



Autoencoders

Machine Learning: Jordan Boyd-Graber
University of Maryland

SLIDES ADAPTED FROM IAN GOODFELLOW

Theme for Today

- Representation
- Generation
- Training

Theme for Today

- Representation
- Generation $p(x, y)$
- Training

Ties in to the generative models and inference we've been talking about!

Theme for Today

- Representation
- Generation
- Training

But with weaker probabilistic assumptions . . .

Theme for Today

- Representation
- Generation
- Training

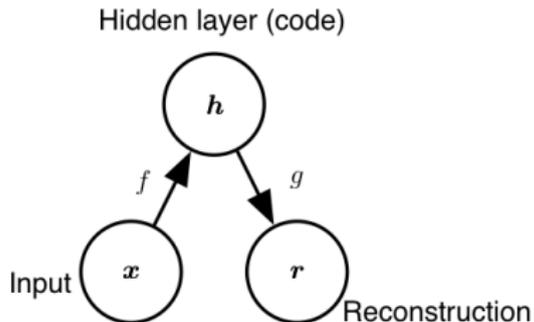
But with weaker probabilistic assumptions . . . okay if you have enough data.

Why autoencoders

- Discover hidden structure
 - Unlike clustering or admixtures, continuous
 - Not always interpretable
- Reconstruct data
- Features for downstream model (a la word2vec)

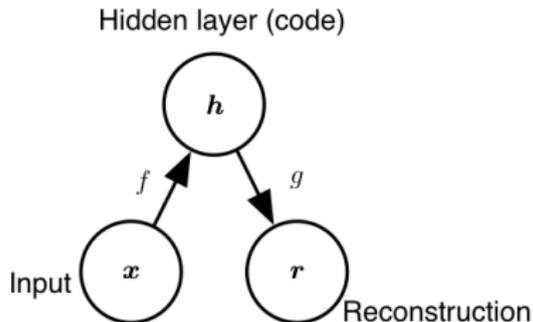
Setup: Encoders and Decoders

- Minimize $L(\vec{x}, g(f(\vec{x})))$
 - Encoder f
 - Decoder g
- Problem: avoid identity!
- Via hidden layer \vec{h}
 - Regularized
 - Smaller dimensionality
 - Probabilistic



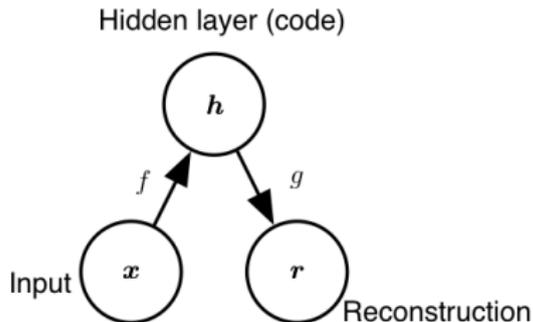
Setup: Encoders and Decoders

- Minimize $L(\vec{x}, g(f(\vec{x})))$
 - Encoder f
 - Decoder g
- Problem: avoid identity!
- Via hidden layer \vec{h}
 - Regularized
 - Smaller dimensionality
 - Probabilistic



Setup: Encoders and Decoders

- Minimize $L(\vec{x}, g(f(\vec{x})))$
 - Encoder f
 - Decoder g
- Problem: avoid identity!
- Via hidden layer \vec{h}
 - Regularized
 - Smaller dimensionality
 - Probabilistic

