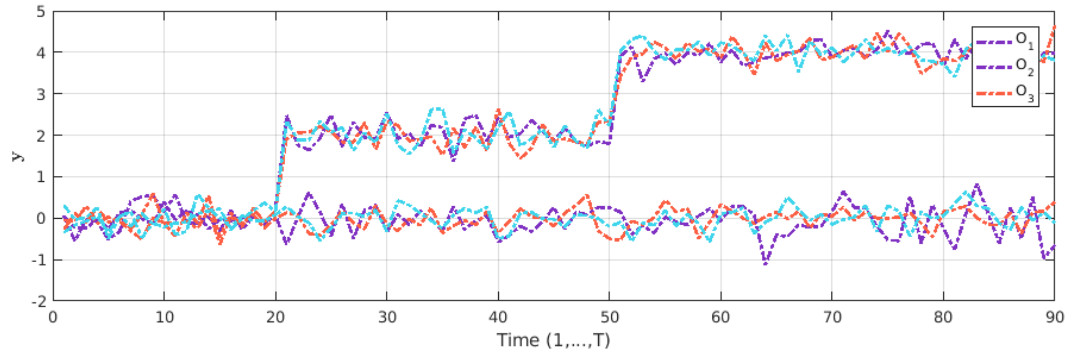


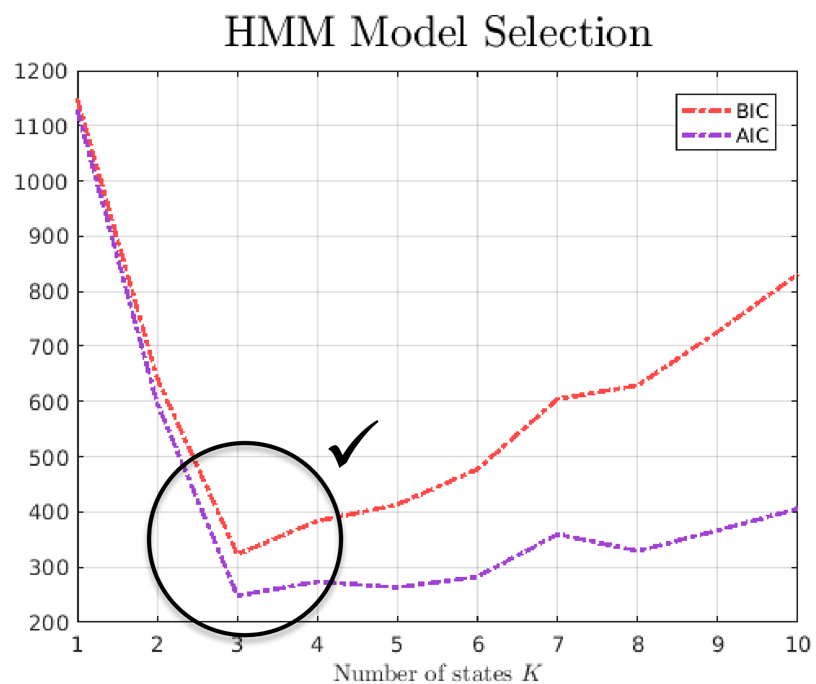
Segmented Data, K:5, loglik:-49.356



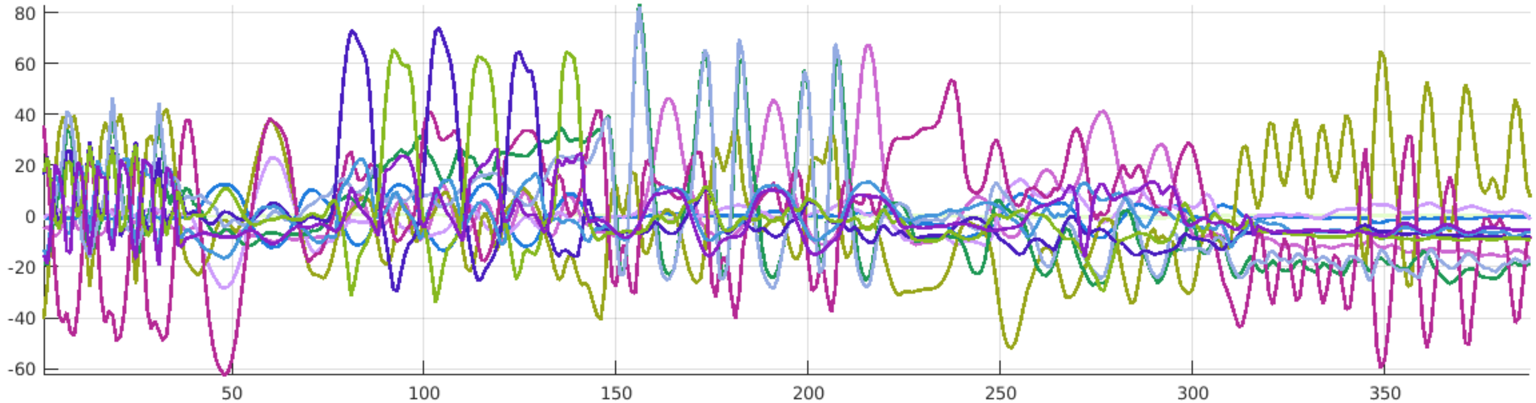
LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model Selection for HMMs

