

Deep Generative Models and Differentiable Inference

Sargur N. Srihari
srihari@cedar.buffalo.edu

Topics in Deep Generative Models

1. What are Generative Models?
2. Types of Generative Models
3. Boltzmann machines
4. Restricted Boltzmann Machines
5. Deep Belief Networks
6. Deep Boltzmann Machines
7. Sigmoid Belief Networks
8. Differentiable Generator Networks
9. Variational Autoencoders
10. Generative Adversarial Networks

What are generative models?

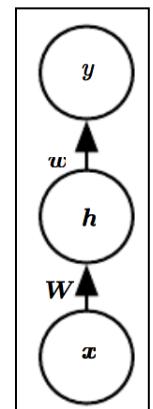
- *Generative modeling* refers to building a model of data, $p(\mathbf{x})$, that we can sample from
 - E.g., \mathbf{x} is an image
- *Discriminative Modeling*, such as regression and classification, tries to estimate conditional distributions such as $p(\text{class}|\mathbf{x})$

Why generative models?

- Even for prediction, generative models useful:
 1. Data efficiency and semi-supervised learning
 2. Model checking by sampling
 3. Understanding
- Each of these reasons are discussed next

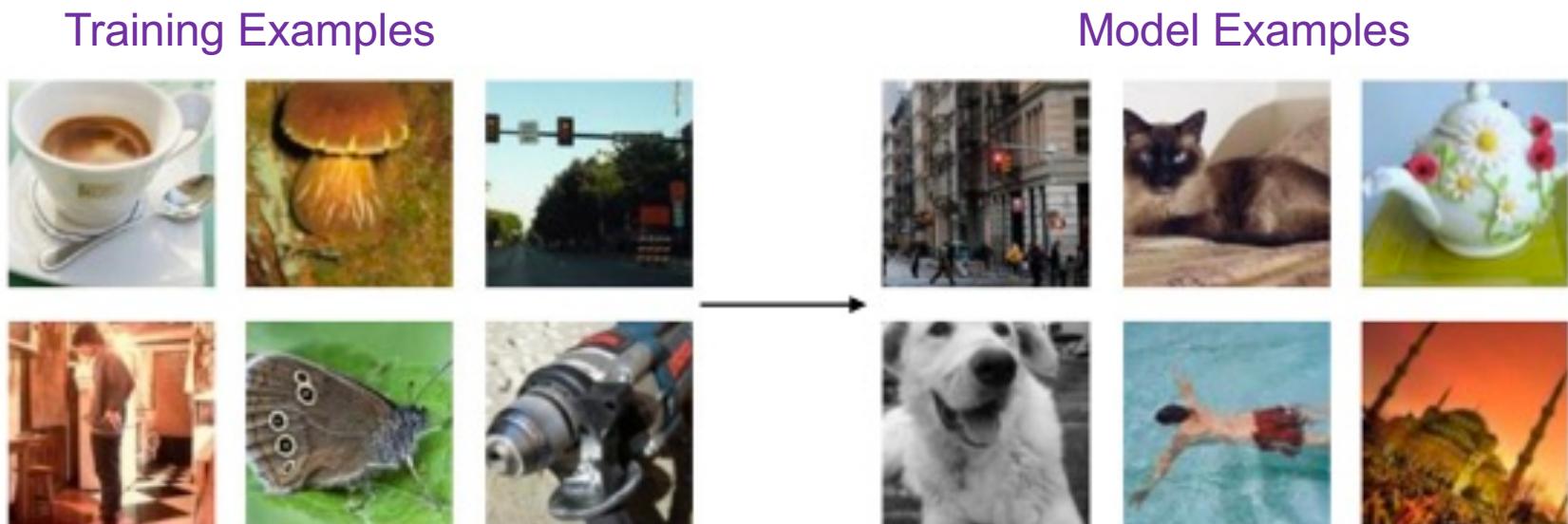
1. Data efficiency

- Generative models can reduce data required
 - Image classifier $p(\text{class}|\mathbf{x})$ has high-dimensionality requiring a lot of data
- If data is modeled as being generated from low-dimensional latent variables \mathbf{h} , as in
 - $p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{h}) p(\mathbf{h}) d\mathbf{h}$
 - Then we need only learn $p(\text{class}|\mathbf{h})$
- Allows taking advantage of unlabeled data
 - Called *semi-supervised* learning



