Abstraction by Structure

Carl Henrik Ek, Danica Kragic
{chek, danik}@csc.kth.se

Royal Institute of Technology

April 18, 2012

Centre for Autonomous Systems

- Head of group, Prof. Danica Kragic, 8 Senior staff
- 12+ Post-docs/Researchers
- 20+ PhD students
- To celebrate our 20th ICRA is coming to town in 2016

Centre for Autonomous Systems

- Head of group, Prof. Danica Kragic, 8 Senior staff
- 12+ Post-docs/Researchers
- 20+ PhD students
- To celebrate our 20th ICRA is coming to town in 2016

Collaborators

Hedvig Kjellström, Niklas Bergström, Marianna Madry, Florian Pokorny, Renaud Detry, Dan Song, Javier Romero, Martin Hjelm, Andrea Baisero, Guoliang Luo, Andreas Damianou, Neil Lawrence, Neill Campbell, Colin Dalton, Alexander Davies

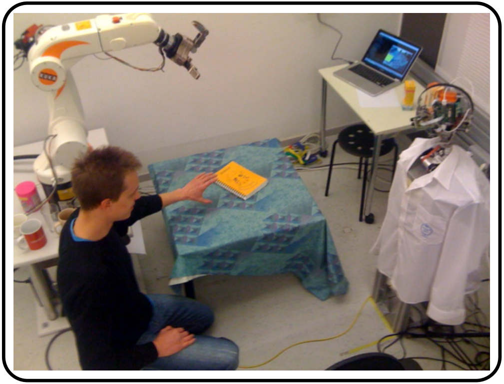
Introduction

Structural Representations

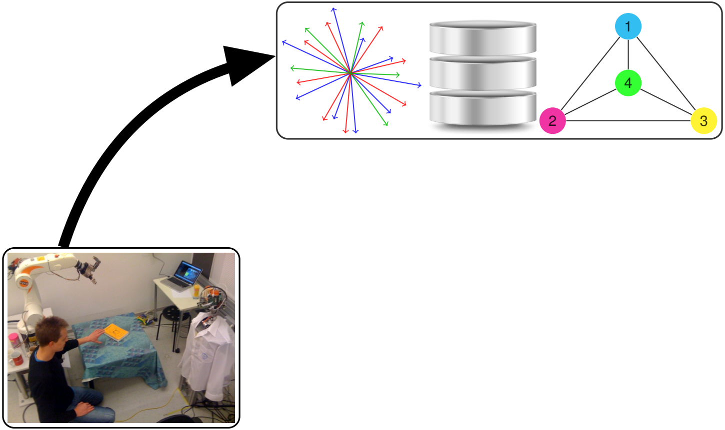
Structural Models

Conclusion

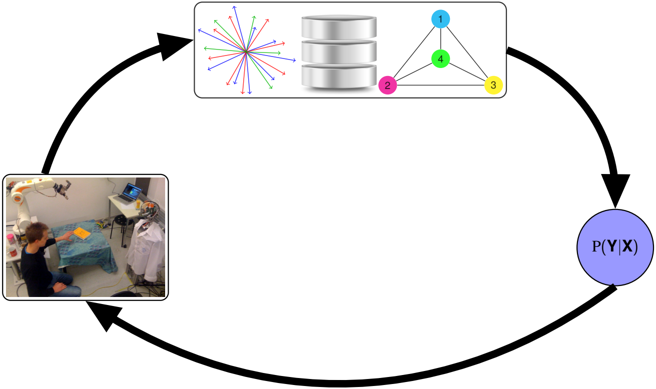
Abstraction



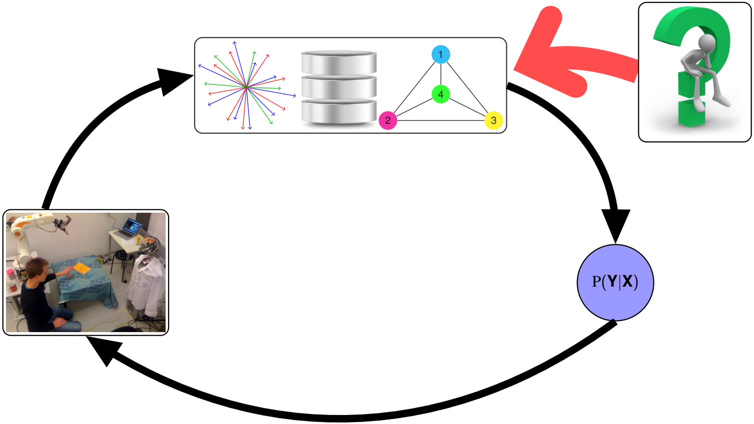
Abstraction



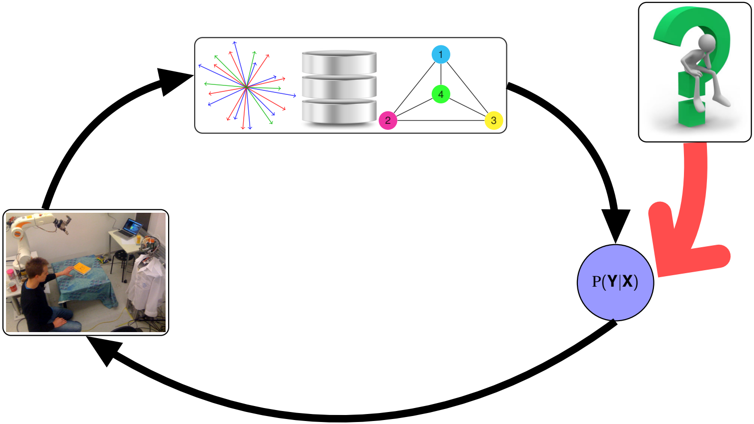
Abstraction



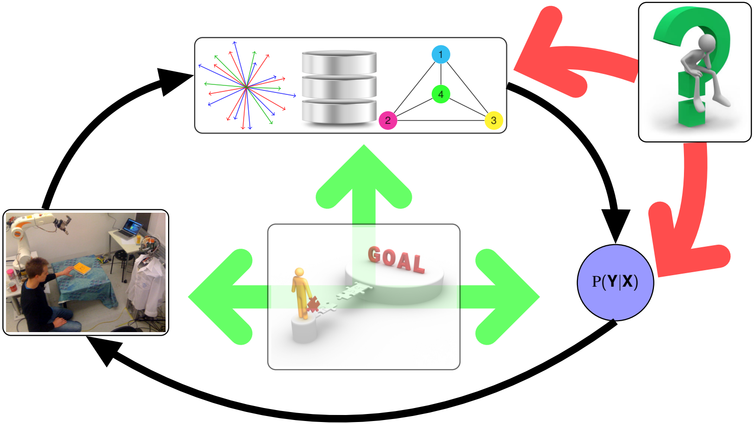
Abstraction



Abstraction



Abstraction



Representation

- Internal representation
  - basis for reasoning
  - unobservable
- External representation
  - Facilitates communication
  - Agreed/negotiated
  - Aware