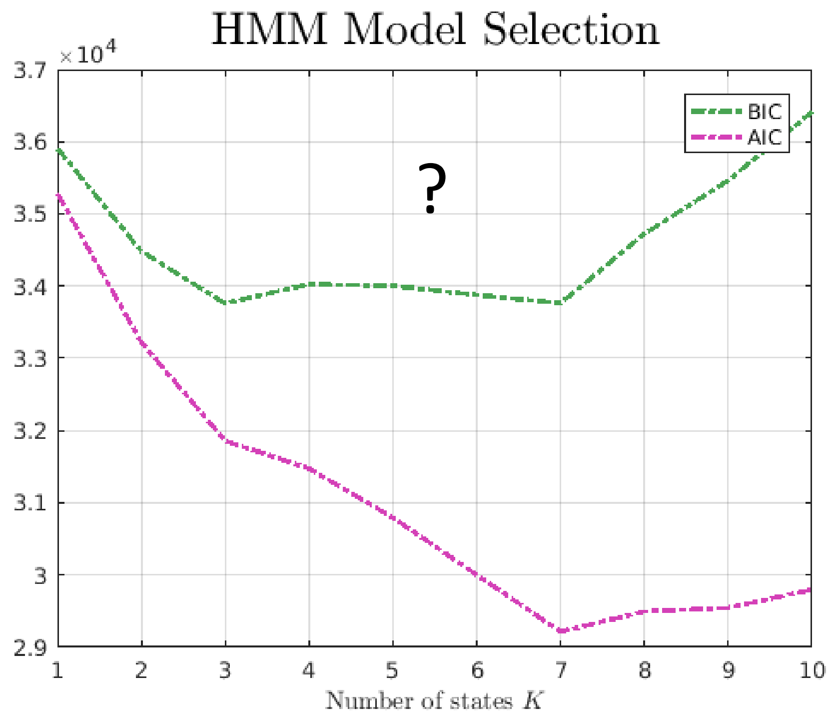




Model Selection for HMMs



$\mathbf{y}_i \in \mathbb{R}^{12}$



EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Limitations of classical finite HMMs for Segmentation

Undefined # hidden states

Cardinality: Choice of hidden states is based on Model Selection heuristics, there is little understanding of the strengths and weaknesses of such methods in this setting [1].

Fixed Transition Matrix

Topology: We assume that all time series share the same set of emission models and switch among them in exactly the same manner [2].

[1] Emily Fox et al., An HDP-HMM for Systems with State Persistence, ICML, 2008
[2] Emily Fox et al., Sharing Features among Dynamical Systems with Beta Processes, NIPS, 2009

Solution: Bayesian Non-Parametrics

Bayesian Non-Parametrics

- ***Bayesian:*** Use *Bayesian inference* to estimate the parameters; i.e. priors on model parameters!

$$\lambda = \{\pi, \overset{\text{Prior}}{\downarrow \atop A}, \overset{\text{Prior}}{\swarrow \atop \Theta}\}$$

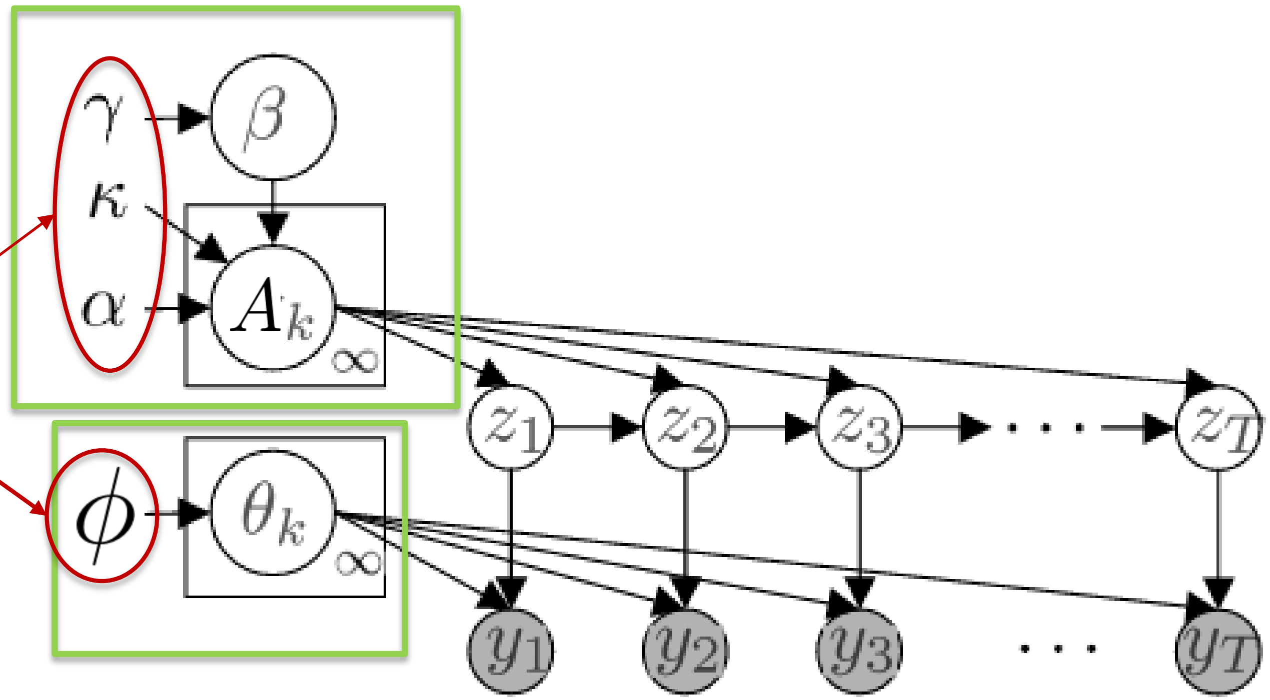
- ***Non-parametric*:** Does NOT mean methods with “no parameters”, rather models whose complexity (# of states, # Gaussians) is inferred from the data.
  1. Number of parameters grows with sample size.
  2. Infinite-dimensional parameter space!

