

COMS 4030A/7047A

Adaptive Computation and Machine Learning

Hima Vadapalli

Semester I, 2022

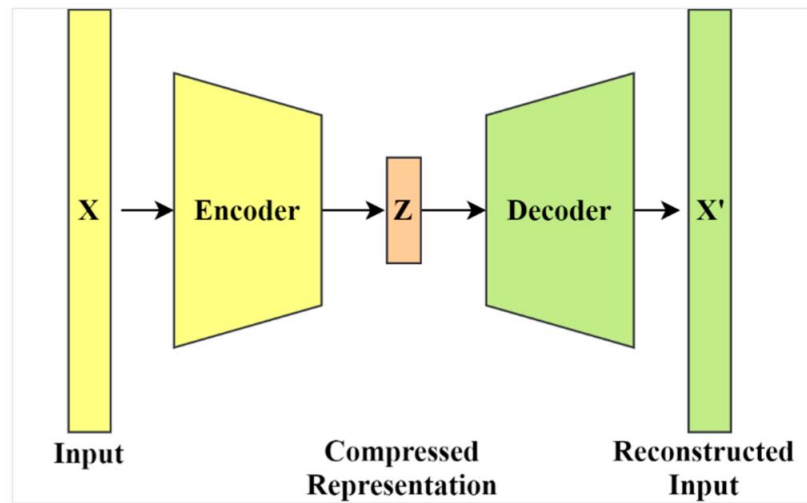
*So far:
CNNs, RNNs, LSTMs (Supervised
Setting)*

