Deep Probabilistic Graphical Models

Hung Bui

Director
VinAI Research
Hanoi, Vietnam

About VinAI Research

- Lab Location: Hanoi, Vietnam. https://www.vinai.io/
- Funded by VinGroup, the largest private company in Vietnam.
- Publication focused.
- The first of its kind in South East Asia (not counting Singapore).
- We're hiring scientists, engineers and AI Residents.

Roadmap

- Motivation
- Deep Probabilistic Graphical Models
- Training Deep PGMs via Variational Inference
- Representation Learning with Deep PGMs
    - Disentangled Representations
    - Planning in Representation Space

Motivation: Unsupervised Learning, Generative Modeling

where can i buy an affordable stationary bike ? try this place , they have every type imaginable with prices to match . http : UNK </s>

if our economy collapses , will canada let all of us cross their border ? no , a country would have to be stupid to let that many people cross their borders and drain their resources . </s>

does the flat earth society still exist ? i 'm curious to know whether the original society still exists . i 'm not especially interested in discussion about whether the earth is flat or round . although there is no currently active website for the society , someone (apparently a relative of samuel UNK) maintains the flat earth society forums . this website , which offers a discussion forum and an on-line archive of flat earth society UNK from the 1970s and 1980s , represents a serious attempt to UNK the original flat earth society . </s>

Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale gan training for high fidelity natural image synthesis." arXiv preprint arXiv:1809.11096 (2018).
Kim, Yoon, et al. "Semi-amortized variational autoencoders." arXiv preprint arXiv:1802.02550 (2018).
Van Den Oord, Aäron, et al. "WaveNet: A generative model for raw audio." SSW 125 (2016).

Motivation: Representation Learning

Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale gan training for high fidelity natural image synthesis." arXiv preprint arXiv:1809.11096 (2018).

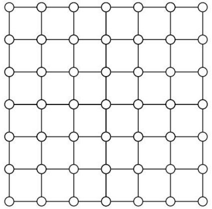
Motivation: Representation Learning

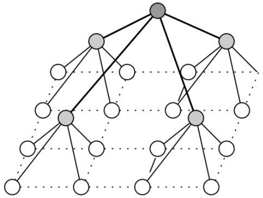
Varying the code of tense	
i thought the movie was too bland and too much	this was one of the outstanding thrillers of the last decade
i guess the movie is too bland and too much	this is one of the outstanding thrillers of the all time
i guess the film will have been too bland	this will be one of the great thrillers of the all time

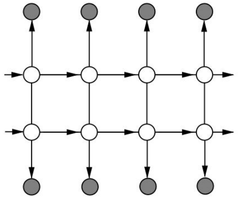
Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale gan training for high fidelity natural image synthesis." arXiv preprint arXiv:1809.11096 (2018).
Hu, Zhiting, et al. "Toward controlled generation of text." Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017.

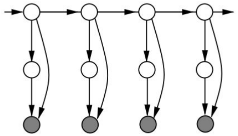
Probabilistic Graphical Models

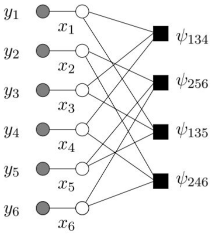
Graphical representation of probabilistic dependency in joint probability distributions

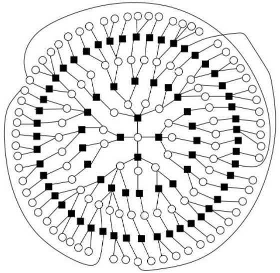






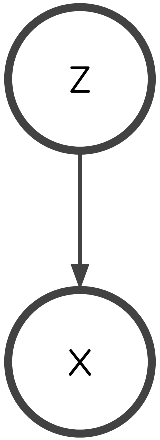






Wainwright, Martin J., and Michael I. Jordan. "Graphical models, exponential families, and variational inference." Foundations and Trends® in Machine Learning 1.1–2 (2008): 1-305.