"Grammar"

- Facilitates reasoning
  - Rules of generalisation
  - Ex *Triangle inequality*
- Preferential Representation
  - "Simple" structurally

Representation

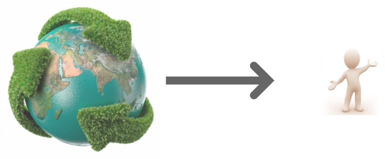
Representation

- Representation result of capturing
- Over-interpretation on
- Same yes, similar no
  - *success of NN, RBF*

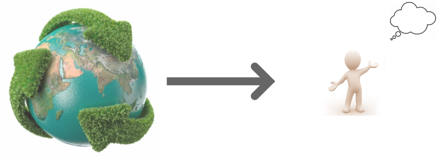
Scenario

Scenario

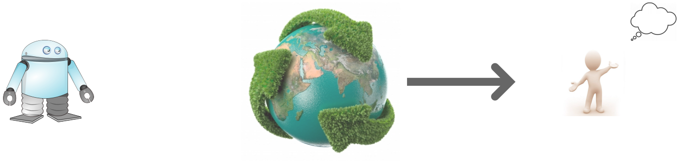
Scenario



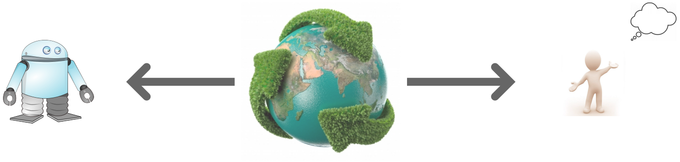
Scenario



Scenario

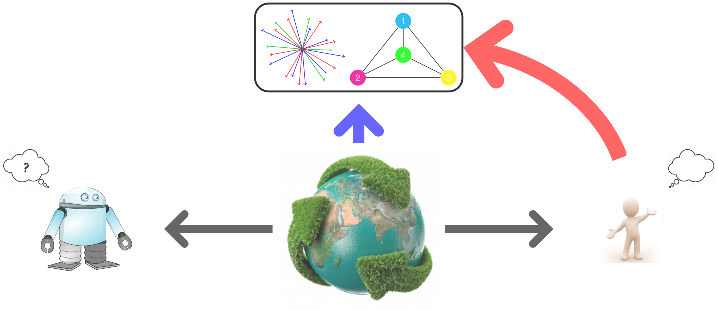


Scenario

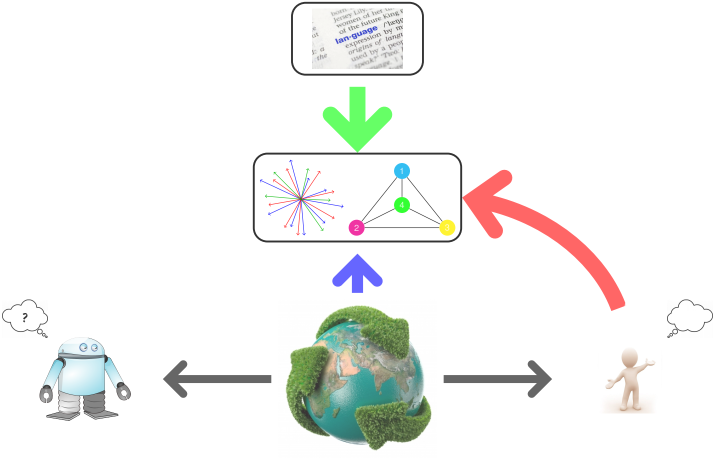


Scenario

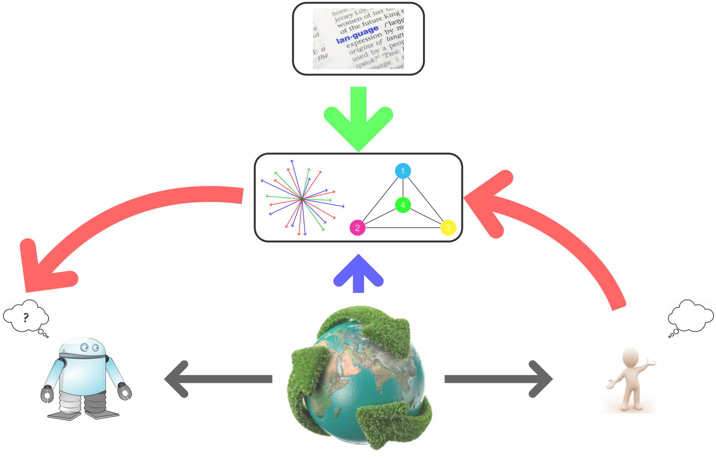
Scenario



Scenario



Scenario



External

Information

- Sensory data
  - images
  - depth
  - ...

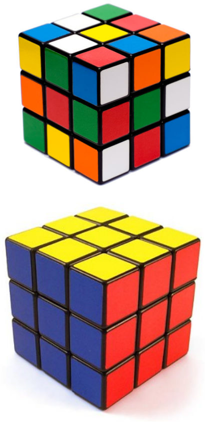
Language/Grammar

- "*Mathematical*"
  - similarity
  - integration/derivation
  - generalisation

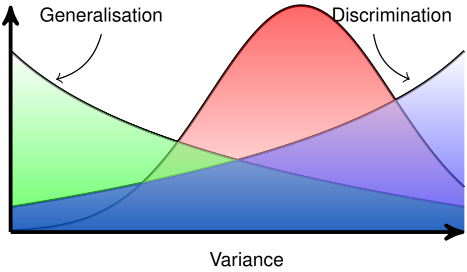
Motivation

What to represent?

- Task dependent
- Very rich sensory domain
  - Generalisation **not** discrimination
- Generalising variance?
  - structure?
  - appearance?