BNP for HMMs: HDP-HMM

☺ Cardinality

- ***Hierarchical Dirichlet Process (HDP)*** prior on the transition Matrix!

Hyper-
parameters!



- ***Normal Inverse Wishart (NIW)*** prior on emission parameters!

Emily Fox et al., An HDP-HMM for Systems with State Persistence, ICML, 2008

LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

BNP for HMMs: HDP-HMM

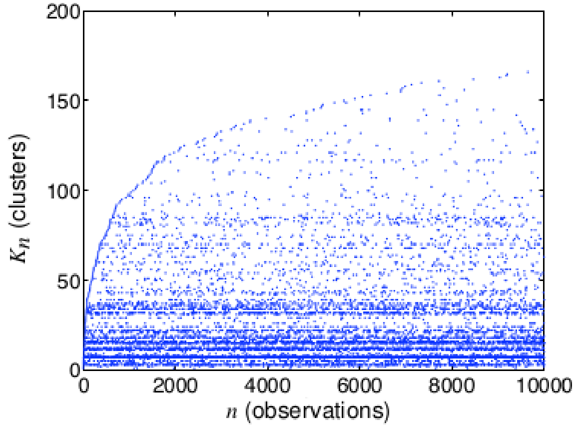
☺ Cardinality

- The **Dirichlet Process (DP)** is a prior distribution over distributions.

$$M(\,.\,) = \sum_{k=1}^{\infty} C_k \delta_{\Theta_k}(\,.\,) \qquad \text{where} \qquad \sum_{k=1}^{\infty} C_k = 1$$

- Used for clustering with infinite mixture models; i.e. instead of setting K in a GMM, the K is learned from data.

$$K_n = \#\text{ clusters in sample of size } n$$



This only gives us an estimate of the K clusters!
Cannot be use directly on transition matrix:

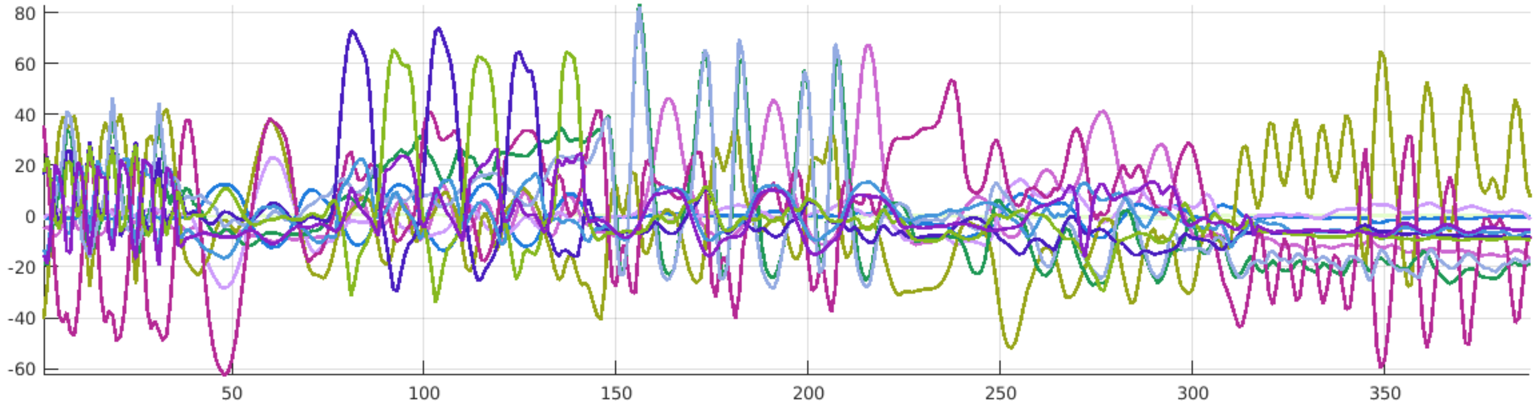
$$A \in \mathbb{R}^{K \times K}$$

- *The **Hierarchical Dirichlet Process (HDP)*** is a hierarchy of DPs!

