Learning Multimodal Transition Dynamics for Model-Based Reinforcement Learning

Thomas Moerland, Joost Broekens, Catholijn Jonker

Delft University of Technology
The Netherlands

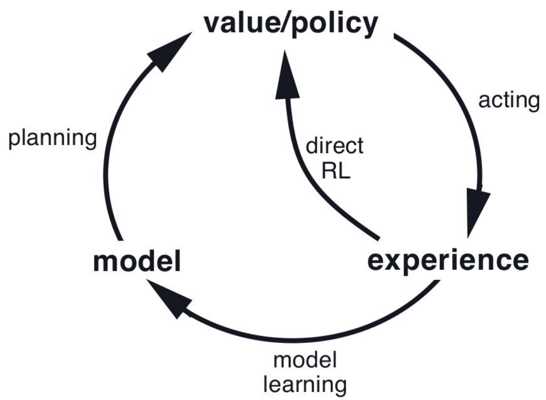
Content

1. Introduction
2. Conditional Variational Inference
3. Experiments
4. Conclusion

Introduction

Model-based RL:

1. Transition dynamics approximation (supervised learning)
2. Planning



Sutton, Richard S., and Andrew G. Barto. *Reinforcement learning: An introduction*. Vol. 1. No. 1. Cambridge: MIT press, 1998.

Introduction

Model-based RL:

1. Transition dynamics approximation (supervised learning)
2. Planning

Benefits:

- Data efficiency
- Targeted exploration

Introduction

Model-based RL:

1. Transition dynamics approximation (supervised learning)
2. Planning

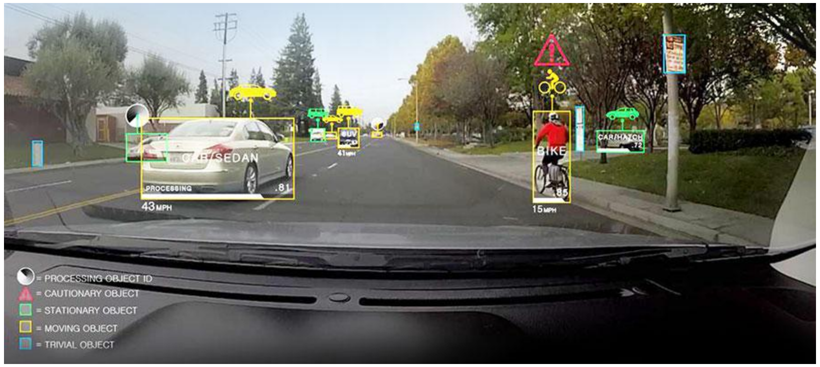
Benefits:

- Data efficiency
- Targeted exploration

Challenges (ad 1.):

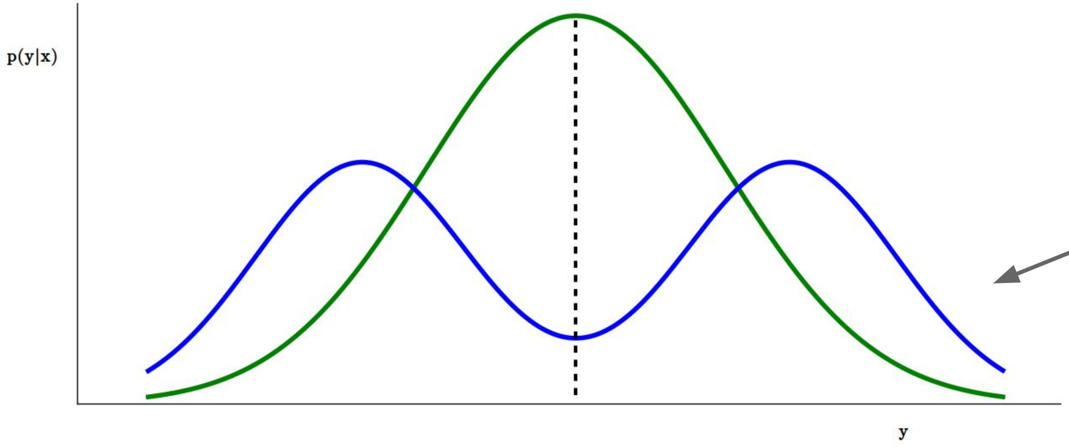
- Stochasticity
- (High-dimensionality)