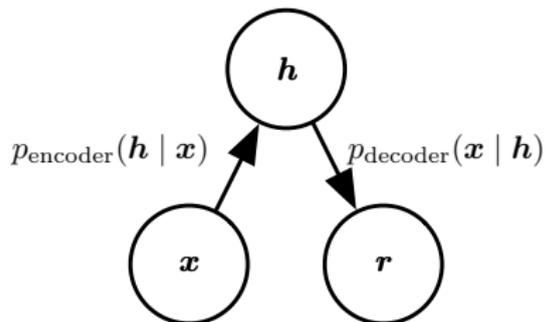


Setup: Encoders and Decoders

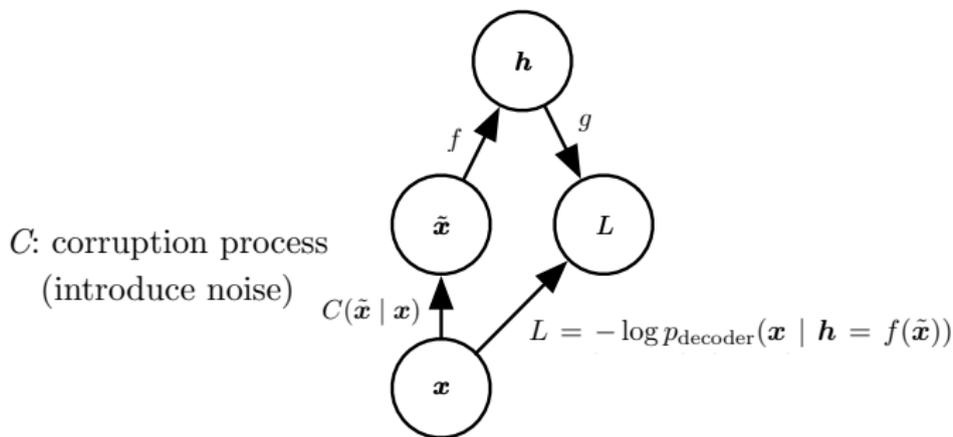
- Minimize $L(\vec{x}, g(f(\vec{x})))$
 - Encoder f
 - Decoder g
- Problem: avoid identity!
- Via hidden layer \vec{h}
 - Regularized
 - Smaller dimensionality
 - Probabilistic

Setup: Encoders and Decoders

- Minimize $L(\vec{x}, g(f(\vec{x})))$
 - Encoder f
 - Decoder g
- Problem: avoid identity!
- Via hidden layer \vec{h}
 - Regularized
 - Smaller dimensionality
 - Probabilistic

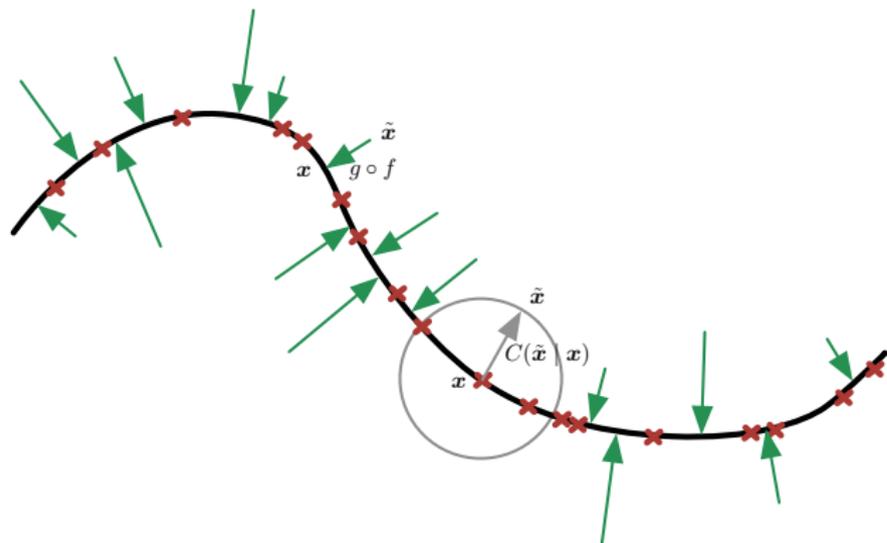


Denoising Autoencoders



- Perturb the input somehow
- Try to correct the noise: high probability of $\tilde{\mathbf{x}}$!
- Learns manifold

Denoising Autoencoders



- Perturb the input somehow
- Try to correct the noise: high probability of \tilde{x} !
- Learns manifold