Unsupervised Representation Learning

DeepMind Lecture

① Why is it important?

Unsupervised learning provides the ability to, without human input, arrange data into compelling representations. We can arrange such representations based on orthogonality, independence, etc.

② How do we evaluate results?

Feature engineering is a longstanding difficulty when it comes to modeling the world. Restricted Boltzman machines were an early variant.

(3) Unsupervised representation learning as providing robust representations of a scene. Harder to fool.

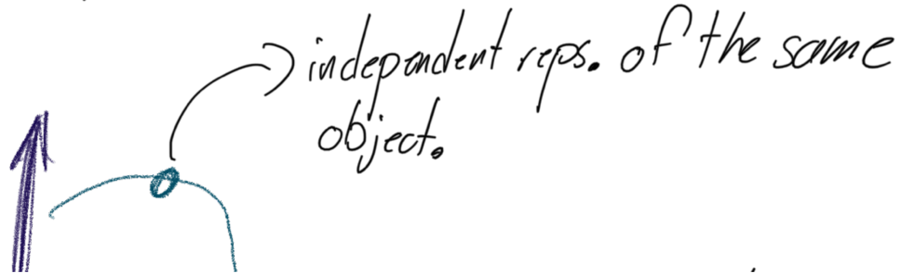
(4) Representations as increasing generalization of model. Lacking "common sense."

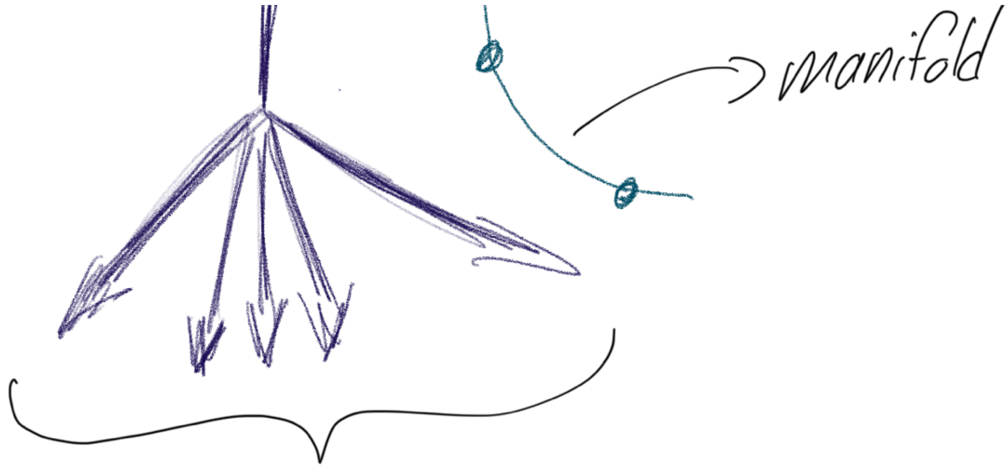
(5) Geoff Hinton, Yann LeCun, Yoshua Bengio all believe robustness will be found in unsupervised representation.

(6) Representational form orthogonal to information content. Makes some computations more efficient.

(7) We seek to identify the shape of the manifold which contains the data within representational space.

(8)





Representation Space

(9) As time progresses, these visual manifolds separate and become less entangled.

(10) Visual processing as reformatting.

(11) We must contextualize representations wrt a task of some kind. Some objective goal.

(12) A state must capture all pertinent data.

(13) Solving tasks requires

(1) Attention

(2) Clustering

3. Latent States -> Occluded Data

14. Compositionality is another key aspect.
"meaning of its constituent expressions and the rules used to combine them"

<u>Compositionality leads to open-endedness.</u>

15. Good Representations Are:

- Untangled
- Attention
- Clustering
- Latent Information
- Compositionality

16. Inspirations also Arise in Physics

We utilize the fundamental notion of symmetry

For our purposes, symmetry represents those stubborn cores that remain unaltered even under transformations that could change them" - Mario Livio, 2012

(17) More formally we seek commutativity

(18) Neural networks use information bottlenecks. (Data processing inequality)

(19) We seek to learn invariant representations, along with equivariant representations.

(20) We ideally seek disentangled representations to capture equivariant representations.

(21) This is similar to untangling in neuroscien

(22) Can be used to capture properties of Symmetry groups.

(23) Evaluating the merit of a representatio

-> Symmetrics -> Effectiveness on mapping groups (equivariant map)

-> Compositionality as tied to original source material

-> Attention -> May be binary mask

-> Clustering -> We assume our representation lives in the vector space and select a helpful metric

If these are cleared, we say we have a good representation.

24 We say such a representation is

good because it helps with:

Data Efficiency

Generalisation

Robustness

Transfer

Common Sense

These are not speculative associations

24) Attenuation is key for generality.

25) Transfer is facilitated by ignoring individual observations.

26) Side Bar: A game theoretic

approach to achieving "good" representati

(27) Big Idea: Is all machine learning representation learning?

(28) Representation Learning Techniques

3 Methods:

Generative Modeling

Contrastive Loss

Self-Supervision

↳ Exploit knowledge of data to design

learning' tasks which lead to useful representations.

↳ Utility signal contained in such a task?

(29) Semi supervised learning: Use learned representations for classification. → Using downstream tasks

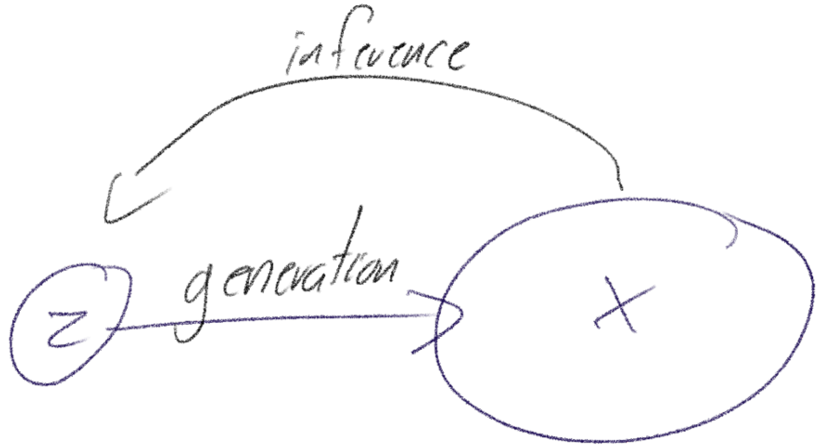
(30) Model Analysis for the sake of interpretable models.

(31) It is essential to keep in mind consistency and temporal abstraction. Similar entities should be represented in similar ways.

(32) Generative Modeling: Model the underlying data distribution

(33) Latent Variable Models:

Model the data generating process as a mapping from low dimensional unknown (latent space to the data distribution.



34 We seek $p(z|x)$

We wtf the underlying factors which generated the data (<u>with uncertainty estimates</u>).

Finding $p(z|x)$ is often intractable, and we have to resort to approximations.

35 We want to learn inference & generation at the same time.

(36) We accomplish this using max liklihood:

given an expectation

$$E_{p^*(x)}[\log p_\theta(x)]$$

We expect high liklihoo

Latent Variable Model

$$\log p_\theta(x) = \log \int p_\theta(x|z)\, p(z)\, dz$$

params

We want the model that is able to explain our data.

$$\log p_\theta(x) \geq \overbrace{E_{q_\eta(z|x)} \log p_\theta(x|z)}^{\text{reconstruct}} - \underbrace{KL(q_\eta(z|x) \,\|\, p(z))}_{\text{stay close to}}$$

prior

Approximate posterior:

$q_\eta(z|x)$

37

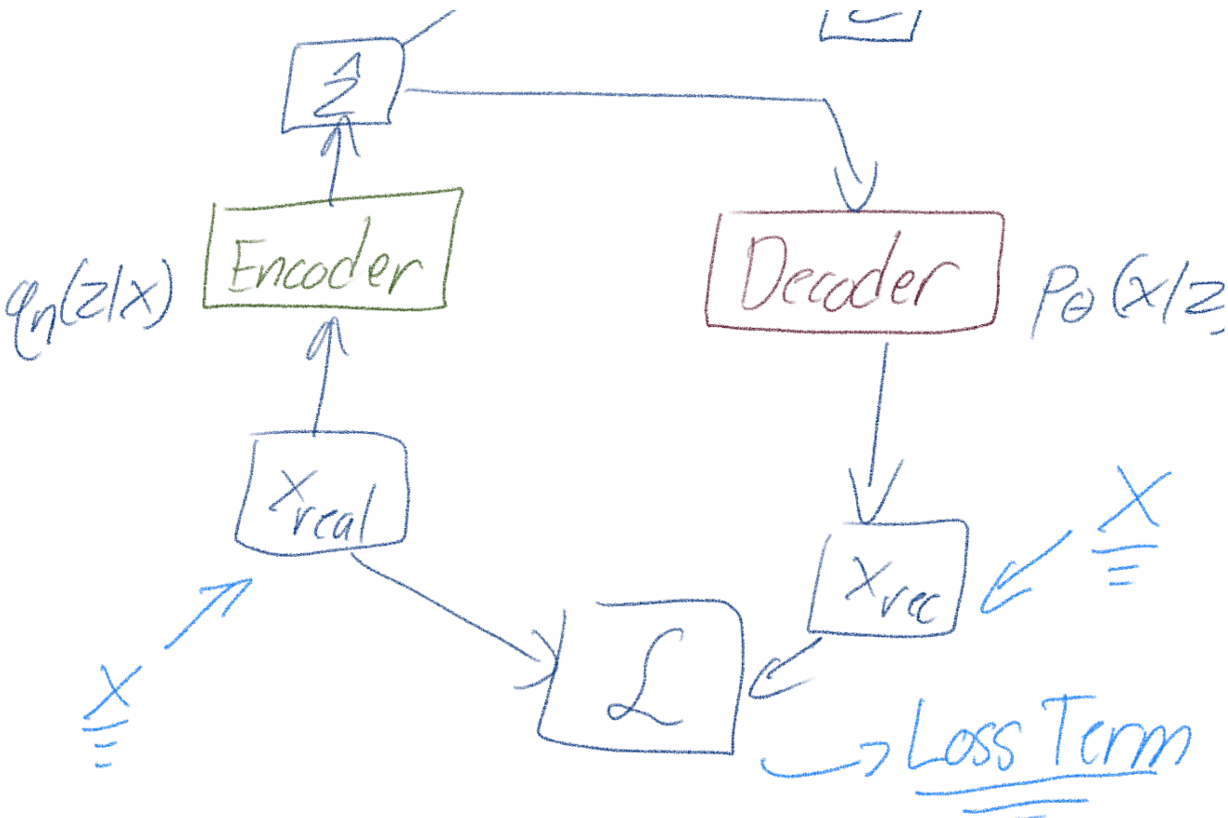
Both the inference and generation models are deep neural networks.

The KL term regularises the approxim posterior to the prior.

$$KL[q_\eta(z|x)] \| [p(z)]$$

38

$\boxed{KL}$



(39) What does the selection of z really look like? What are we regularizing against?

(40) We may change the weight of the KL term to encourage disentangled representations.

(41) beta-VAE

Learns disentangled continuous representation
encoding semantic information.

(42) Disentanglement reveals encodings as ascribed to particular objects.

(43) beta-VAE: Learns disentangled continuous encoding semantic information.

(44) Integrating beta-VAEs into reinforcem learning agents improves generalization and transf

(45) Sequential VAEs - ConvDraw
-> Recurrent Component
-> Recurrence helps: iteratively refine and add details
-> Latents: Spatial and Temporal

(46) Monet
-> [illegible]

-> using attention in a multi level process leads to a generative model which learns concepts (objects) unsupervis

-> Latent traversals show that Monet learns to encode the position of an object into a single latent.

(47) Generative Query Networks (GQN)

-> Use learned representation to condition a recurrent generative model to generate how the scene looks from a different view.

(48) Vector Quantized VAEs

Learning Discrete latent variables is challenging. So embed them into continuous embedding space.

Results in high compression rate

Results in high compression rates.

(49) Game inverted with BigBiGan and we arrive at embeddings devoid of reconstruction loss.

(50) Latent variables in such models encode very high level structures.

(51) GPT learns robust representations which may be used for many downstream tasks.

(52) Contrastive Learning (An easier alternative)

-> Uses classification loss

-> Model trains on positive and negative examples

(53) Contrastive Predictive Coding

-> Maximize mutual information between data and learned representations

Side Bar: Mutual information as loss is

tolerated.

-> Uses supervised learning to model density ratios.

54 SimCLR

-> Use contrastive losses to maximize mutual information between representations of data under differant transformations.

55 Unsupervised Visual Representation Learning By Context Prediction

-> Predicts Selected Patches

-> Unsupervised Object Discovery