*Today:
Autoencoders (Unsupervised Setting)*

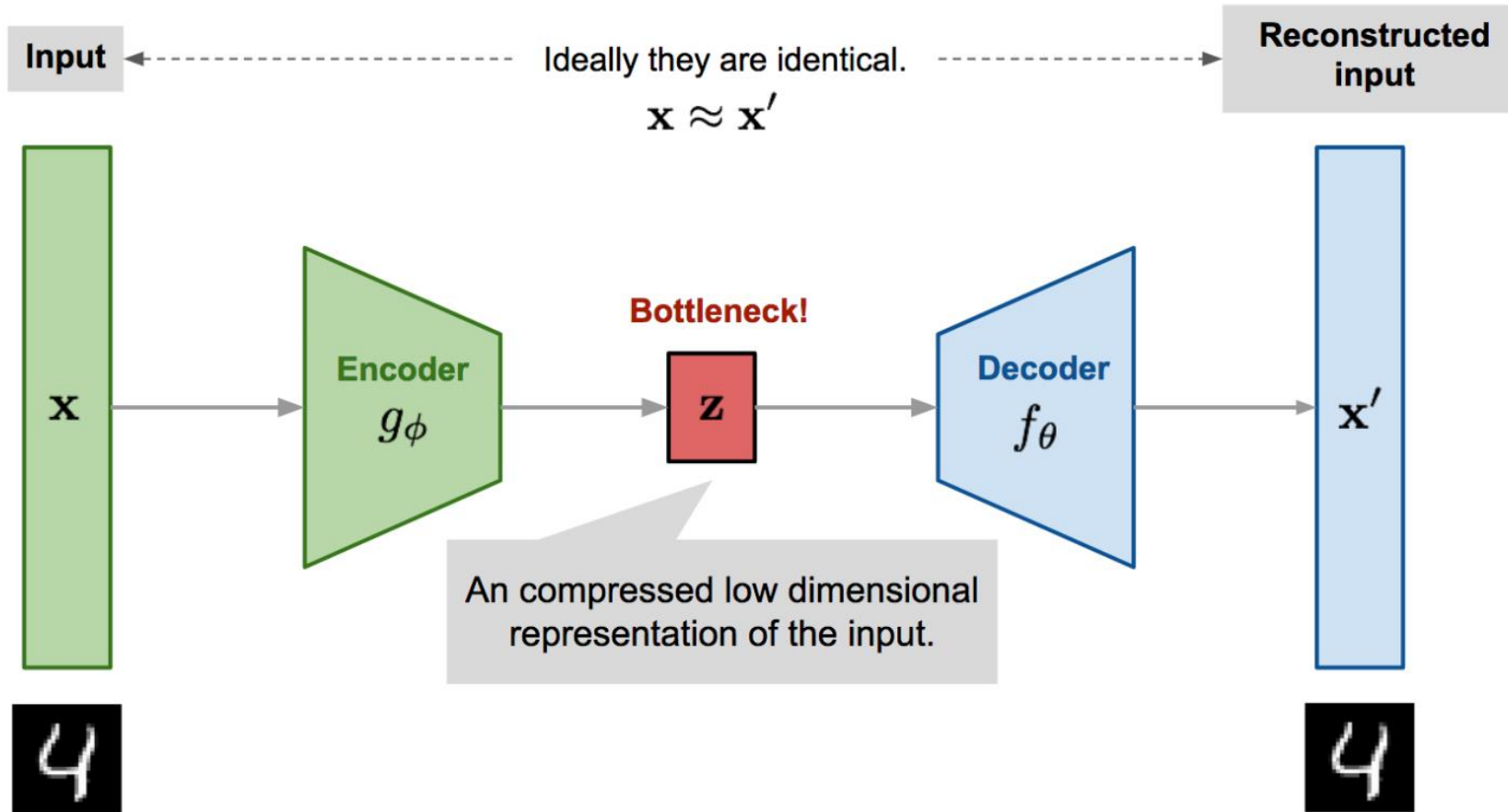
Autoencoders

Special type of feedforward neural networks where input is the same as output.

- compress input into a lower-dimensional *code* (latent space representation)



Autoencoders

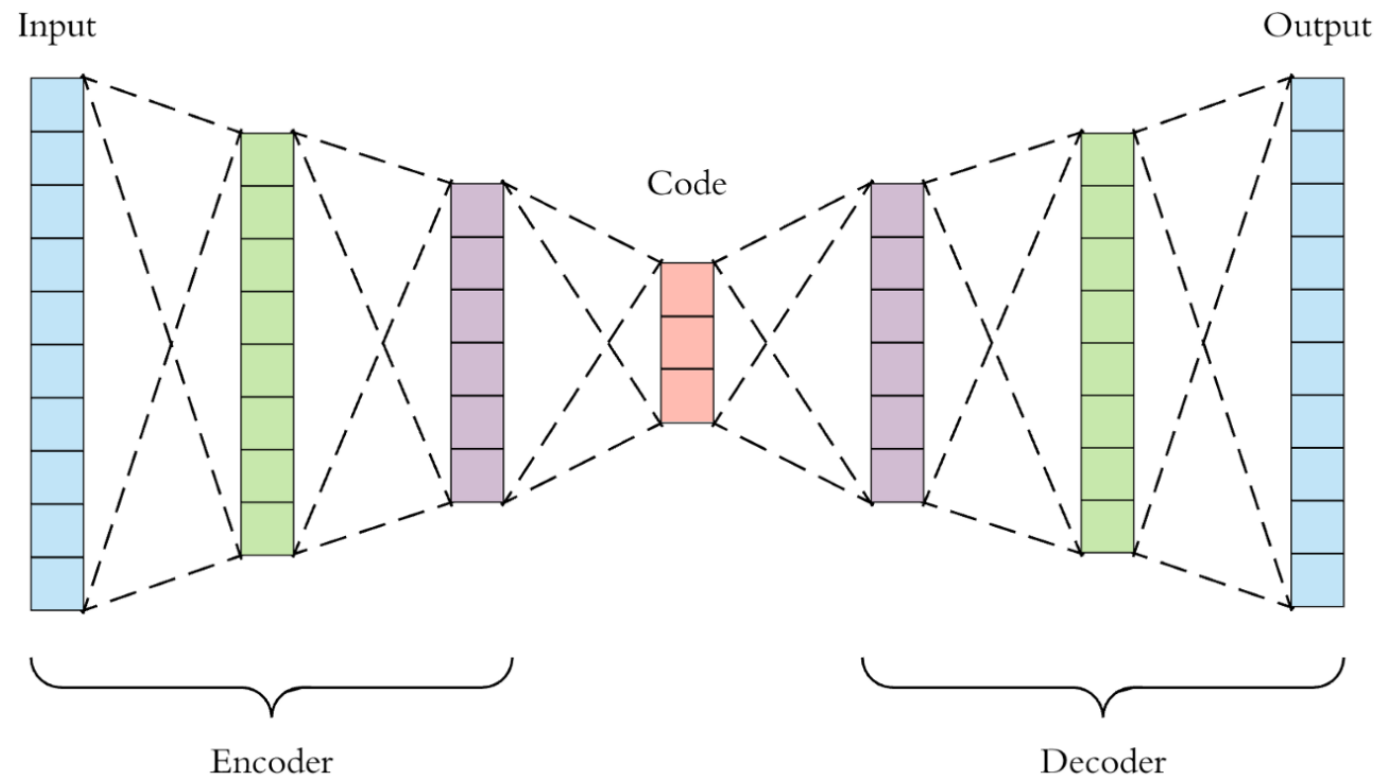


$$L_{\text{AE}}(\theta, \phi) = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}^{(i)} - f_\theta(g_\phi(\mathbf{x}^{(i)})))^2$$