Stochasticity



Stochasticity

$$f:\ x \rightarrow p(y)$$

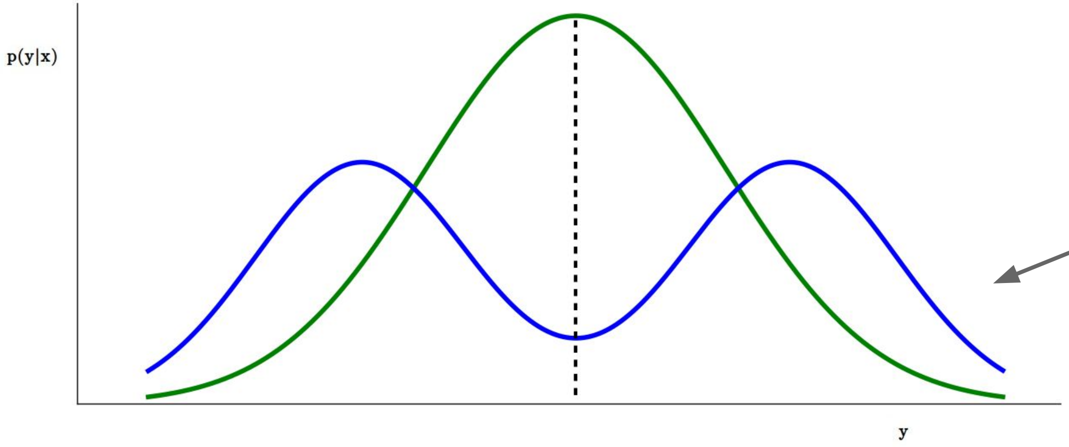


Multimodal stochasticity:

multiple local maxima

Stochasticity

$$f:\ x \rightarrow p(y)$$



Multimodal stochasticity:
multiple local maxima

1) Mean-squared error (MSE)/deterministic prediction fails
2) Most density estimation techniques (e.g. Gaussian mixtures) don't scale

Solution: Deep Generative Model

Variational Auto-Encoder (VAE)[1]

= generative model

for **p(y)**

↓

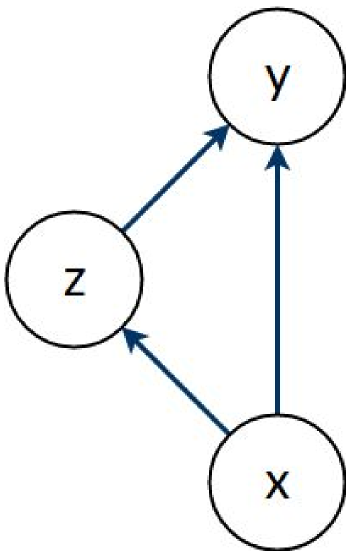
Modify to the conditional setting

p(y|x)

1. Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).

2. Conditional Variational Auto-Encoder

Introduce latent variables **z** to obtain more expressivity in the marginal:



$$p(y|x) = \int p(y|z,x)p(z|x)dz$$

marginal (multimodal)

decoder (*unimodal*)

prior (*unimodal*)

2. Conditional Variational Auto-Encoder

Problem:

a) **z** is not observed (*what value to plug in?*)
b) Posterior p(z|x,y) is not tractable

2. Conditional Variational Auto-Encoder

Problem:

a) **z** is not observed (*what value to plug in?*)
b) Posterior $p(z|x,y)$ is not tractable

Solution:

a) Parametric inference network $q(z|x,y)$ that approximates $p(z|x,y)$

2. Conditional Variational Auto-Encoder

Problem:

a) **z** is not observed (*what value to plug in?*)
b) Posterior p(z|x,y) is not tractable

Solution:

a) Parametric inference network q(z|x,y) that approximates p(z|x,y)
b) Maximize the Evidence Lower Bound (ELBO):

$$\log p(y|x) \geq \mathbb{E}_{z\sim q(\cdot|x,y)}[\log p_\theta(y|z,x)] - D_{\mathrm{KL}}[q_\phi(z|x,y)||p_\phi(z|x)] = \mathcal{L}(y|x;\theta,\phi)$$

2. Conditional Variational Auto-Encoder

Problem:

a) **z** is not observed (*what value to plug in?*)
b) Posterior p(z|x,y) is not tractable

Solution:

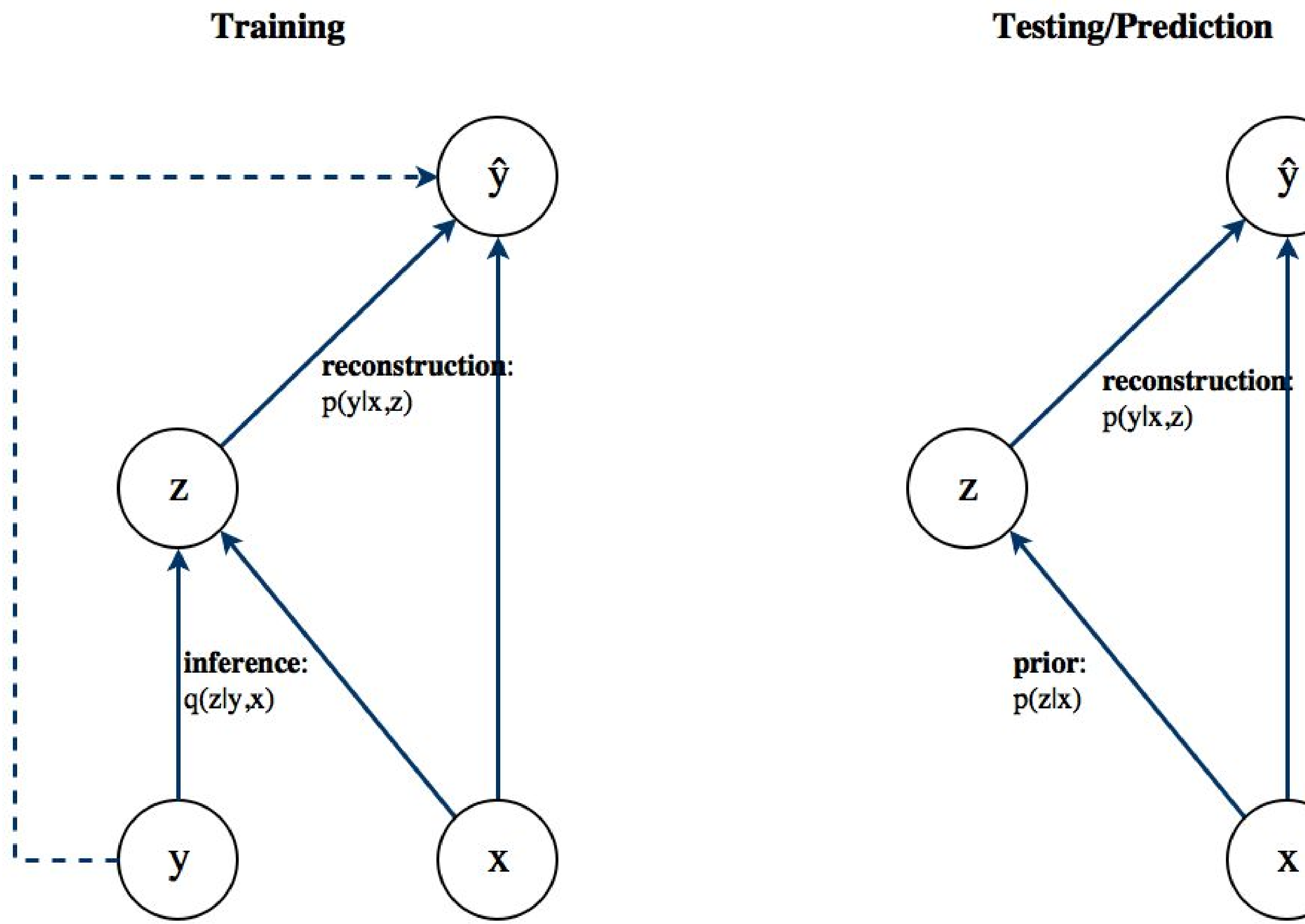
a) Parametric inference network q(z|x,y) that approximates p(z|x,y)
b) Maximize the Evidence Lower Bound (ELBO):

$$\log p(y|x) \geq \mathbb{E}_{z\sim q(\cdot|x,y)}[\log p_\theta(y|z,x)] - D_{\mathrm{KL}}[q_\phi(z|x,y)||p_\phi(z|x)] = \mathcal{L}(y|x;\theta,\phi)$$

