Autoencoder Architectures

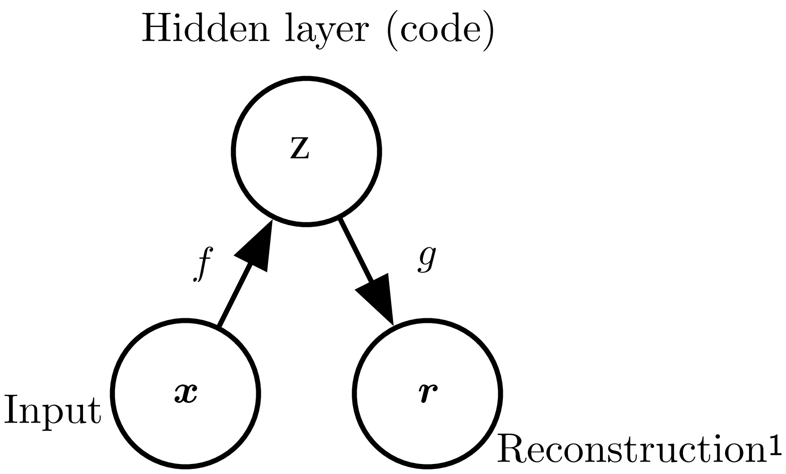
Kate Farrahi

ECS Southampton

March 25, 2019

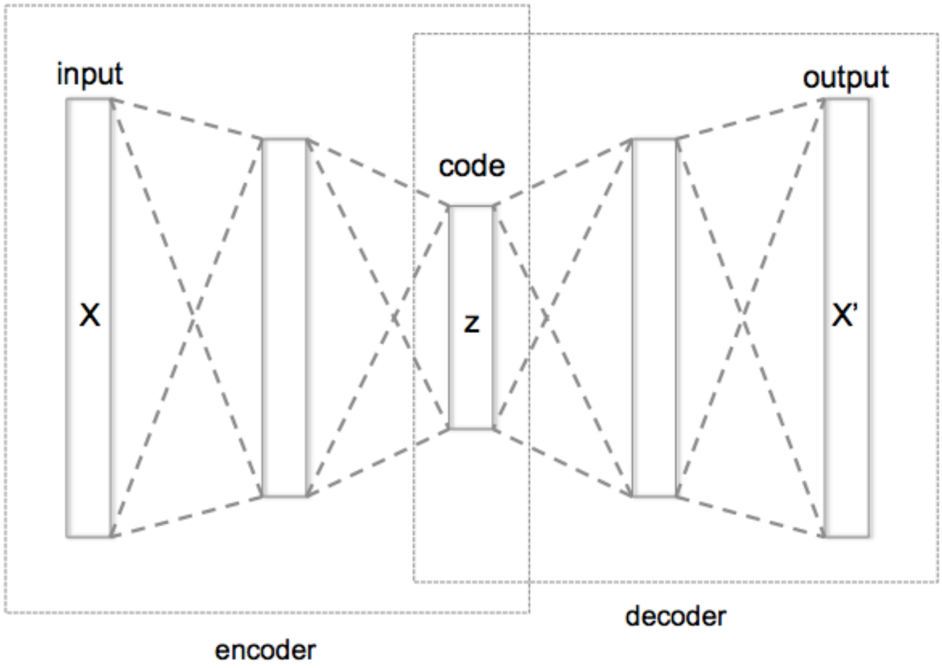
Autoencoder

The aim of an autoencoder is to learn a representation (encoding) for a set of data, typically for dimensionality reduction, by training the network to ignore signal noise.



[1] Image taken from Deep Learning book by Goodfellow et al.

Deep Autoencoder



[2]

[2] Image taken from wikipedia

Denoising Autoencoder

- Denoising autoencoders take a partially corrupted input and train to recover the original undistorted input.
- To train an autoencoder to denoise data, it is necessary to perform a preliminary stochastic mapping to corrupt the data ($x \rightarrow \tilde{x}$).
- A normal autoencoder is used with $\tilde{x}$ is used as input and x as output.
- In a denoising autoencoder, the loss should be computed on $\mathcal{L}(x, \hat{x})$ as opposed to $\mathcal{L}(\tilde{x}, \hat{x})$.

Sparse Autoencoder

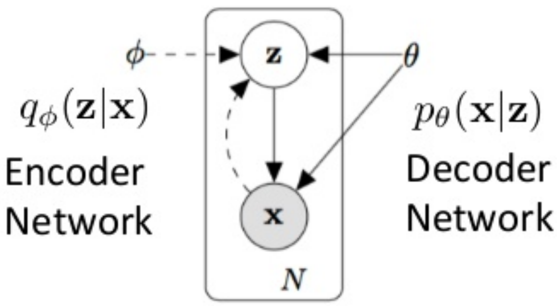
- In a sparse autoencoder, there are more hidden units than inputs, but only a small number of the hidden units are allowed to be active at the same time.

Autoencoder Applications

- Any basic AE (or its variant) is used to learn a compact representation of data.
- You can learn automatic features from data. E.g. HAR with sensor data.
- Denoising can help generalise over the test set since the data is distorted by adding noise.
- Pretraining networks by learning your network weights using a stacked AE.

Variational Autoencoders (VAEs)

Variational Autoencoder



Minimize: $D_{KL}[q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z}|\mathbf{x})]$

Intractable: $p_\theta(\mathbf{z}|\mathbf{x}) = \dfrac{p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})}{p_\theta(\mathbf{x})}$

[3]

[3] Slide taken from https://www.slideshare.net/ckmarkohchang/variational-autoencoder

Variational Autoencoder

- In VAEs, data is generated by a directed graphical model $p(x|z)$.
- The encoder is learning an approximation $q_\phi(z|x)$ to the posterior distribution $p_\theta(x|z)$, where ϕ and θ denote the parameters of the encoder and decoder, respectively.
- The objective of the VAE has the following form $\mathcal{L} = D_{KL}(q_\phi(z|x)||p_\theta(z)) - \mathbb{E}_{q_\phi}(log p_\theta(x|z))$