LASA
Learning Algorithms and Systems Laboratory

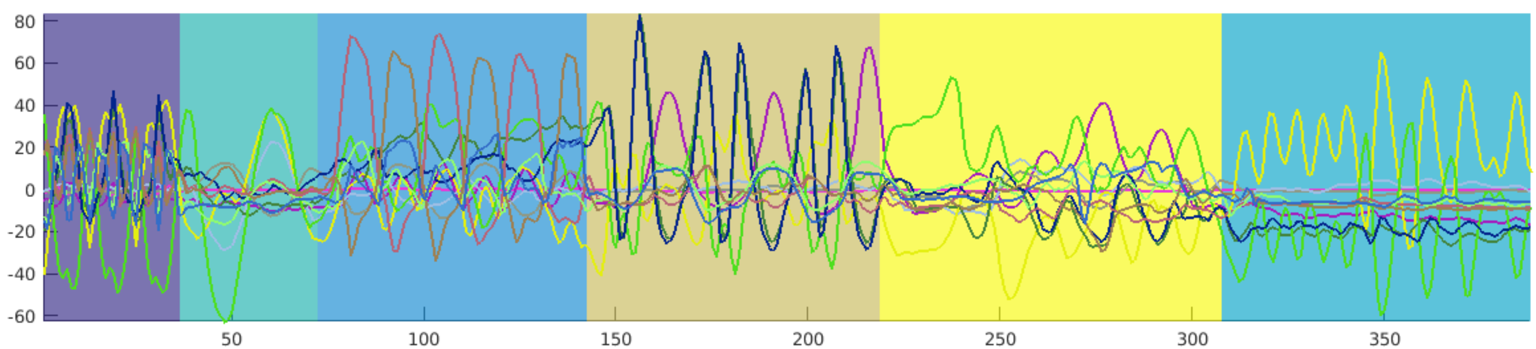
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Segmentation with HDP-HMM