Latent Variable Models



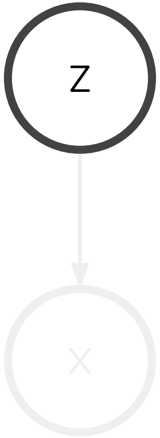
Node denotes an unconditional probability distribution

$Pr(Z)$

Linked Node denotes a conditional probability distribution

$Pr(X \mid Z = z)$

Latent Variable Models

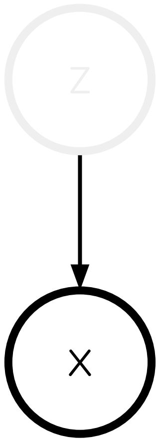


Node denotes an unconditional probability distribution
Pr(Z)

How to represent an unconditional probability distribution:

Parametric distributions Pr(Z ; γ)
Discrete: *Categorical*, *Binomial*, *Bernoulli*, etc
Continuous: *Gaussian*, *Gamma*, *Beta*, *Dirichlet*, etc

Latent Variable Models



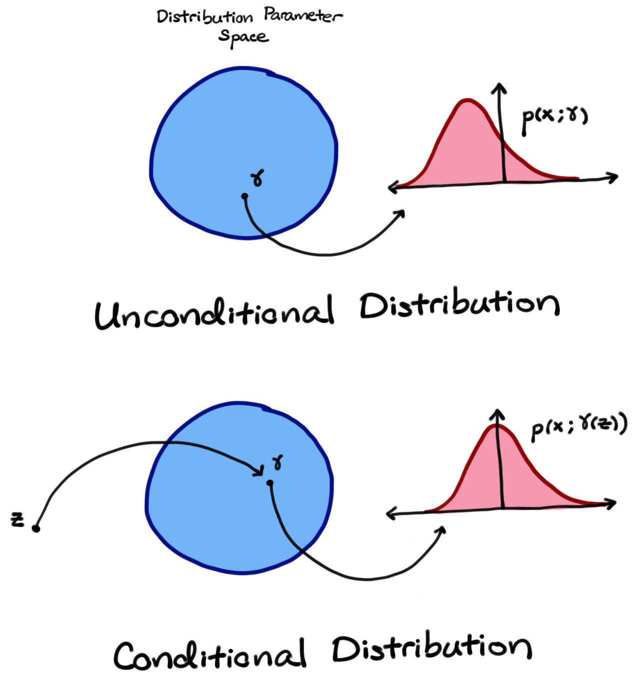
Linked Node denotes a conditional probability distribution
$Pr(X \mid Z = z)$

How to represent a conditional probability distribution:

- A mapping function $f: z \rightarrow \gamma$
- A parametric distribution $Pr(X ; \gamma)$

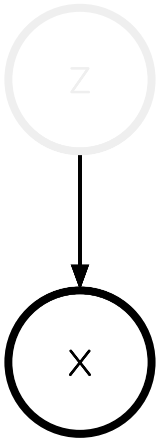
$Pr(X \mid Z = z) := Pr(X ; f(z))$

Probabilistic Graphical Models



Rui Shu. "Density Estimation: Variational Autoencoders." ruishu.io/2018/03/14/vae (2018).

Deep Probabilistic Graphical Models



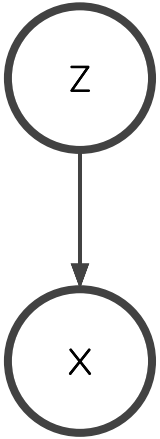
Linked Node denotes a conditional probability distribution
Pr(X | Z = z)

How to represent a conditional probability distribution:
- A mapping function f: Z → γ
- A parametric distribution Pr(X ; γ)

Pr(X | Z = z) := Pr(X ; f(z))

<u>f is a neural network</u>

Deep Probabilistic Graphical Models

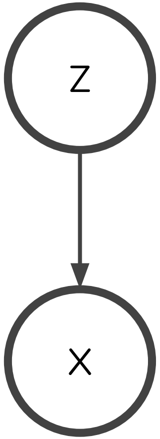


Node denotes an unconditional probability distribution
Pr(Z)

Linked Node denotes a conditional probability distribution
Pr(X | Z = z)

<u>f is a neural network</u>

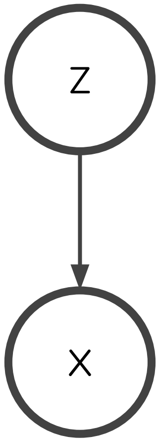
Training a Deep PGM



Given a dataset (D), and a deep PGM (P)

Minimize "distance" between D and P

Training a Deep PGM



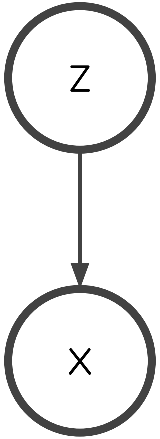
Given a dataset (D), and a deep PGM (P)

Minimize "distance" between D and P

Choices of distance:

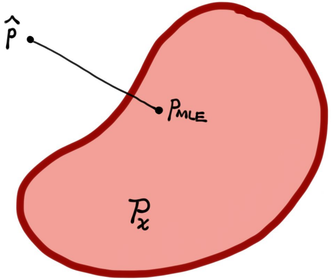
- Kullback-Leibler Divergence
- Jensen-Shannon Divergence
- Wasserstein Distance
- etc

Training a Deep PGM: Kullback-Leibler

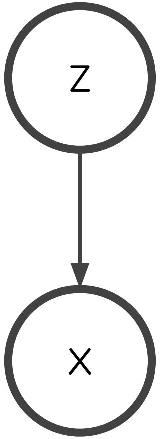


Given a dataset (D), and a deep PGM (P)

Minimize KL divergence between D and P



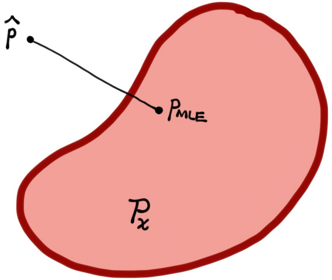
Training a Deep PGM: Kullback-Leibler



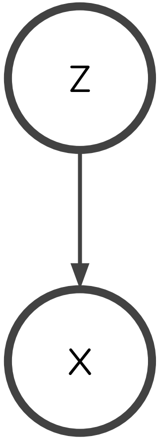
Given a dataset (D), and a deep PGM (P)

Minimize <u>KL divergence</u> between D and P <u>is equivalent to</u>

Maximize log-likelihood of ln P(x) where x is sampled from D



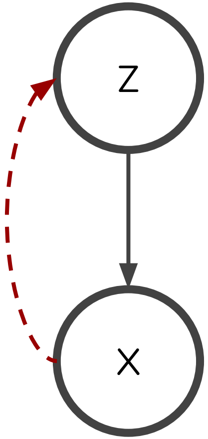
Training a Deep PGM: Kullback-Leibler



Computing ln P(x) is intractable

$$\ln p(x) = \ln \int p(x, z)\, \mathrm{d}z \geq \mathbb{E}_{q(z|x)} \ln \frac{p(x, z)}{q(z \mid x)}$$

Training a Deep PGM: Kullback-Leibler



q(z | x) is the variational approximation

Computing ln P(x) is intractable

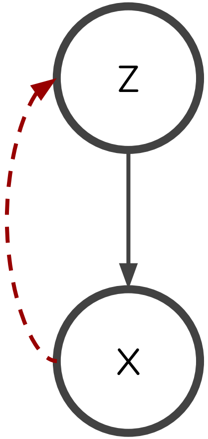
$$\ln p(x) = \ln \int p(x, z)\, \mathrm{d}z \geq \mathbb{E}_{q(z|x)} \ln \frac{p(x, z)}{q(z \mid x)}$$

Training a Deep PGM: Kullback-Leibler

Computing ln P(x) is intractable

$$\ln p(x) = \ln \int p(x, z)\, \mathrm{d}z \geq \mathbb{E}_{q(z|x)} \ln \frac{p(x, z)}{q(z \mid x)}$$

$$\underset{p}{\text{maximize}} \quad \ln p(x) \qquad \longrightarrow \qquad \underset{p,q}{\text{maximize}} \quad \mathbb{E}_{q(z|x)} \ln \frac{p(x, z)}{q(z \mid x)}$$



q(z | x) is the variational approximation

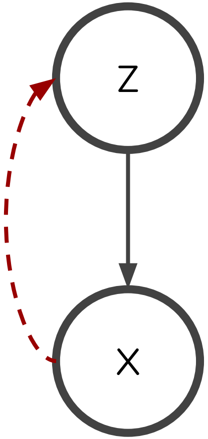
Evidence Lower Bound (ELBO)

$$\mathbb{E}_{q_\phi(z|x)} \ln \frac{p_\theta(x, z)}{q_\phi(z|x)}$$

$$\mathbb{E}_{q_\phi(z|x)} \ln p_\theta(x|z) - D_{KL}(q(z|x)||p(z))$$

$$\ln p_\theta(x) - D_{KL}(q(z|x)||p(z|x))$$

Training a Deep PGM: Kullback-Leibler



q(z | x) is the variational approximation

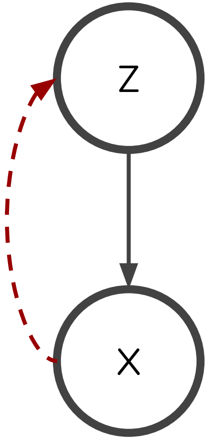
Computing ln P(x) is intractable

$$\ln p(x) = \ln \int p(x, z)\,\mathrm{d}z \geq \mathbb{E}_{q(z|x)} \ln \frac{p(x, z)}{q(z \mid x)}$$

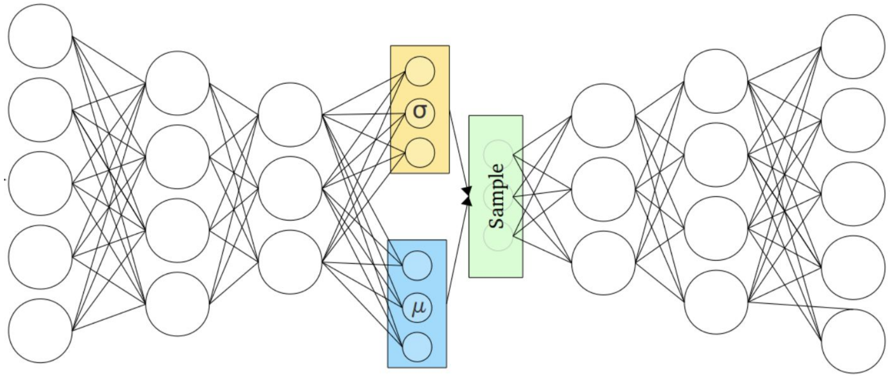
$$\underset{p}{\text{maximize}} \quad \ln p(x) \qquad \longrightarrow \qquad \underset{p,q}{\text{maximize}} \quad \mathbb{E}_{q(z|x)} \ln \frac{p(x, z)}{q(z \mid x)}$$

Variational Autoencoder

Training a Deep PGM: Kullback-Leibler



q(z | x) is the variational approximation



$$\underset{p}{\text{maximize}} \quad \ln p(x) \qquad \longrightarrow \qquad \underset{p,q}{\text{maximize}} \quad \mathbb{E}_{q(z|x)} \ln \frac{p(x,z)}{q(z \mid x)}$$

Variational Autoencoder

Irhum Shafkat. "Intuitively Understanding Variational Autoencoders." (2018).

Reparameterization Trick

$$\nabla_\phi \mathbb{E}_{q_\phi(x)} f(x, \phi)$$

Cannot push the gradient operator inside the expectation

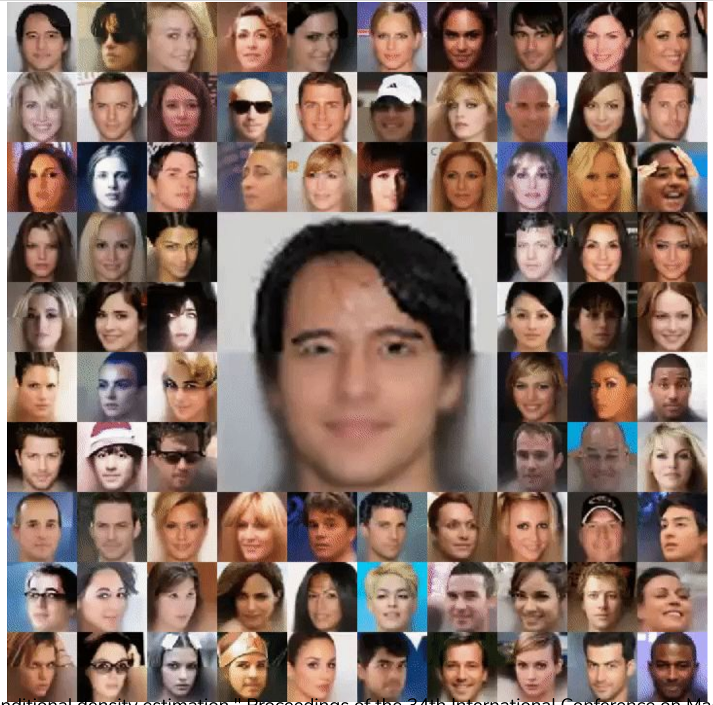
$$\epsilon \sim q_0, x = r(\epsilon, \phi)$$

Reparameterize the distribution

Rewrite the gradient, push gradient operator inside

$$\nabla_\phi \mathbb{E}_{\epsilon \sim q_0} f(r(\epsilon, \phi), \phi) = \mathbb{E}_{\epsilon \sim q_0} \nabla_\phi f(r(\epsilon, \phi), \phi)$$

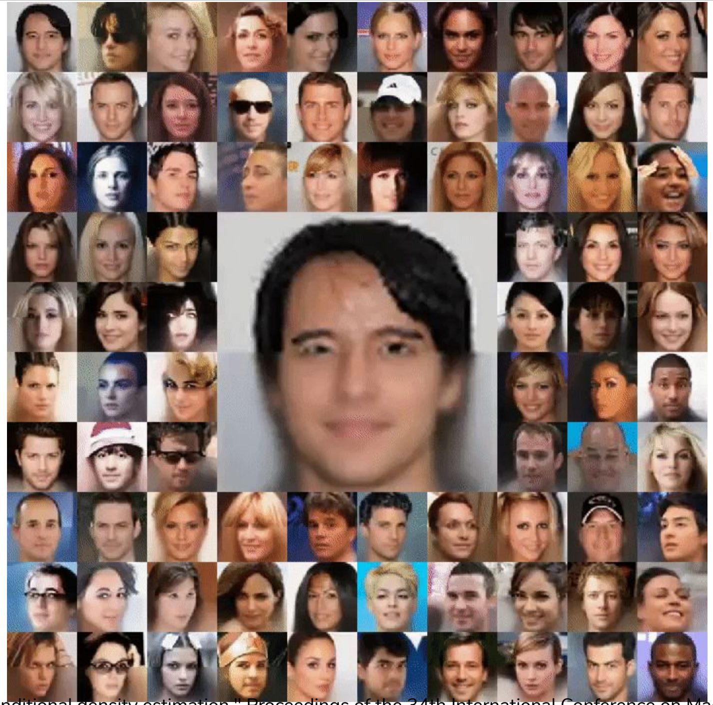
Toward Disentangled Face Representations



Shu, Rui, Hung H. Bui, and Mohammad Ghavamzadeh. "Bottleneck conditional density estimation." Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017.

Toward Disentangled Face Representations

Hung Bui
VinAI Research Director



Shu, Rui, Hung H. Bui, and Mohammad Ghavamzadeh. "Bottleneck conditional density estimation." Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017.

Style and Content Disentanglement

Model 1

Model 2

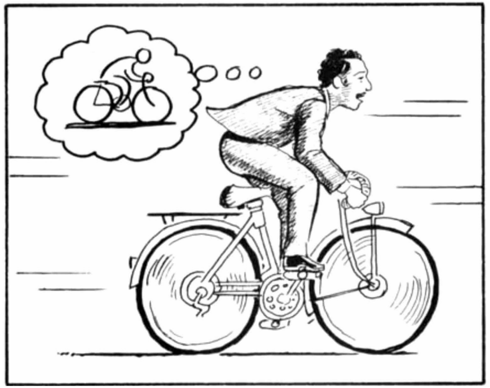
1. Why does our optimizer coincide with human preference?
2. How do we modify the loss function so that the loss function favors Model 1?

Shu, Rui, Shengjia Zhao, and Mykel J. Kochenderfer. "Rethinking Style and Content Disentanglement in Variational Autoencoders." (2018).

Planning in Representation Space

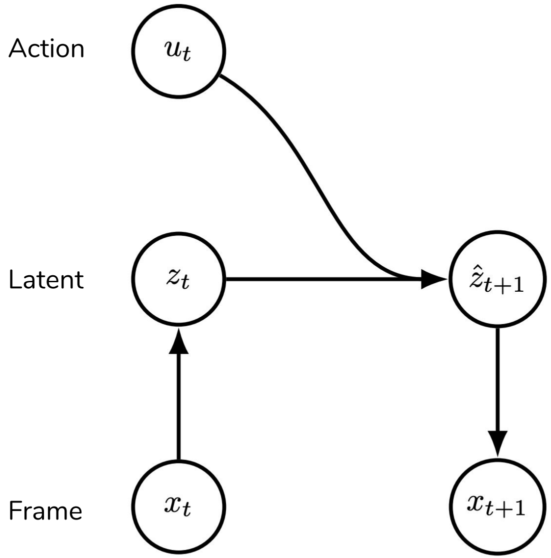
If you can **predict** what will happen if you perform an action, then you can **plan** ahead

- (Prediction) Use a generative model to do prediction in **<u>representation space</u>**
- (Planning) Apply optimal control algorithm in **<u>representation space</u>**



McCloud, Scott. Understanding Comics: The Invisible Art. Tundra Publishing, 1993. URL https://en.wikipedia.org/wiki/ Understanding_Comics.

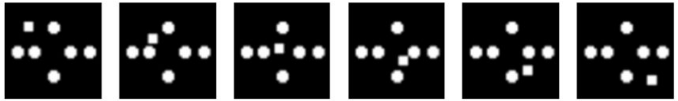
Planning in Representation Space



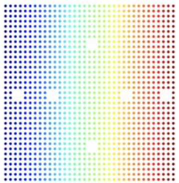
Predict next frame via a bottlenecked model

Nir Levine, Yinlam Chow, Rui Shu, Mohammad Ghavamzadeh, Hung Bui. “Prediction, Consistency, Curvature: Representation Learning for Locally-Linear Control.” (2019)

Planning in Representation Space

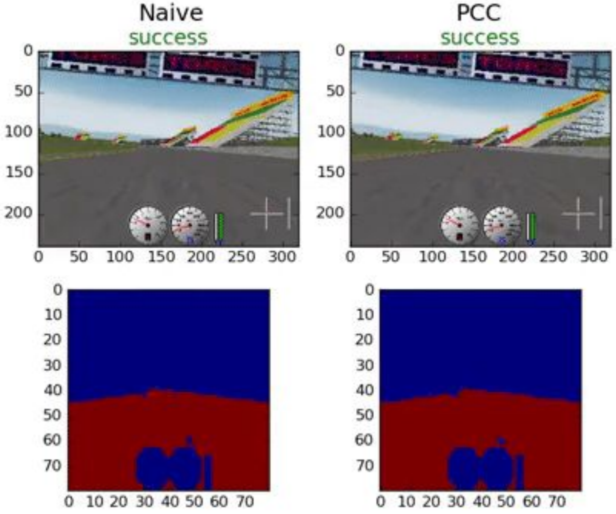


True map



Learned latent space

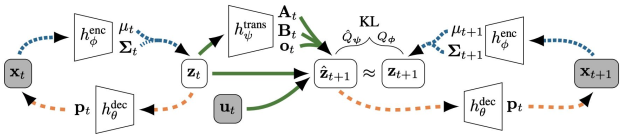
Planar System Environment

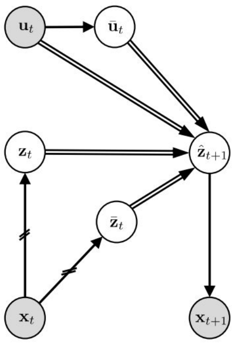


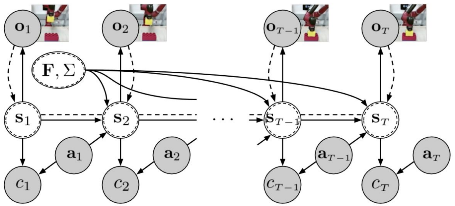
TORCS Environment

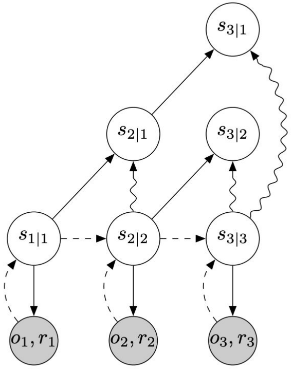
Nir Levine, Yinlam Chow, Rui Shu, Mohammad Ghavamzadeh, Hung Bui. “Prediction, Consistency, Curvature: Representation Learning for Locally-Linear Control.” (2019)

Planning in Representation Space









1. What is the best way to learn a representation space for control?
2. Should our choice of controller influence how we choose to learn the representation space?

Watter, Manuel, et al. "Embed to control: A locally linear latent dynamics model for control from raw images." Advances in neural information processing systems. (2015).
Banijamali, E., Shu, R., Ghavamzadeh, M., Bui, H., & Ghodsi, A. “Robust locally-linear controllable embedding.” arXiv preprint arXiv:1710.05373. (2017).
Zhang, Marvin, et al. "Solar: Deep structured latent representations for model-based reinforcement learning." arXiv preprint arXiv:1808.09105 (2018).
Hafner, Danijar, et al. "Learning latent dynamics for planning from pixels." arXiv preprint arXiv:1811.04551 (2018).

Recap

- Motivation
  - Generative modeling
  - Representation learning
- Deep Probabilistic Graphical Models
  - A neural network mapping function
- Training Deep PGMs
  - Variational Inference
- Representation Learning with Deep PGMs
  - Disentangled Representations
  - Planning in Representation Space