The Importance of Structure[1]

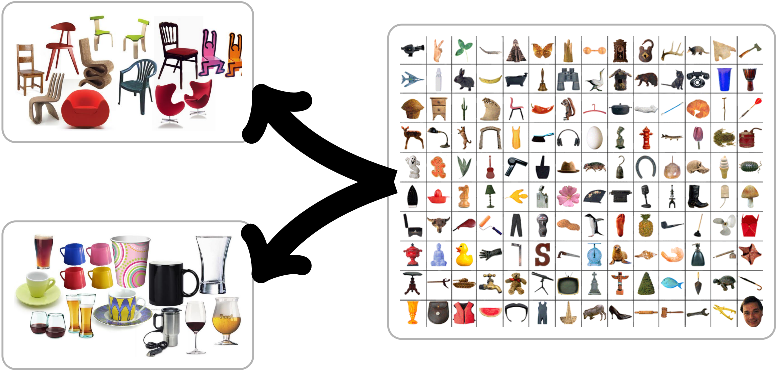


[1] Ek and Kragic [2011]

The Importance of Structure[1]

[1] Ek and Kragic [2011]

The Importance of Structure[1]



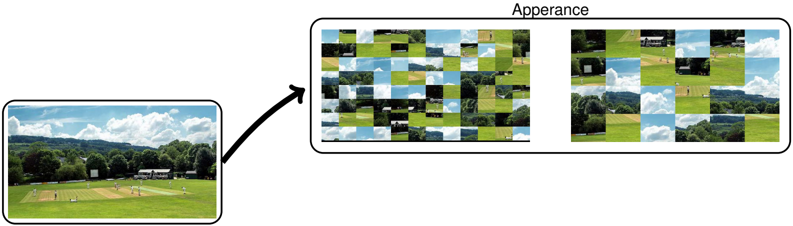
[1] Ek and Kragic [2011]

The Importance of Structure[1]

[1] Ek and Kragic [2011]

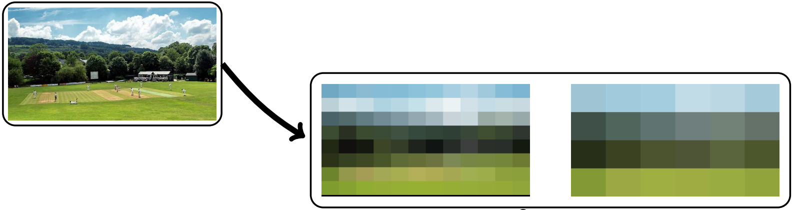
Current approach

Current approach



- 438 – 5, Alastair Cook Gray Nicolls bat

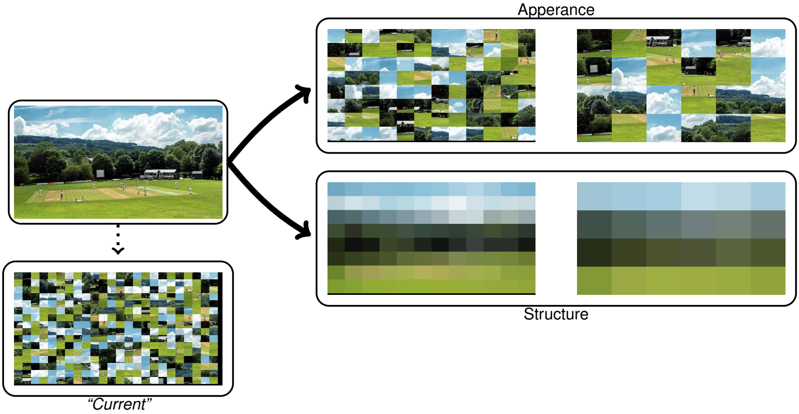
Current approach



Structure

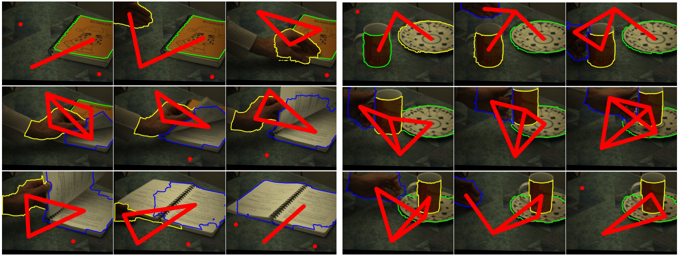
- Summer, Field, Outdoor

Current approach



- *Worst of both worlds*

Action Representation

 2

---

2 Aksoy et al. [2010]

Text analysis

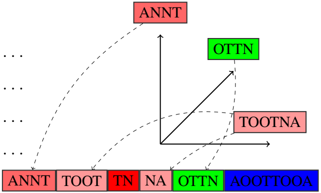
Amiga demos are demos created for the Commodore Amiga home computer.

A "demo" is a demonstration of the multimedia capabilities of a computer (or more to the point, a demonstration of the skill of the demo's constructors). There was intense rivalry during the 1990s among the best programmers, graphic artists and computer musicians to continually outdo each other's demos. Since the Amiga's hardware was more or less fixed (unlike today's PC industry, where arbitrary combinations of hardware can be put together), there was competition to test the limits of that hardware and perform theoretically "impossible" feats by refactoring the problem at hand. The Amiga was the undisputed leader of mainstream multimedia computing in the late 1980s and early 1990s, though it was inevitably overtaken by PC architecture.

- “No” appearance problem
- All information in ordering (1D Structure)

String Feature Space[3]

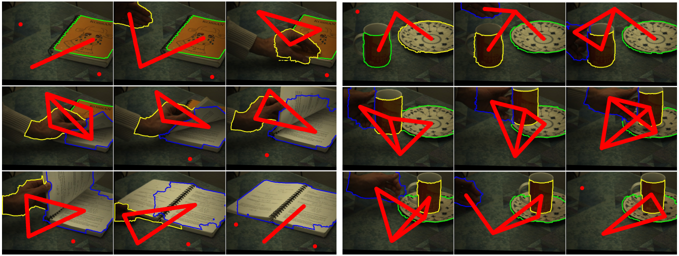
$$\begin{array}{lcllllll}
 & & \mathrm{AA} & \mathrm{NN} & \mathrm{AN} & \mathrm{AT} & \mathrm{NT} & \dots \\
\Phi(\mathrm{ANT}) & = & [\ 0 & 0 & \lambda^2 & \lambda^3 & \lambda^2 & \dots \\
\Phi(\mathrm{ANNT}) & = & [\ 0 & \lambda^2 & \lambda^2+\lambda^3 & \lambda^4 & \lambda^2+\lambda^3 & \dots \\
\Phi(\mathrm{AATN}) & = & [\ \lambda^2 & 0 & \lambda^3+\lambda^4 & \lambda^2+\lambda^3 & 0 & \dots
\end{array}$$