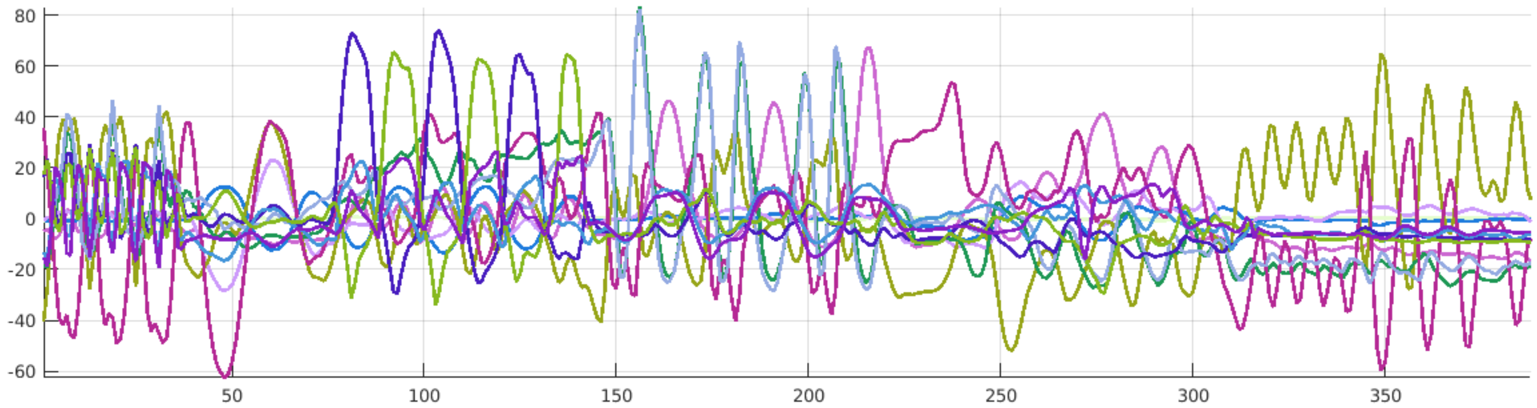
$$\mathbf{y}_i \in \mathbb{R}^{12}$$

✓

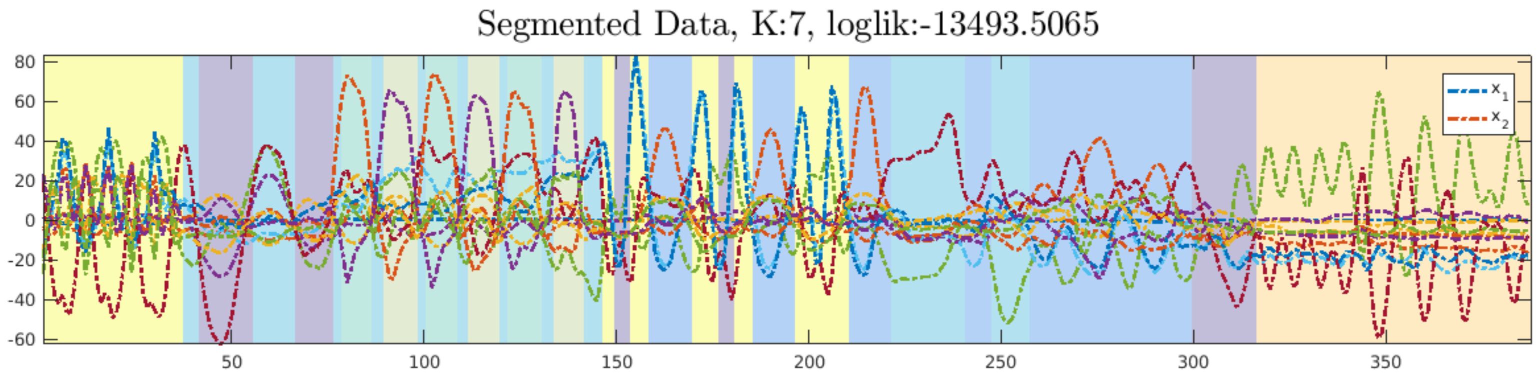


Emily Fox et al., An HDP-HMM for Systems with State Persistence, ICML, 2008

Compare to Model Selection with Classical HMMs



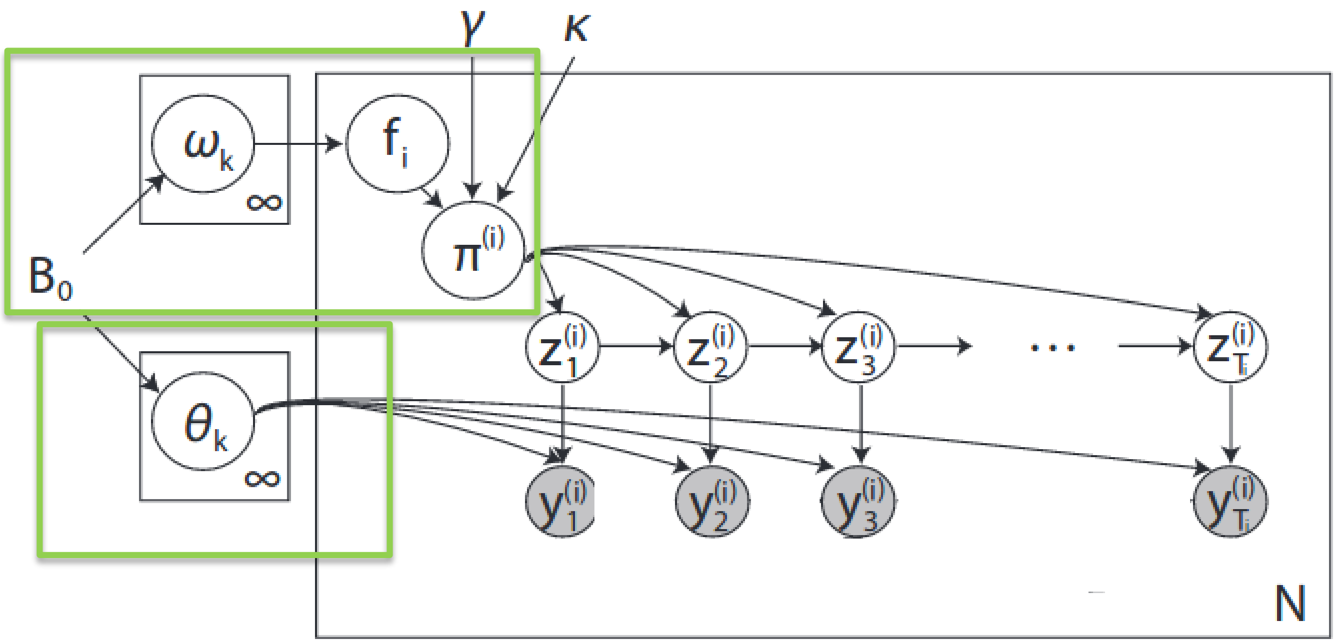
$$\mathbf{y}_i \in \mathbb{R}^{12}$$



BNP for HMMs: BP-HMM

☺ Cardinality
☺ Topology

- *The **Beta Process (BP)*** prior on the transition Matrix!



- ***Normal Inverse Wishart (NIW)*** prior on emission parameters!

Emily Fox et al., Sharing Features among Dynamical Systems with Beta Processes, NIPS, 2009

LASA
Learning Algorithms and Systems Laboratory
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

BNP for HMMs: BP-HMM

☺ Cardinality

☺ Topology

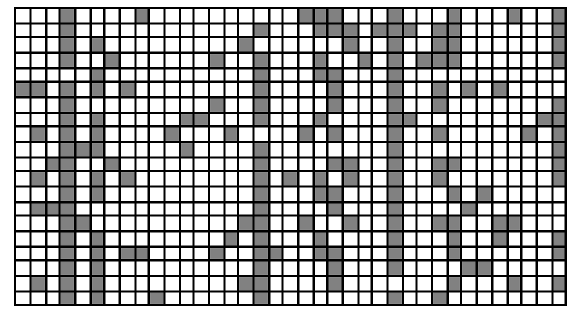
- *The **Beta Process (BP)*** prior on the transition Matrix!

Beta Process (BP)

Distribution on objects of the form

$$\theta = \sum_{k=1}^{\infty} w_k \delta_{\phi_k} \qquad \text{with } w_k \in [0, 1] \ .$$

Time-Series



Features (i.e. shared HMM States)