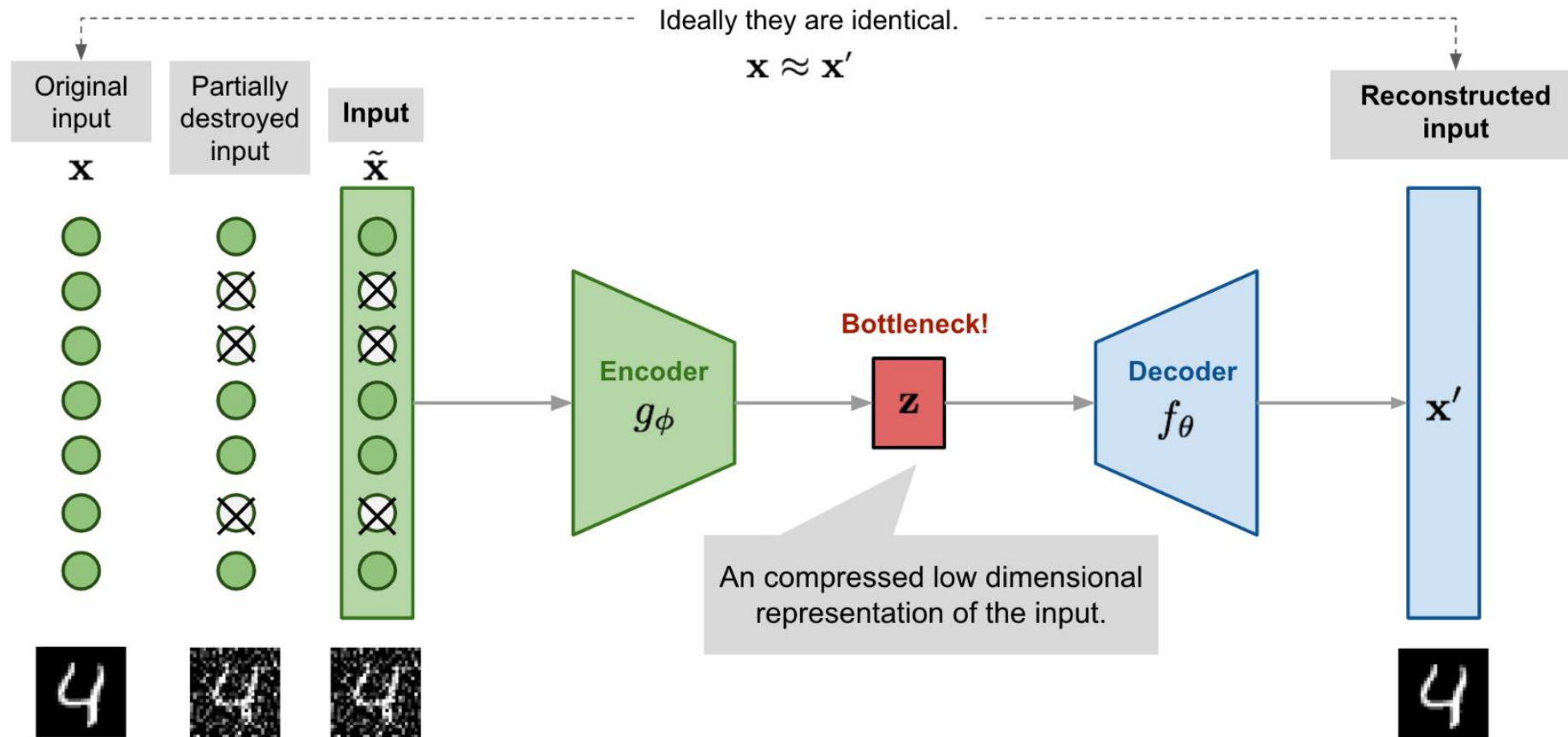
Autoencoders : Properties

- Data specific
 - Can only meaningfully compress data that was seen before
- Lossy
 - Output not exactly the same as input
- Unsupervised
 - Don't need explicit labels
 - Using training input data as labels (self supervised)

Deep Autoencoders



Denoising autoencoder



$$L_{\text{DAE}}(\theta, \phi) = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}^{(i)} - f_\theta(g_\phi(\tilde{\mathbf{x}}^{(i)})))^2$$

Variational autoencoder