- Infinite dimensional representation **space**
- Kernel finite dimensional
- Inner-product efficiently computed

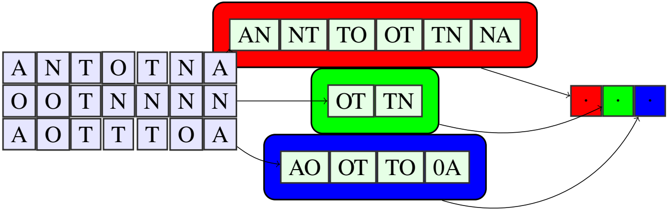
[3] Lodhi et al. [2002]

Action Representation

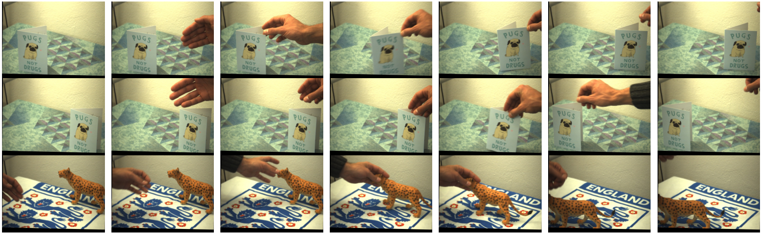
 [4]

[4] Aksoy et al. [2010]

Action Representation



Action Representation

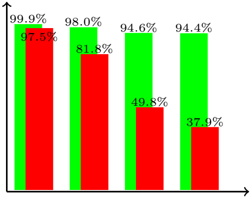


[4]

---

[4] Luo et al. [2011]

Action Representation



94.8	1.6	0.4	3.2
0.4	99.2	0.4	0
2.0	1.2	96.8	0
0.8	0	0	99.3

75.2	4.0	18.8	2.0
6.0	81.2	12.8	0
10.0	6.8	81.2	2.0
1.6	0.4	8.4	89.6

47.6	7.2	11.6	33.6
34.8	34.0	10.0	21.2
40.8	6.4	30.0	22.8
10.0	1.2	1.2	87.6

57.2	3.6	0.8	38.4
53.6	10.8	6.4	29.2
61.6	0.8	8.0	29.6
22.8	0.8	0.8	75.6

99.7	0	0.3	0
0	100	0	0
0	0.1	99.9	0
0	0	0	100

99.3	0	0.6	0.1
0	99.7	0.3	0
1.6	1.4	97	0
1.3	2.7	0.1	95.9

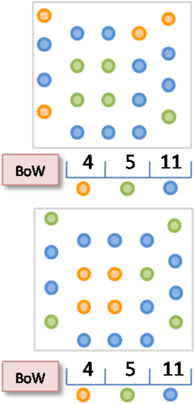
91.7	2.8	4.4	1.1
1.3	96.7	1.6	0.4
4.2	2.4	93.3	0.1
1.7	1.7	0.1	96.5

90.2	2.2	7.1	0.4
0.4	96.4	2.9	0.3
6.6	1.2	92.1	0.3
0.5	0.7	0	98.8

Objects[4]

Feature Representation

- Local representation
- Encode order
- Distribution of order

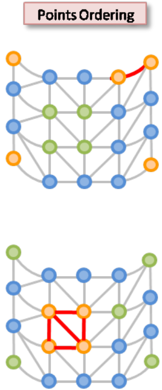


[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Objects[4]

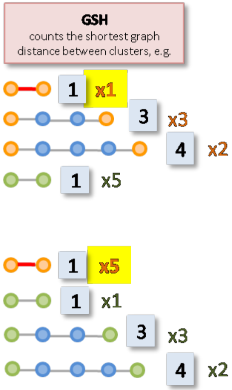
Feature Representation

- Local representation
- Encode order
- Distribution of order



[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Objects[4]

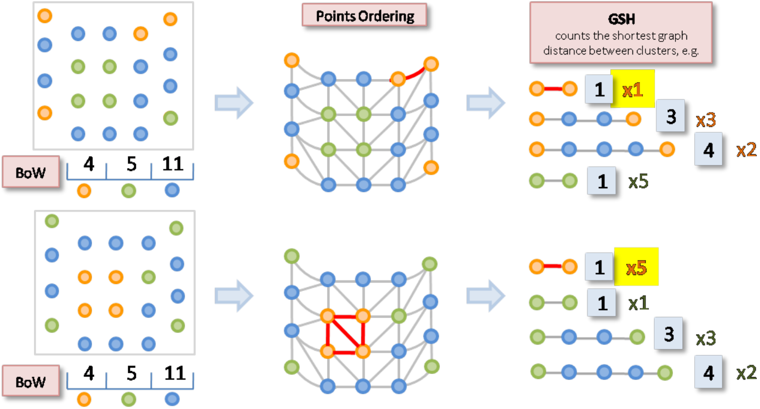


Feature Representation

- Local representation
- Encode order
- Distribution of order

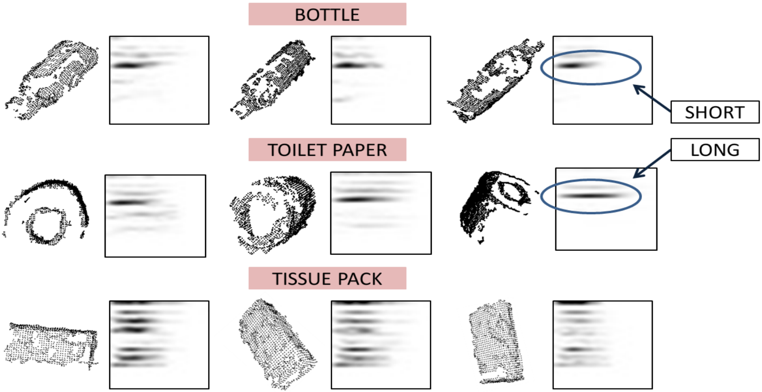
[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Objects[4]



[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Objects[4]

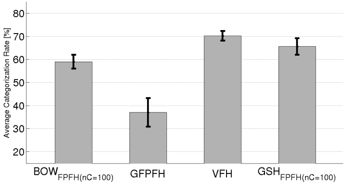


[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Objects[4]

Test Setting

- Synthetic
- Real
- Real & different pose,scale
- Synthetic & full,partial

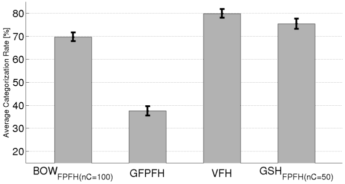


[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Objects[4]