LASA
Learning Algorithms and Systems Laboratory

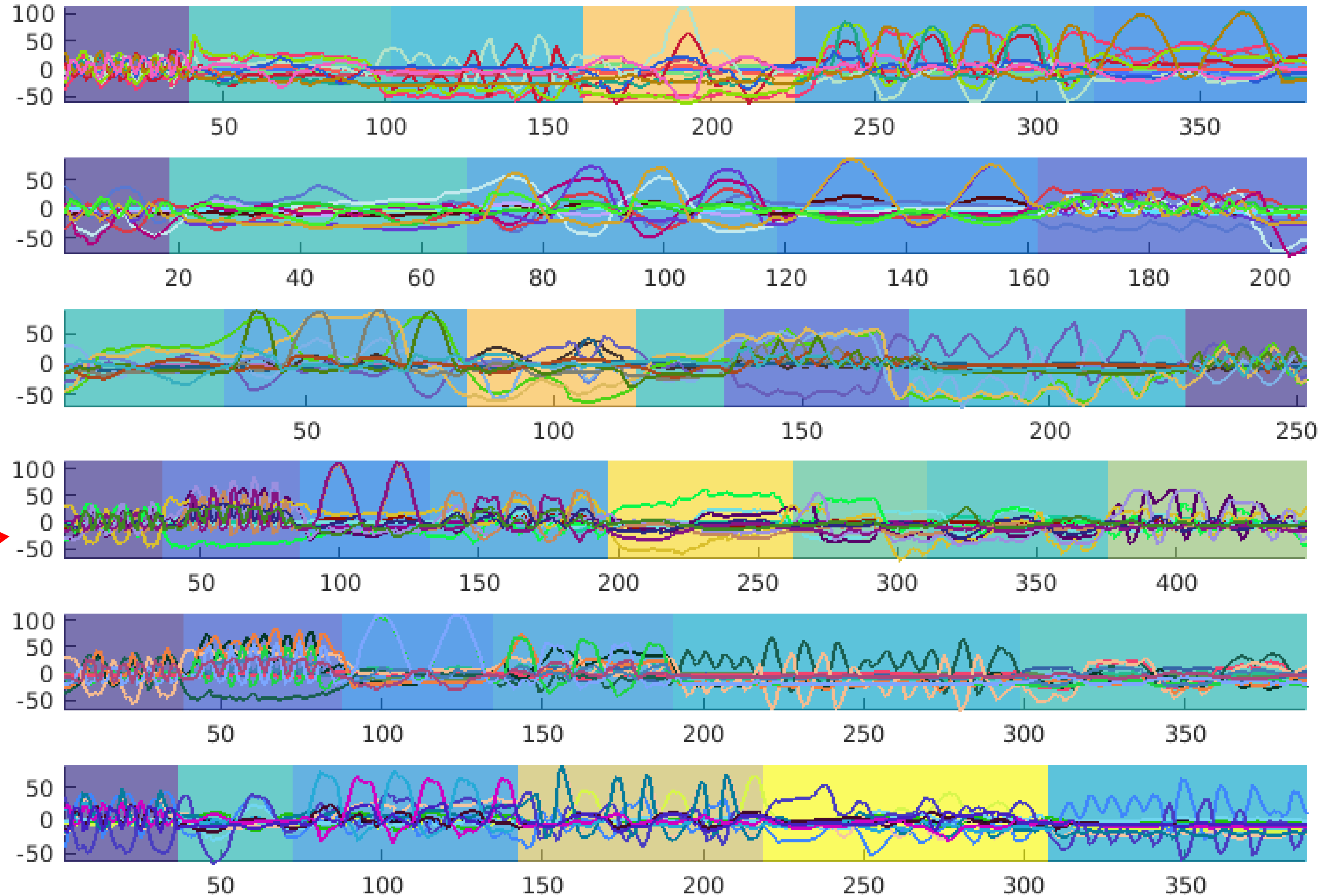
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Segmentation in Human Motion Data

<u>12 Variables</u>

- Torso position
- Waist Angles (2)
- Neck Angle
- Shoulder Angles
- ..

$$\mathbf{y}_i \in \mathbb{R}^{12}$$



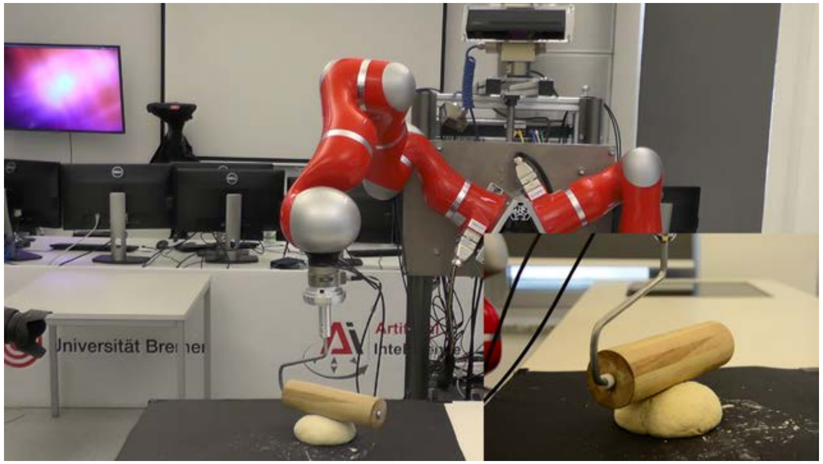
Emily Fox et al., Sharing Features among Dynamical Systems with Beta Processes, NIPS, 2009

LASA Learning Algorithms and Systems Laboratory

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Applications in Robotics

Learning Complex Sequential Tasks from Demonstration



LASA
Learning Algorithms and Systems Laboratory

EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Learning Complex Sequential Tasks from Demonstrations: Pizza Dough Rolling