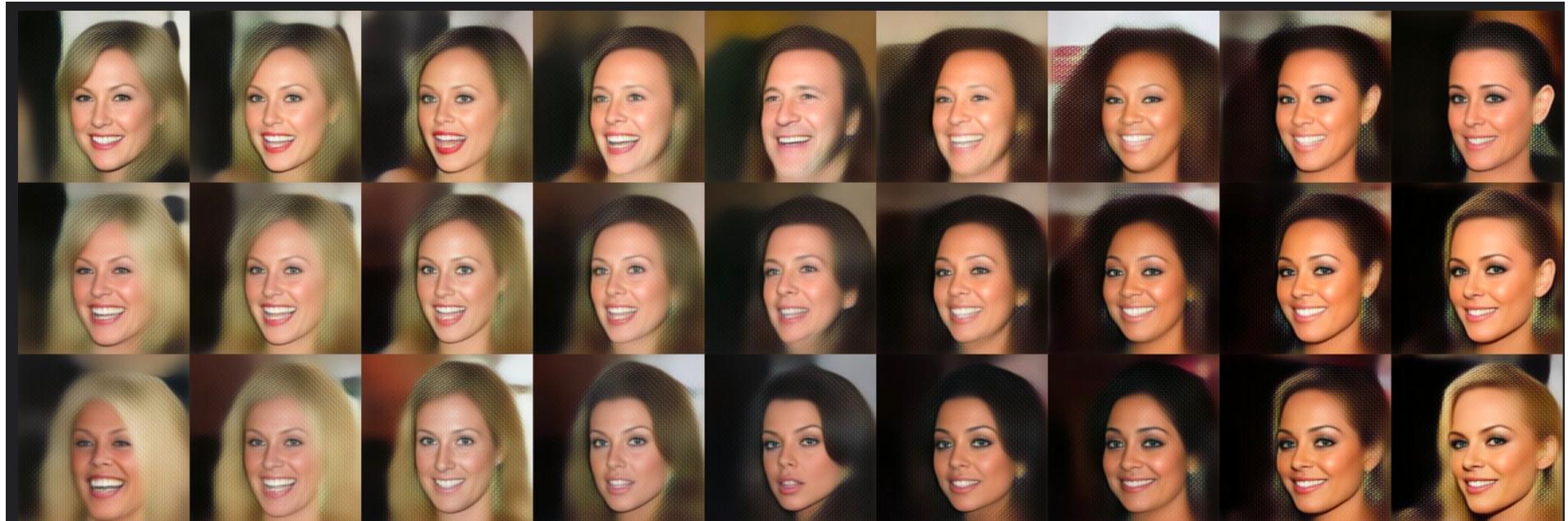
2. Model checking by sampling

- Understanding complex regression and classification models is hard
 - Unclear as to what model learns and what is missed
- Compare data sampled from model to real data



3. Understanding

- Generative models assume data point generated from low-dimensional latent variable
- Latent variables are often interpretable
 - May tell us about hidden causes of a phenomenon
 - Allow interpolating between samples



Differentiable Inference

- Generative models are hard to apply to real data sets
 - Inference such as MCMC not scalable on data sets
- Recent generative models:
 1. VAEs: approximate inference and output $q(z|x)$
 2. GANs: optimize generative model by fooling classifier
 3. Invertible density estimation: transform latent distribution by a series of invertible functions
 4. Autoregressive models: $p(x) = p(x_1)p(x_2|x_1)p(x_3|x_2, x_1)\dots$
RNNs use this approach do not support all queries
- Common thread:
 - Can scale to high dimensions because loss functions are end-to-end differentiable

Types of Deep Generative Models

- There are several deep generative models
- They are built and trained using techniques of :
 - Probabilistic graphical models
 - Monte Carlo methods,
 - Partition Functions,
 - Approximate Inference

Types of Deep Generative Models

- All represent probability distributions
 - Some allow distribution to be evaluated explicitly
 - Others do not allow distribution to be evaluated but allow sampling
 - Some described by graphs and factors, others not