Test Setting

- Synthetic
- Real
- Real & different pose,scale
- Synthetic & full,partial

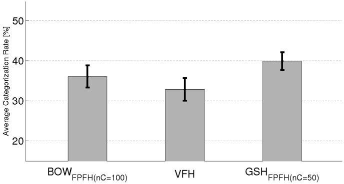


[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Objects[4]

Test Setting

- Synthetic
- Real
- Real & different pose,scale
- Synthetic & full,partial

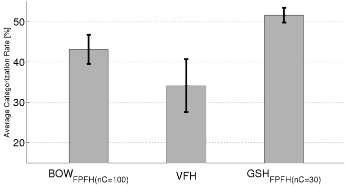


[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Objects[4]

Test Setting

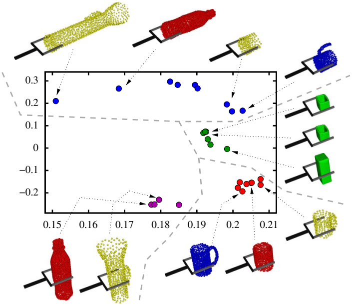
- Synthetic
- Real
- Real & different pose,scale
- Synthetic & full,partial



[4] In submission: Marianna Madry, Renaud Detry, Kaiyu Hang

Grasping[5]

- Pre-segmentation of objects
- Exploit structure in joint object and grasp space
- Part based generalisation



[5] Detry et al. [2012]

Grasping[5]

[5] Detry et al. [2012]

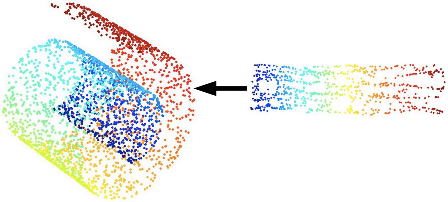
Introduction

Structural Representations

Structural Models

Conclusion

Re-representations

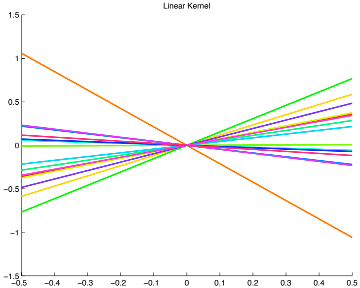


- Preference: low-dimensional
- Linearity
- Observed data $\mathbf{Y} \in \Re^{N \cdot D}$
- Underlying intrinsic representation $\mathbf{X} \in \Re^{N \cdot q}$
- Generative mapping: $\mathbf{y}_i = f(\mathbf{x}_i)$

Gaussian Processes: Prior

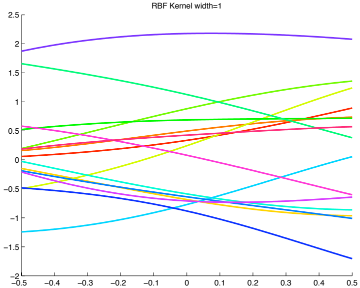
- Distribution over infinite objects: **functions.**

Gaussian Processes: Prior



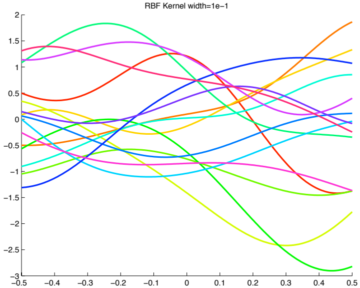
- Distribution over infinite objects: **functions.**

Gaussian Processes: Prior



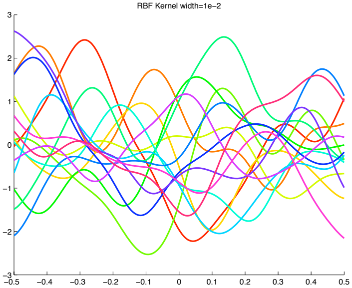
- Distribution over infinite objects: **functions.**

Gaussian Processes: Prior



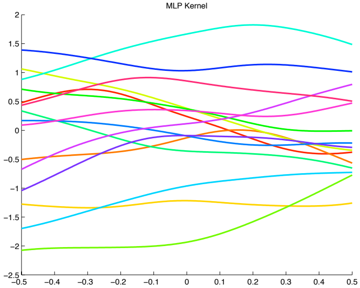
- Distribution over infinite objects: **functions.**

Gaussian Processes: Prior



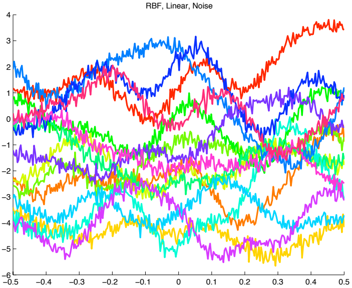
- Distribution over infinite objects: **functions.**

Gaussian Processes: Prior



- Distribution over infinite objects: **functions.**

Gaussian Processes: Prior



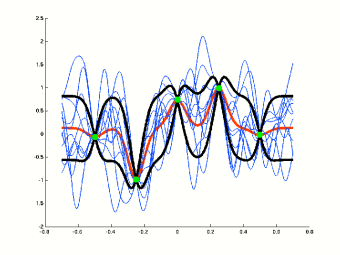
- Distribution over infinite objects: **functions.**

Gaussian Processes: Posterior

Combine prior with observed data

$$
\begin{aligned}
\mathbf{y}_*|\mathbf{X}_*,\mathbf{X},\mathbf{y} \sim & \mathcal{N}(K(\mathbf{X}_*,\mathbf{X})K(\mathbf{X},\mathbf{X})^{-1}\mathbf{y}, \\
& , K(\mathbf{X}_*,\mathbf{X}_*) - K(\mathbf{X}_*,\mathbf{X})K(\mathbf{X},\mathbf{X})^{-1}K(\mathbf{X},\mathbf{X}_*))
\end{aligned}
$$

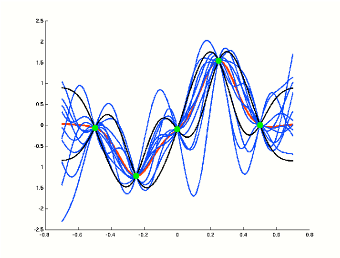
Gaussian Processes: Posterior



Combine prior with observed data

$$\mathbf{y}_*|\mathbf{X}_*,\mathbf{X},\mathbf{y} \sim \mathcal{N}(K(\mathbf{X}_*,\mathbf{X})K(\mathbf{X},\mathbf{X})^{-1}\mathbf{y},$$
$$,K(\mathbf{X}_*,\mathbf{X}_*) - K(\mathbf{X}_*,\mathbf{X})K(\mathbf{X},\mathbf{X})^{-1}K(\mathbf{X},\mathbf{X}_*))$$