Nadia Figueroa, Lucia Pais and Aude Billard

LEARNING ALGORITHMS AND SYSTEMS LABORATORY

LASA

EPFL

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

robohow

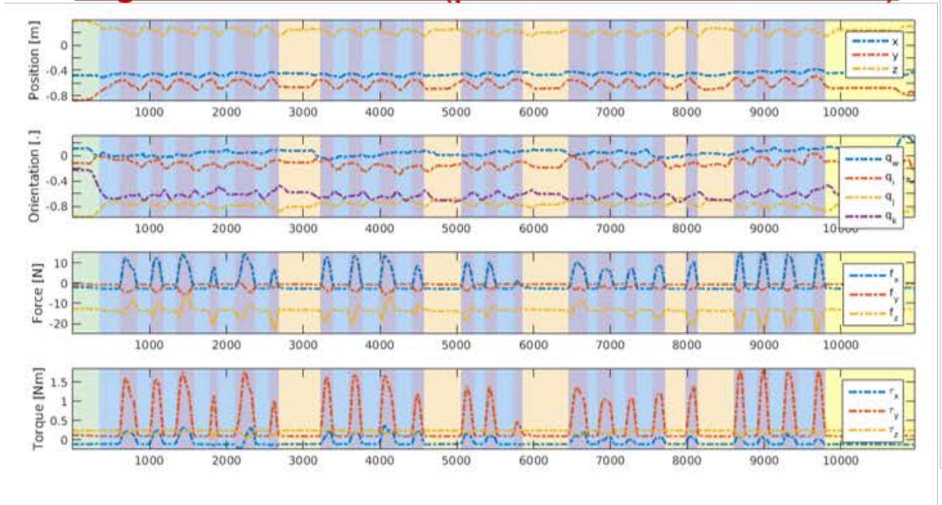
LASA

Learning Algorithms and Systems Laboratory

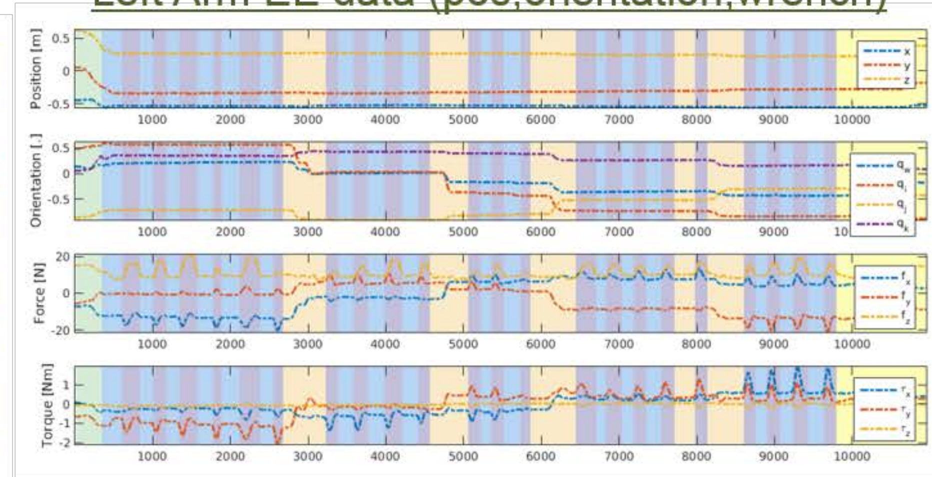
EPFL

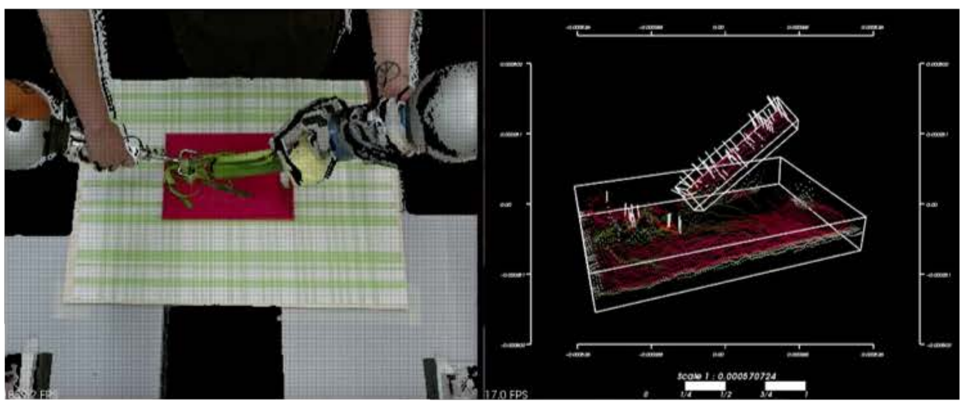
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

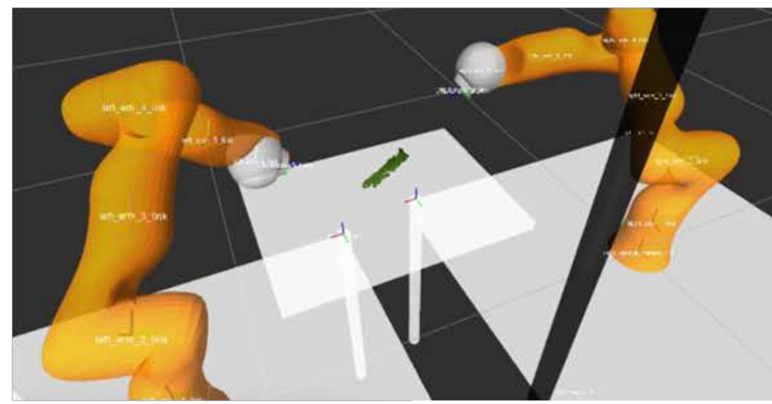
Right Arm EE data (pos,orientation,wrench)