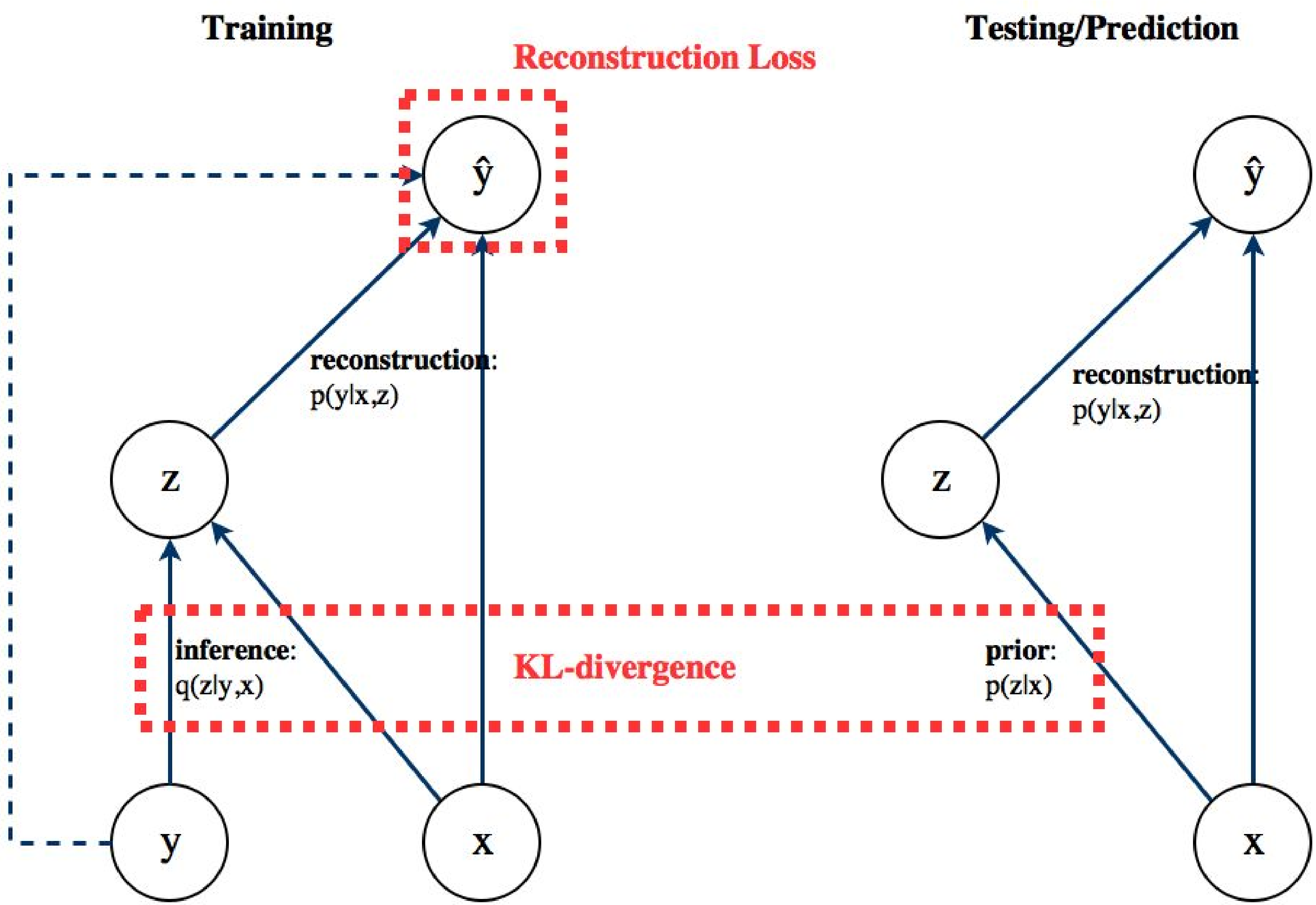
Reconstruction

KL-term/Regularization

2. CVAE: Computation



2. CVAE: Computation

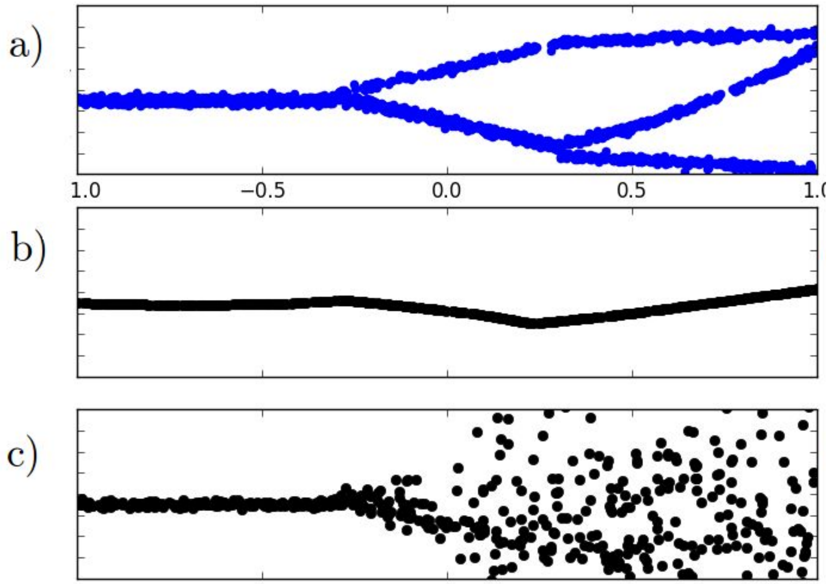


2. CVAE

I. Scales to larger dimensions

II. Training details in the paper:
   A. Reparametrization of z variables
   B. Continuous versus discrete latent variables z
   C. Importance sampling
   D. α-divergence training

3. Experiments

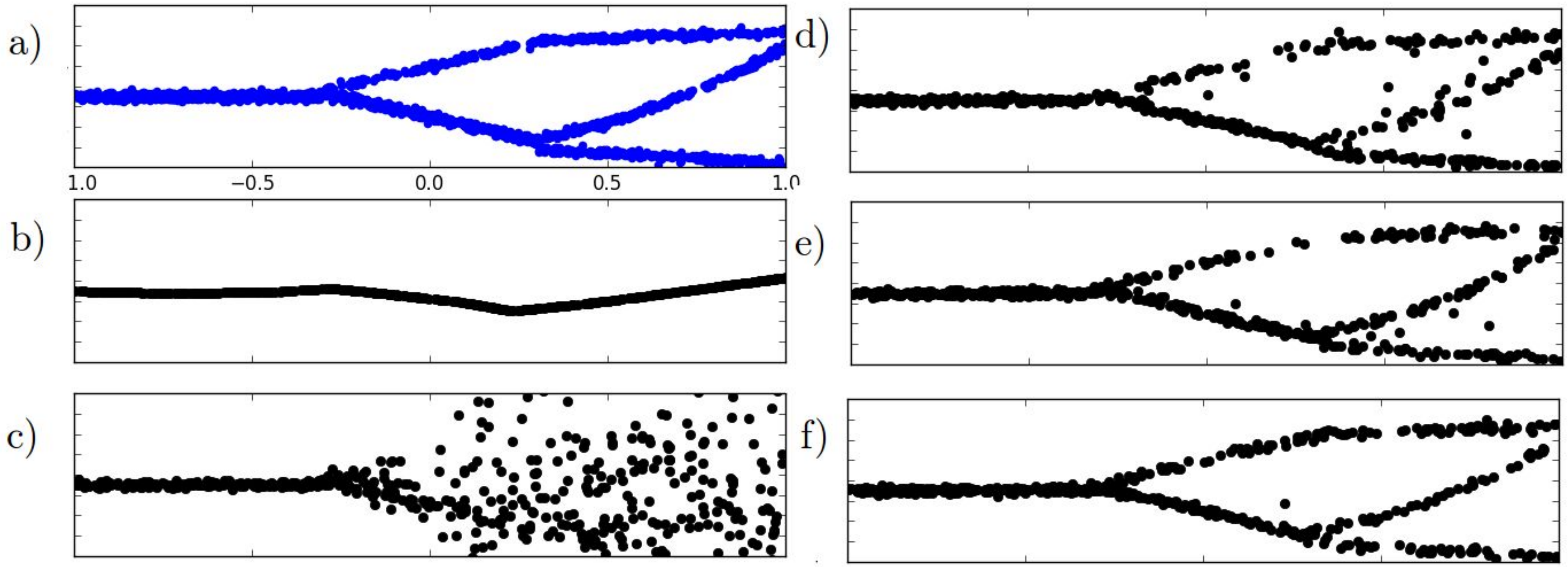


a) True data

b) Mean-squared error

c) MLP with noise input **z**

3. Experiments



a) True data

b) Mean-squared error

c) MLP with noise input **z**

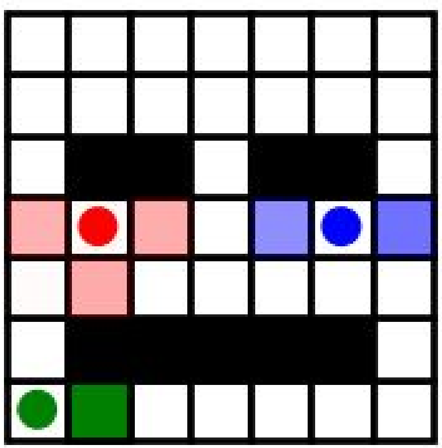
d) **CVAE** with contin. **z**

e) **CVAE** with contin. **z** + flow

f) **CVAE** with discrete **z**

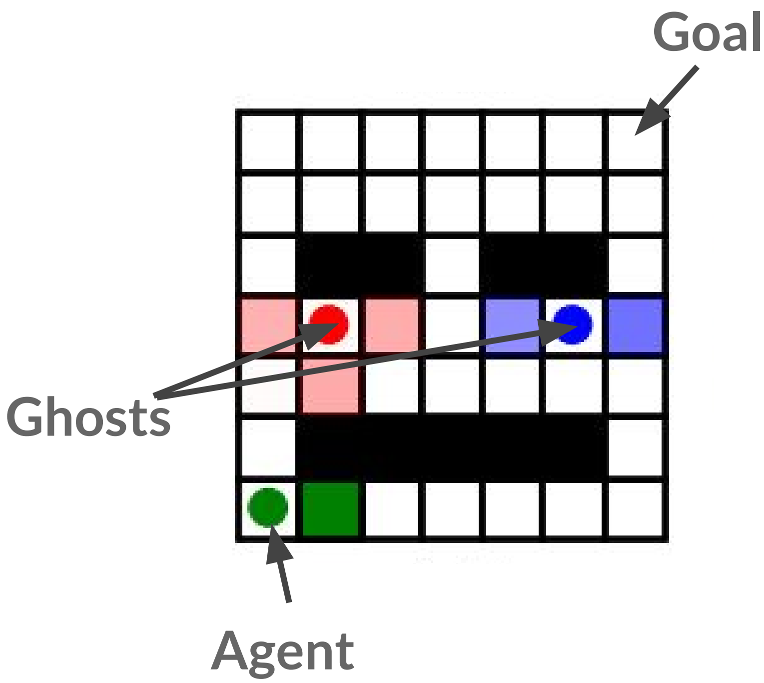
3. Experiments

Gridworld



3. Experiments

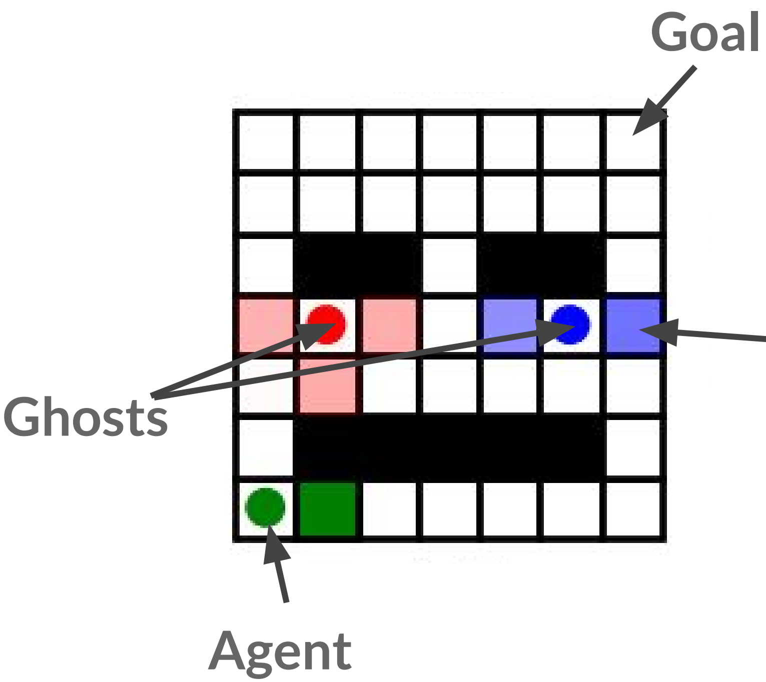
Gridworld



Agents = circles

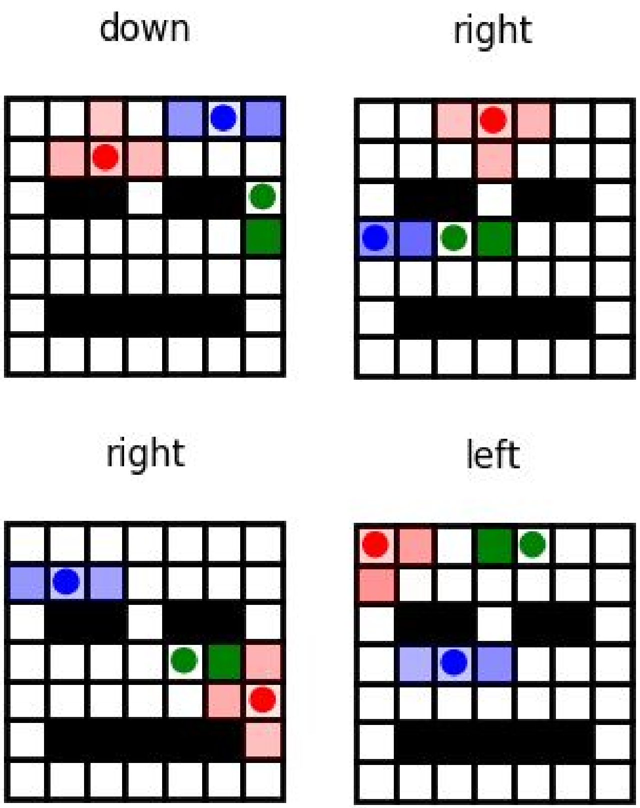
3. Experiments

Gridworld

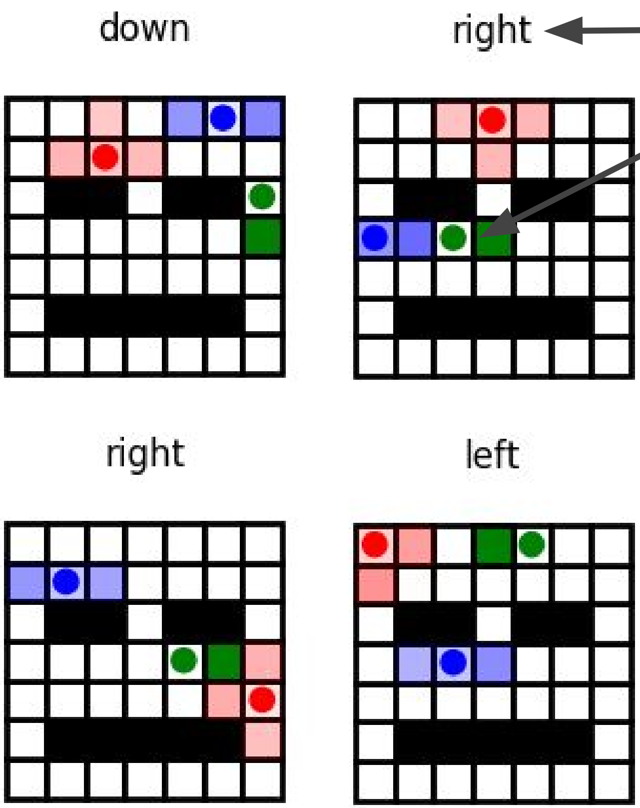


Agents	=	circles
Predictions	=	coloured boxes
		(prob. = intensity)

3. Experiments

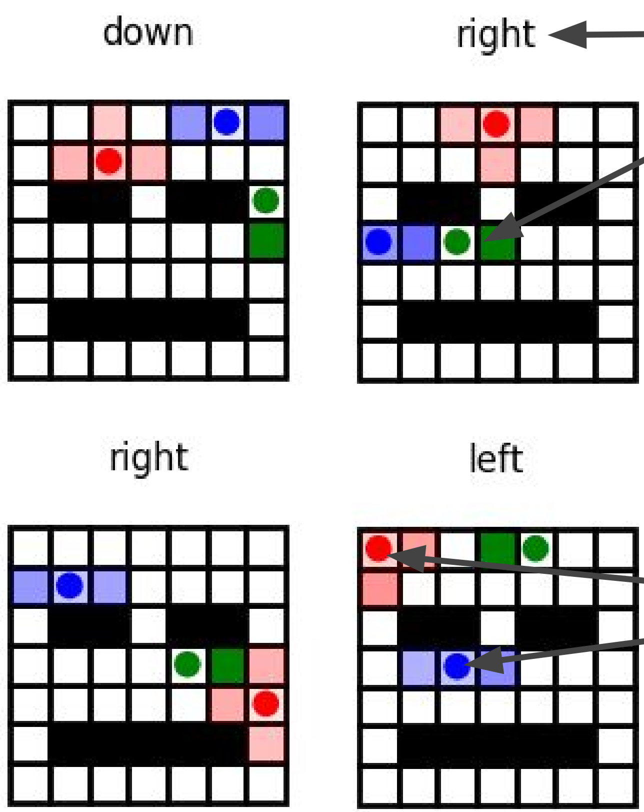


3. Experiments



Learns deterministic agent (action-conditional)

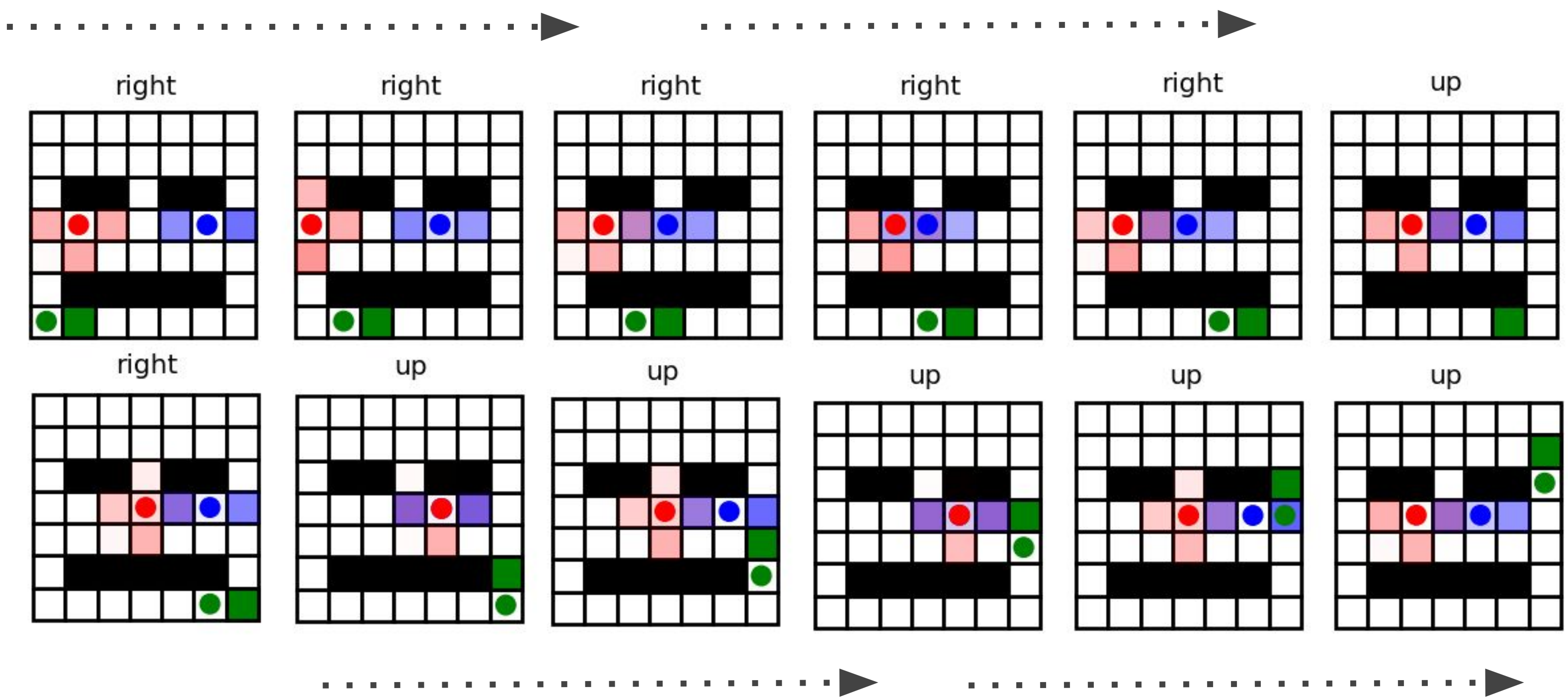
3. Experiments



Learns deterministic agent (action-conditional)

Discriminates stochastic agents (red: all four directions, blue: left-right)

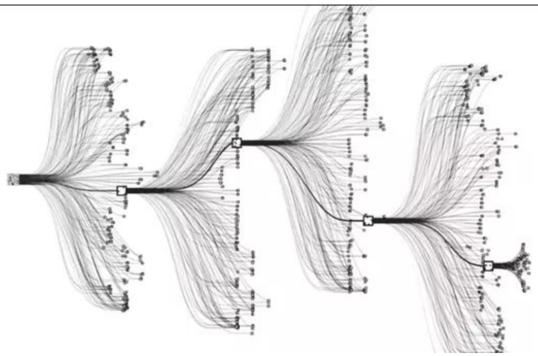
3. Experiments: on-policy predictions



Full roll-out in model

4. Future Work

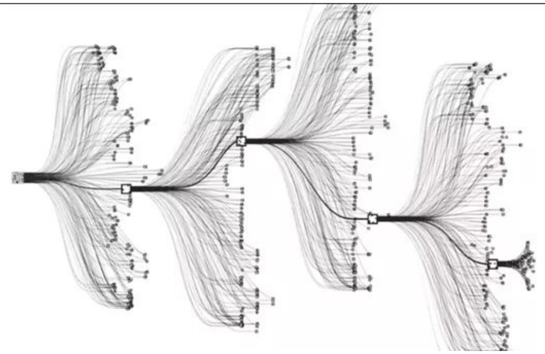
1 **2** **3**



Planning
(under uncertainty)

4. Future Work

1



Planning
(under uncertainty)

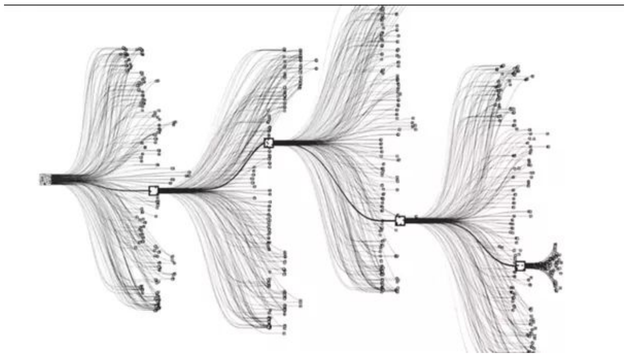
2

Higher-dimensions

3

4. Future Work

1



Planning
(under uncertainty)

2

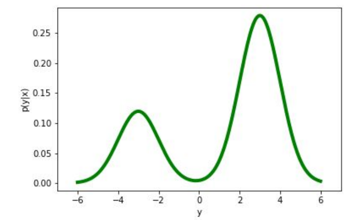
Higher-dimensions

3

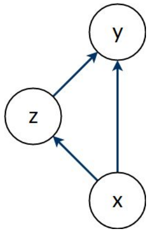
Memory/
Partial-observability

4. Conclusion

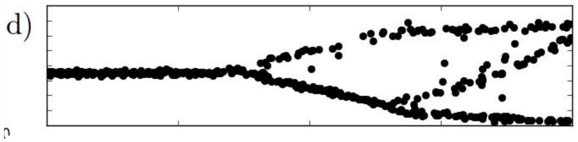
1. Stochasticity is a fundamental problem in model-based RL



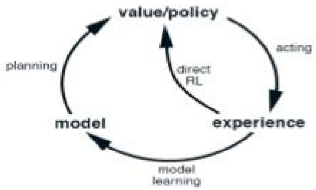
2. Conditional Variational Auto-Encoder (CVAE) learns complex p(y|x) in high dimensions



3. Experiments show multimodal predictions



4. Useful for model-based RL researchers



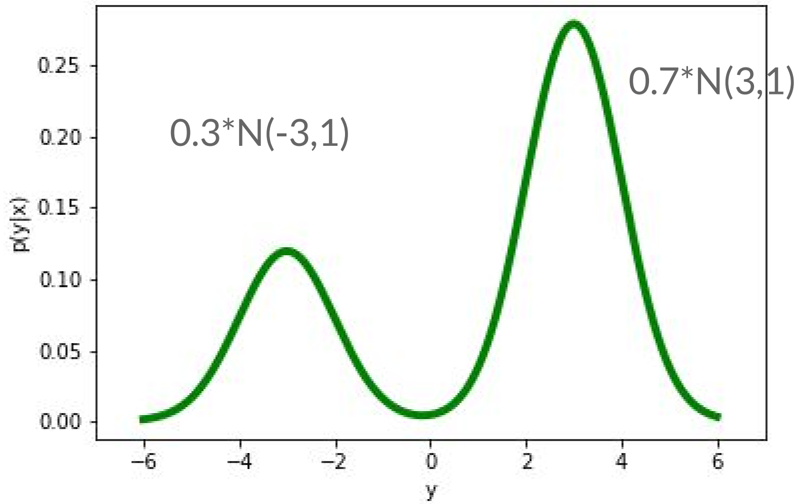
Thanks!

Any questions?

Can always reach me at:
T.M.Moerland@tudelft.nl

Full code online:
www.github.com/tmoer/multimodal_varinf

2. CVAE: Illustration



1) Specify size of **z**-space: **z** in {0,1}
2) Present datapair (x=x,y=3)
3) Inference network predicts we should sample z=1
4) Recognition network predicts (given the sampled z) to sample from N(3,1)
5) Repeat over datapairs (mini-batches): KL divergence with prior will learn p_0=0.3,p_1=0.7
6) At test time: sample from prior, and then from the conditional Gaussian