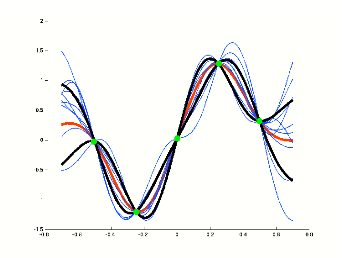
Gaussian Processes: Posterior



Combine prior with observed data

$$\mathbf{y}_*|\mathbf{X}_*,\mathbf{X},\mathbf{y} \sim \mathcal{N}(K(\mathbf{X}_*,\mathbf{X})K(\mathbf{X},\mathbf{X})^{-1}\mathbf{y},$$
$$,K(\mathbf{X}_*,\mathbf{X}_*) - K(\mathbf{X}_*,\mathbf{X})K(\mathbf{X},\mathbf{X})^{-1}K(\mathbf{X},\mathbf{X}_*))$$

Gaussian Processes: Posterior



Combine prior with observed data

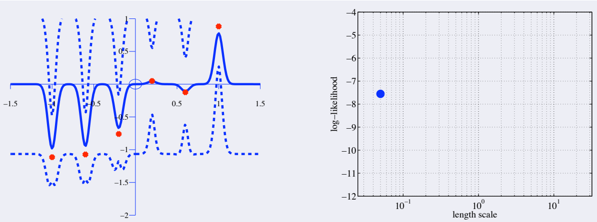
$$\mathbf{y}_*|\mathbf{X}_*, \mathbf{X}, \mathbf{y} \sim \mathcal{N}(K(\mathbf{X}_*, \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y}, \\ , K(\mathbf{X}_*, \mathbf{X}_*) - K(\mathbf{X}_*, \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}K(\mathbf{X}, \mathbf{X}_*))$$

Gaussian Processes: Marginal Likelihood[6]

$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{\textit{data-fit}}-\underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{\textit{complexity}}-\frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

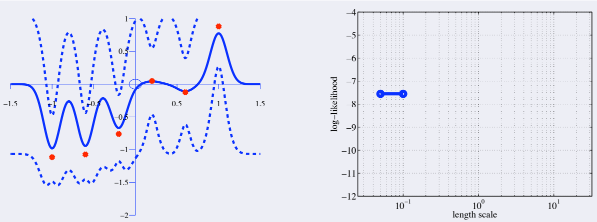
Gaussian Processes: Marginal Likelihood[6]



$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{data-fit}-\underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{complexity}-\frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

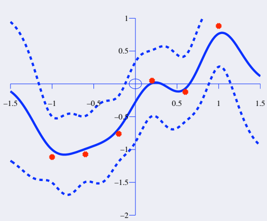
Gaussian Processes: Marginal Likelihood[6]

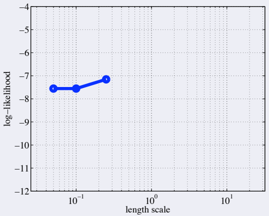


$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{data-fit}-\underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{complexity}-\frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

Gaussian Processes: Marginal Likelihood[6]

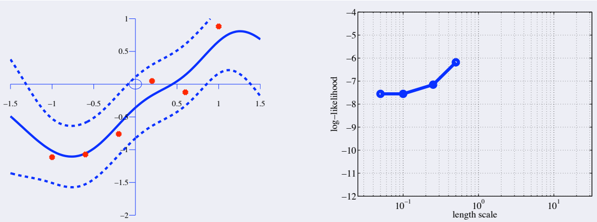




$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{\textit{data-fit}} - \underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{\textit{complexity}} - \frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

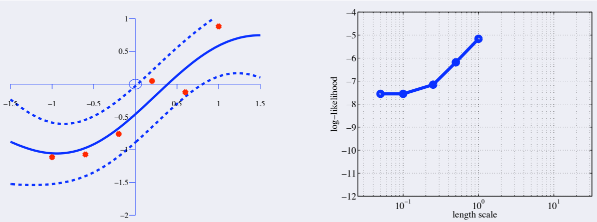
Gaussian Processes: Marginal Likelihood[6]



$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{data-fit} - \underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{complexity} - \frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

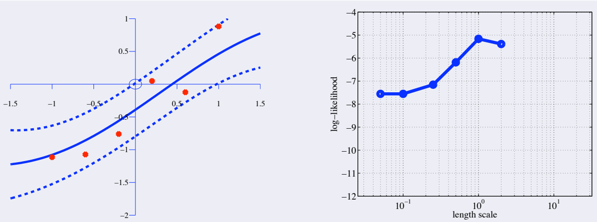
Gaussian Processes: Marginal Likelihood[6]



$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{\textit{data-fit}}-\underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{\textit{complexity}}-\frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

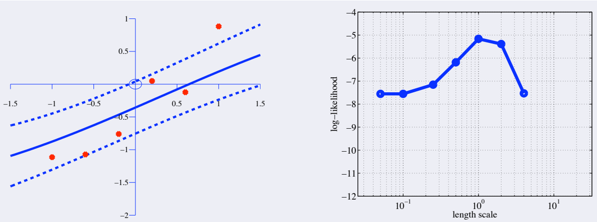
Gaussian Processes: Marginal Likelihood[6]



$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{data-fit}-\underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{complexity}-\frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

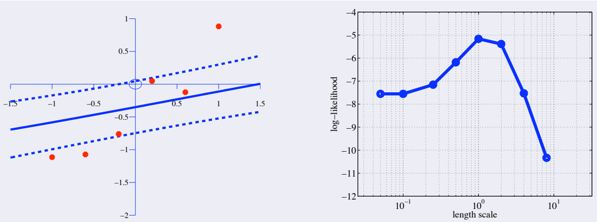
Gaussian Processes: Marginal Likelihood[6]



$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{data-fit}-\underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{complexity}-\frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

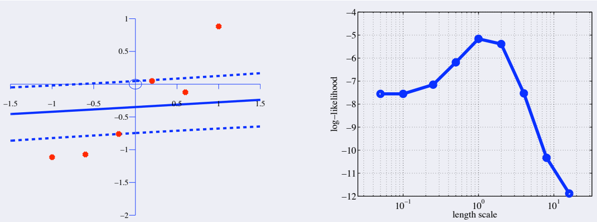
Gaussian Processes: Marginal Likelihood[6]



$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{\text{data-fit}}-\underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{\text{complexity}}-\frac{N}{2}\log 2\pi$$

[6] Images: Neil Lawrence

Gaussian Processes: Marginal Likelihood[6]



$$\underbrace{-\frac{1}{2}\mathrm{tr}\left(\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\beta^{-1}\mathbf{I})^{-1}\mathbf{y}\right)}_{\textit{data-fit}} - \underbrace{\frac{1}{2}\log\left(\det(\mathbf{K}+\beta^{-1}\mathbf{I})\right)}_{\textit{complexity}} - \frac{N}{2}\log 2\pi$$

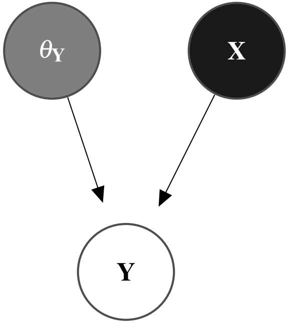
[6] Images: Neil Lawrence

Re-representation

GP-LVM[a]

[a] Lawrence [2005]

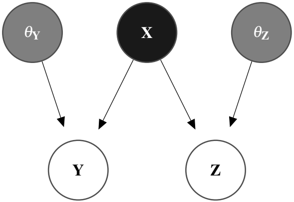
- Occam's Razor
  - Dimensionality
  - Co-variance function
- Sufficiently regularises problem



Shared Representations[7]

GP-LVM

- Fully shared
  - not CCA style
- Shared/Private

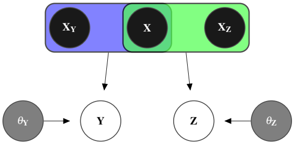


[7] Ek [2008]Salzmann et al. [2010]

Shared Representations[7]

GP-LVM

- Fully shared
  - not CCA style
- Shared/Private



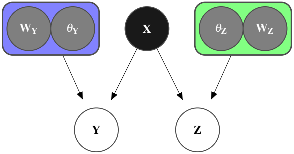
[7] Ek [2008]Salzmann et al. [2010]

Factorized Variance[8]

- Bayesian GP-LVM[a]
  - Prior on **X**
  - ARD

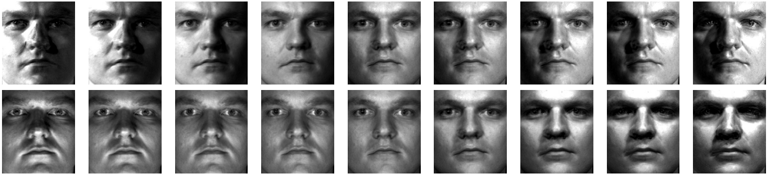
$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = (\sigma^Y_{ard})^2 e^{-\frac{1}{2}\sum_{q=1}^{Q} w_q^Y \left(x_{i,q} - x_{j,q}\right)^2}$$

[a] Titsias and Lawrence [2010]



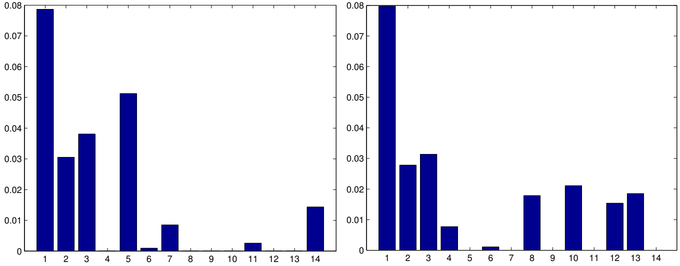
[8] In submission: Damianou, Lawrence, Titsias

Factorized Variance[8]



[8] In submission: Damianou, Lawrence, Titsias

Factorized Variance[8]

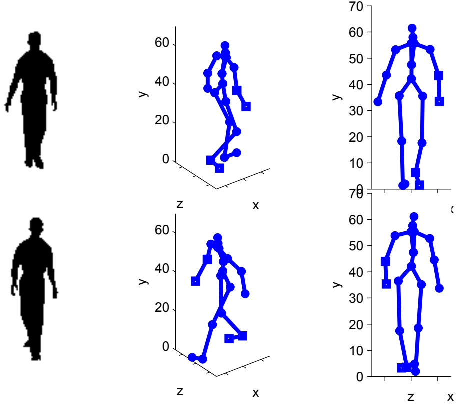


[8] In submission: Damianou, Lawrence, Titsias

Factorized Variance[8]

[8] In submission: Damianou, Lawrence, Titsias

Factorized Variance[8]



[8]In submission: Damianou, Lawrence, Titsias

Factorised Density[9]

Dimensionality Reduction I

- Conditional dependency structures,

$$p(\mathbf{X}) = \prod_i p(\mathbf{x}_i | \pi(\mathbf{x}_i), \theta_i, \mathbf{S})$$

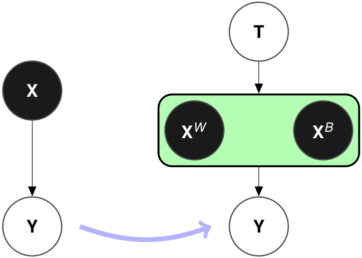
- Learning,
  - Parameters: θ_i
  - Structure: $\mathbf{S}$
    - Priors?
    - Carnality
- Heuristics for discrete data

[9] Song, Huebner, Hjelm

Factorised Density[9]

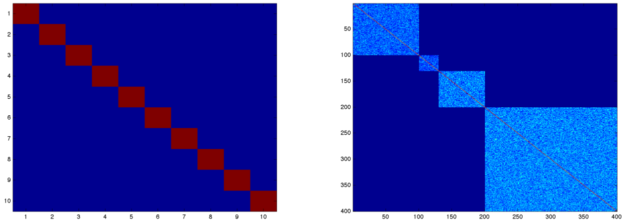
Dimensionality Reduction II

- Very ill-defined
- Re-representation
  - *"a mapping and configuration"*
- Prefer "clustered" re-representation



[9] Song, Huebner, Hjelm

Factorised Density[9]

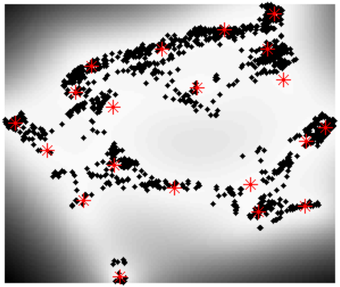


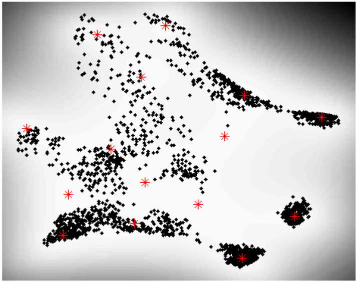
Objective

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{U}|\theta) = \int p(\mathbf{Y}|\mathbf{f}, \theta) p(\mathbf{f}|\mathbf{f}_U, \mathbf{X}, \theta) p(\mathbf{f}_U|\mathbf{U}|\theta) p(\mathbf{X}) p(\mathbf{U}|\theta) \mathrm{d}\mathbf{f} \mathrm{d}\mathbf{f}_U$$

[9] Song, Huebner, Hjelm

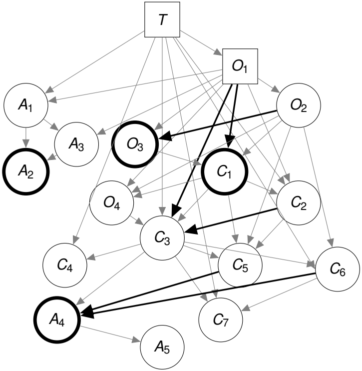
Factorised Density[9]

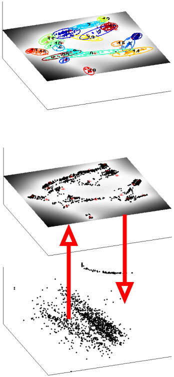




[9]Song, Huebner, Hjelm

Factorised Density[9]





[9] Song, Huebner, Hjelm

Factorised Density[10]

Glass

Hammer

Knife

Hand Over

[10] Ek et al. [2011]Song et al. [2011b]Song et al. [2011a]

Factorised Density[10]

Glass

Hammer

Knife

Pouring

[10] Ek et al. [2011]Song et al. [2011b]Song et al. [2011a]

Factorised Density[10]

Glass

Hammer

Knife

Tool-Use

[10] Ek et al. [2011]Song et al. [2011b]Song et al. [2011a]

Factorised Density[10]

Glass

Hammer

Knife

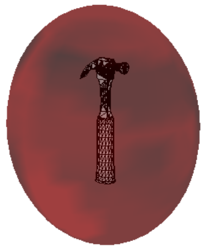
Hand Over

[10] Ek et al. [2011]Song et al. [2011b]Song et al. [2011a]

Factorised Density[10]

Glass

Hammer



Knife

Pouring

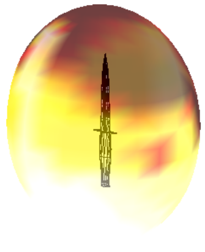
[10] Ek et al. [2011]Song et al. [2011b]Song et al. [2011a]

Factorised Density[10]

Glass

Hammer

Knife



Tool-Use

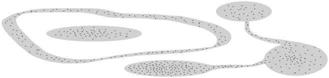
[10] Ek et al. [2011]Song et al. [2011b]Song et al. [2011a]

Structural Density[11]

Topology Respecting

- Structural properties
- Geometrical notion irrelevant
- Topological information
- Barcodes[a]

[a] Carlsson [2009]



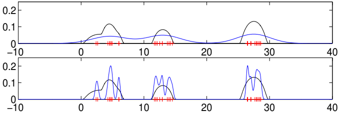
[11] In submission: Pokorny, Kjellström

Structural Density[11]

Topology Respecting

- Structural properties
- Geometrical notion irrelevant
- Topological information
- Barcodes[a]

[a] Carlsson [2009]



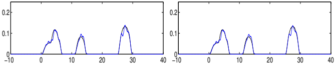
[11] In submission: Pokorny, Kjellström

Structural Density[11]

Topology Respecting

- Structural properties
- Geometrical notion irrelevant
- Topological information
- Barcodes[a]

[a] Carlsson [2009]



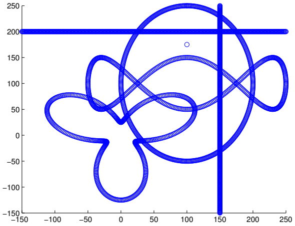
[11] In submission: Pokorny, Kjellström

Structural Density[11]

Topology Respecting

- Structural properties
- Geometrical notion irrelevant
- Topological information
- Barcodes[a]

[a] Carlsson [2009]



[11] In submission: Pokorny, Kjellström

Introduction

Structural Representations

Structural Models

Conclusion

Conclusions

- Generalisation not discrimination
- Less is sometimes more
- Model relevance

Future Work

- Multidimensional structure
- Different generalisations
- Latent space priors
- New kernels
  - know the characteristics of the space

e.o.f.

References I

E. Aksoy, A. Abramov, F. Wörgötter, and B. Dellen. Categorizing Object-Action Relations from Semantic Scene Graphs. In *IEEE International conference on robotics and automation*, 2010.

G. Carlsson. Topology and data. *American Mathematical Society*, 2009.

R. Detry, C. H. Ek, M. Pronobis, J. Piater, and D. Kragic. Generalizing Grasps Across Partly Similar Objects. In *IEEE International conference on robotics and automation*, 2012.

C. H. Ek. GP-LVM for Data Consolidation. *Neural Information Processing Systems: Workshop on Learning from multiple sources*, 2008.

References II

C. H. Ek and D. Kragic. The importance of structure. *International Symposium on Robotic Research*, 2011.

C. H. Ek, D. Song, and D. Kragic. Learning Conditional Structures in Graphical Models from a Large Set of Observation Streams through efficient Discretisation. In *IEEE International Conference on Robotics and Automation, Workshop on Manipulation under Uncertainty*, 2011.

N. D. Lawrence. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *The Journal of Machine Learning Research*, 2005.

H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins. Text classification using string kernels. *The Journal of Machine Learning Research*, 2002.

References III

G. Luo, N. Bergström, C. H. Ek, and D. Kragic. Representing actions with Kernels. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.

M. Salzmann, C. H. Ek, R. Urtasun, and T. Darrell. Factorized Orthogonal Latent Spaces. *International Conference on Artificial Intelligence and Statistics*, 2010.

D. Song, C. H. Ek, K. Huebner, and D. Kragic. Embodiment-specific representation of robot grasping using graphical models and latent-space discretization. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011a.

References IV

D. Song, C. H. Ek, K. Huebner, and D. Kragic. Multivariate discretization for bayesian network structure learning in robot grasping. In *IEEE International conference on robotics and automation*, 2011b.

M. Titsias and N. D. Lawrence. Bayesian Gaussian Process Latent Variable Model. In *International Conference on Artificial Intelligence and Statistics*, 2010.