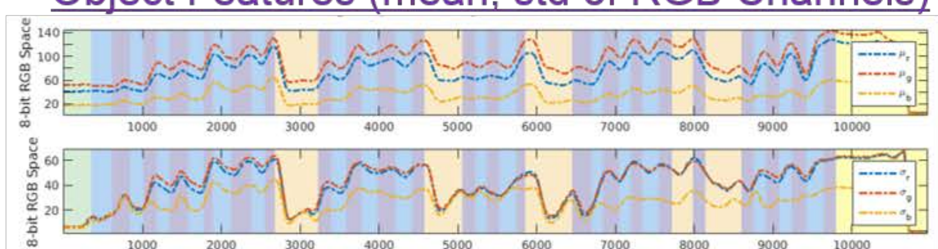
Left Arm EE data (pos,orientation,wrench)





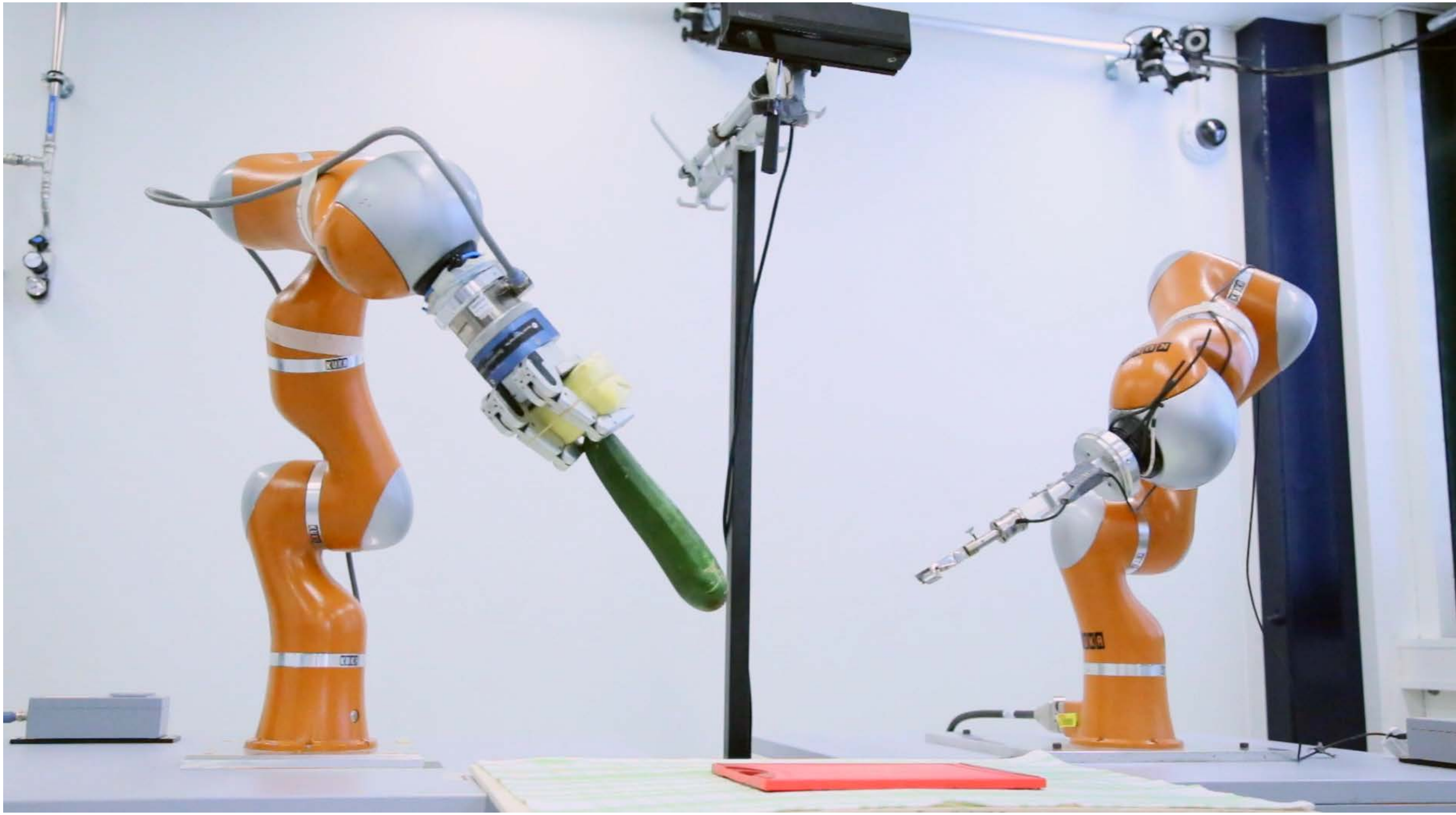


Object Features (mean, std of RGB Channels)



LASA
Learning Algorithms and Systems Laboratory

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE



LASA
Learning Algorithms and
Systems Laboratory

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE