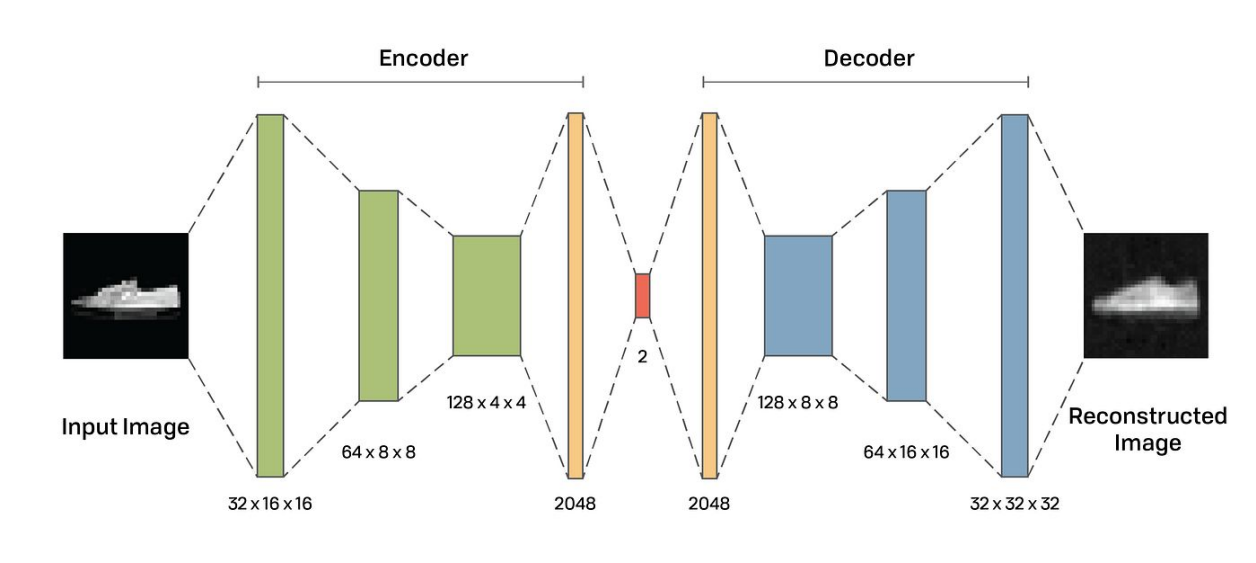
Autoencoder

Walter Fenske, Markos Meitarjyan, Anh Phan



We have learned some different technique for unsupervised learning before (NMF, PCA, etc.)

But do you know

There is a method that might be even better than these methods

That uses neural networks?

That’s autoencoder

What is an autoencoder?

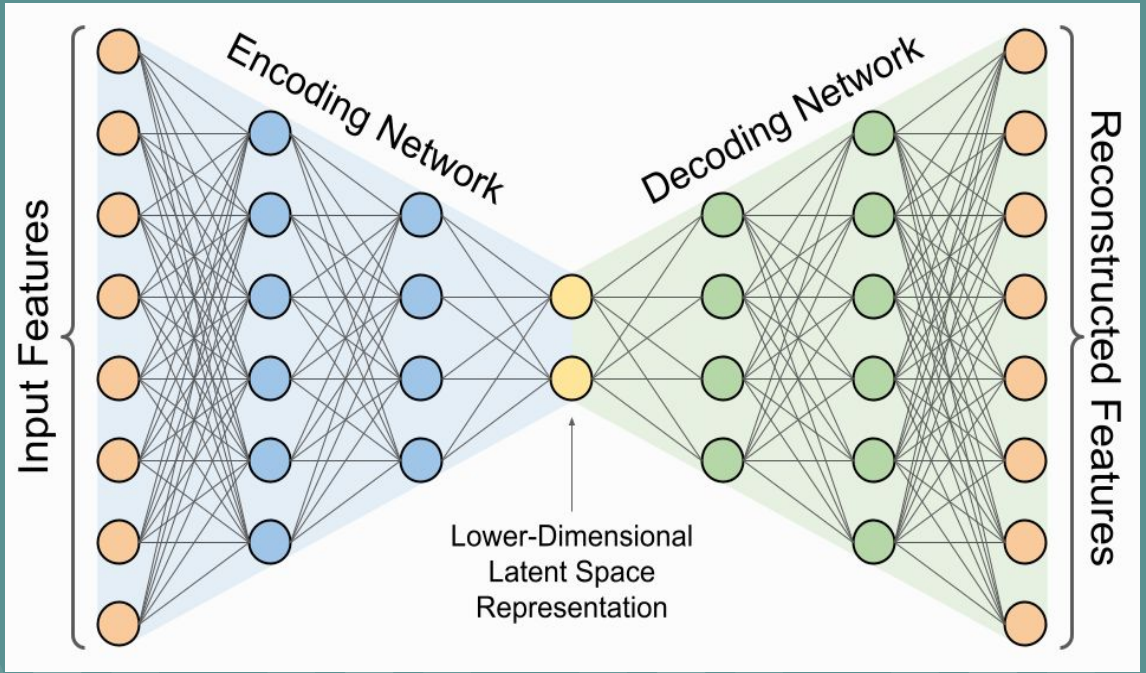
- Autoencoder = Feed-forward neural network
- that learn to **recreate input from input**?
- But why do we want to copy the input to output?

  => Autoencoder do learn to recreate input from input, but it will not learn how to do it perfectly

  => Only copy approximately, and only copy input that resembles the training data

  => Learn the most important features of our data! (like dimension reduction)

Architecture of autoencoder

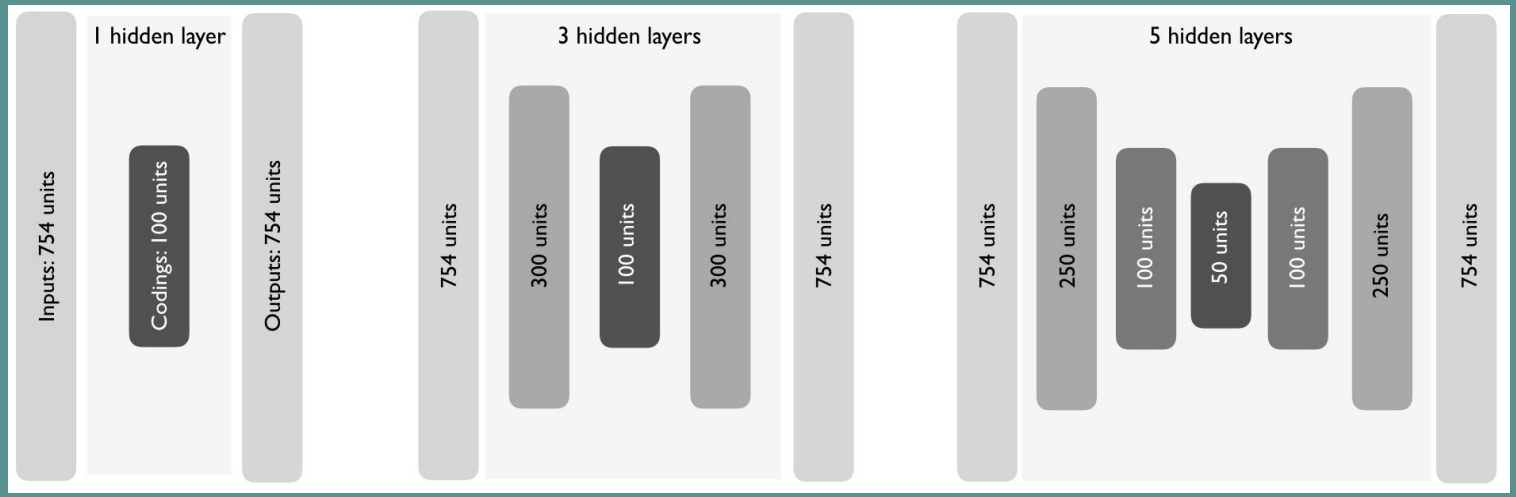


- Autoencoder = Encoder + Decoder
  - + Both are feed-forward NN
- Encoder: Input to Code
  - + Lower-dimension representation of input => “Undercomplete” AE
- Decoder: Code to (approximately) Input

=> The code will represent the most important features

=> Better at learning non-linear features comparing to PCA

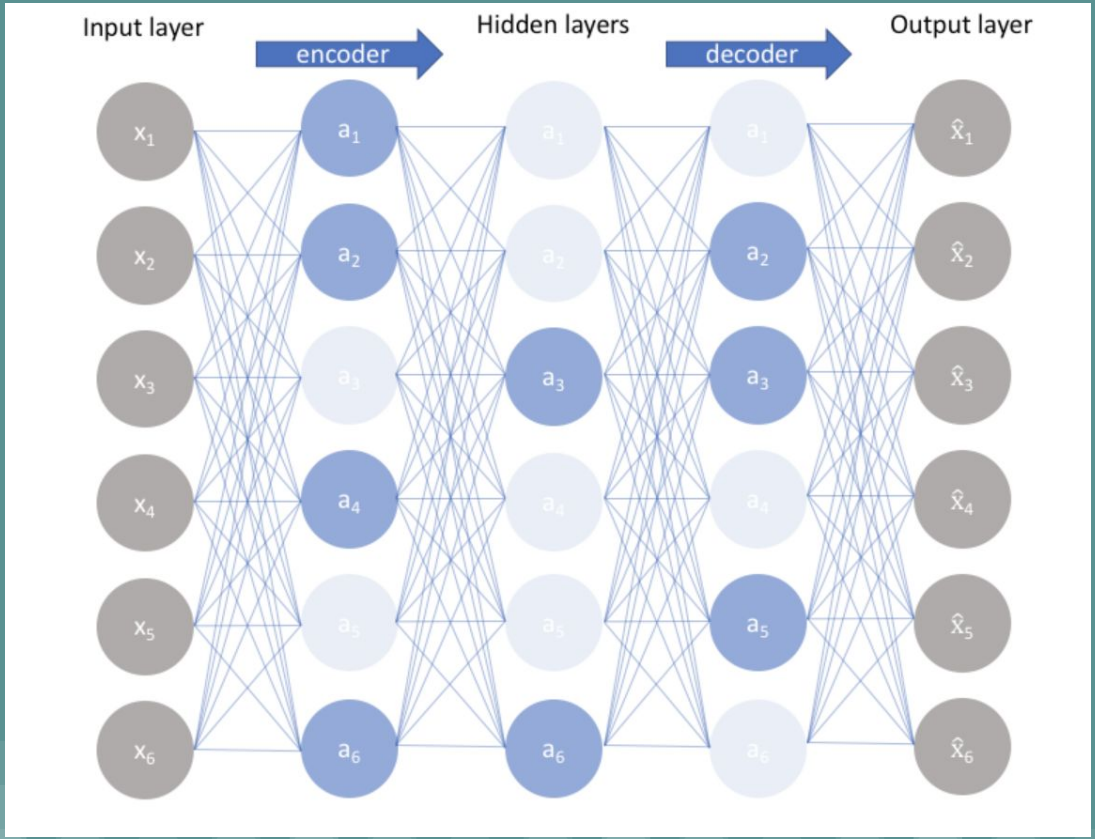
(More) Architecture of autoencoder

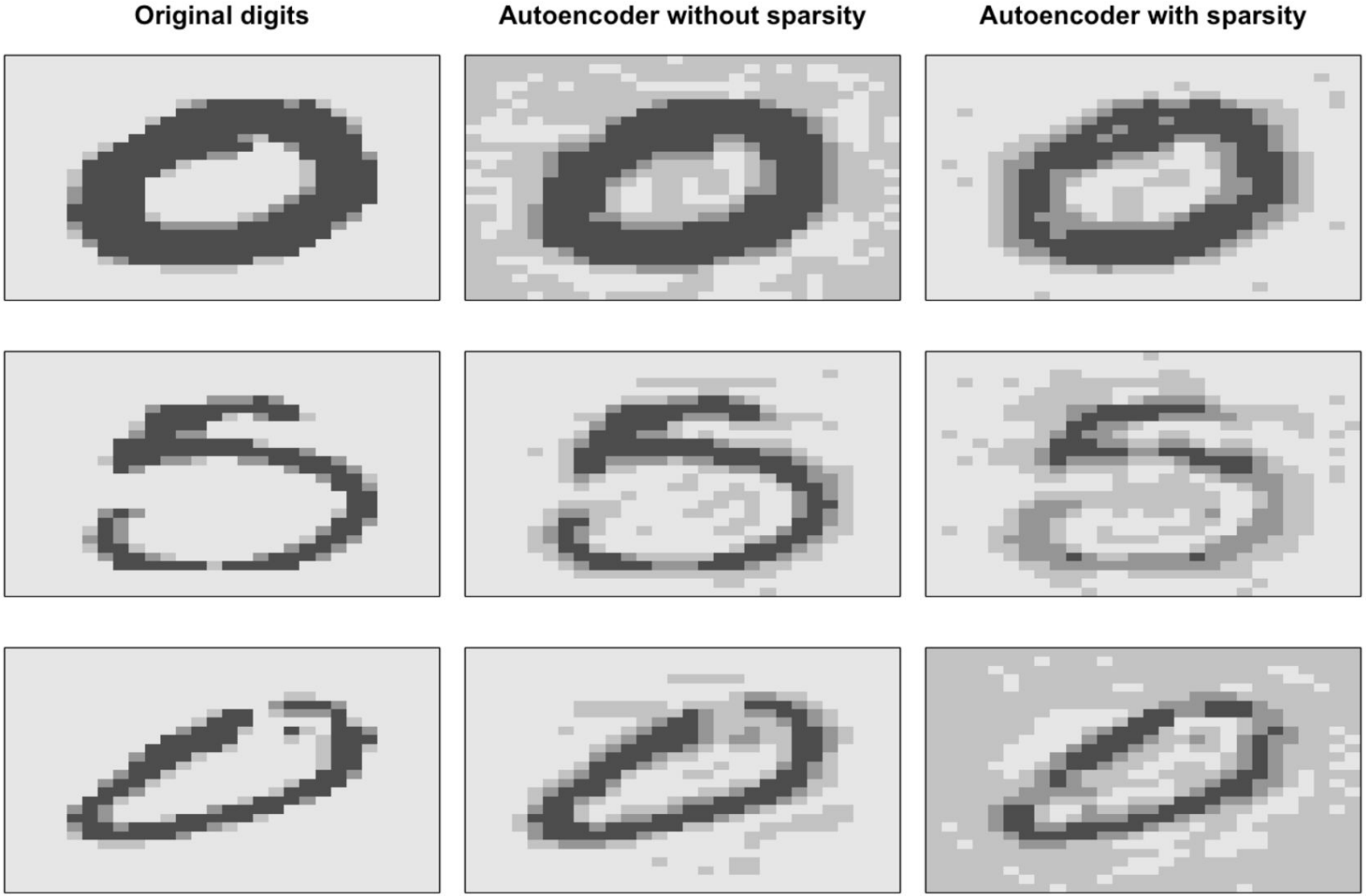


- The simplest one can just go directly from Input -> Code
- But we can have more layers => latent space of smaller dimension + less computing
- But if the model is **too good** => autoencoder = identity function => cannot learn anything
  => Here come regularization autoencoder, with a focus on **Sparse Autoencoder**

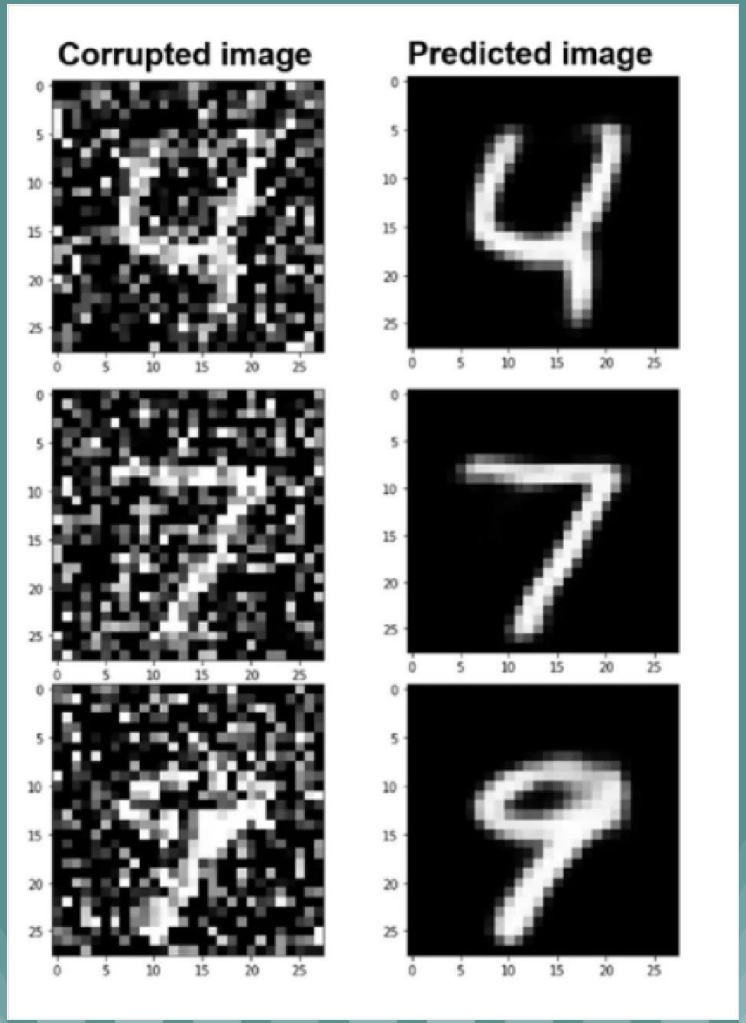
Sparse autoencoders incorporate a sparsity constraint during training.

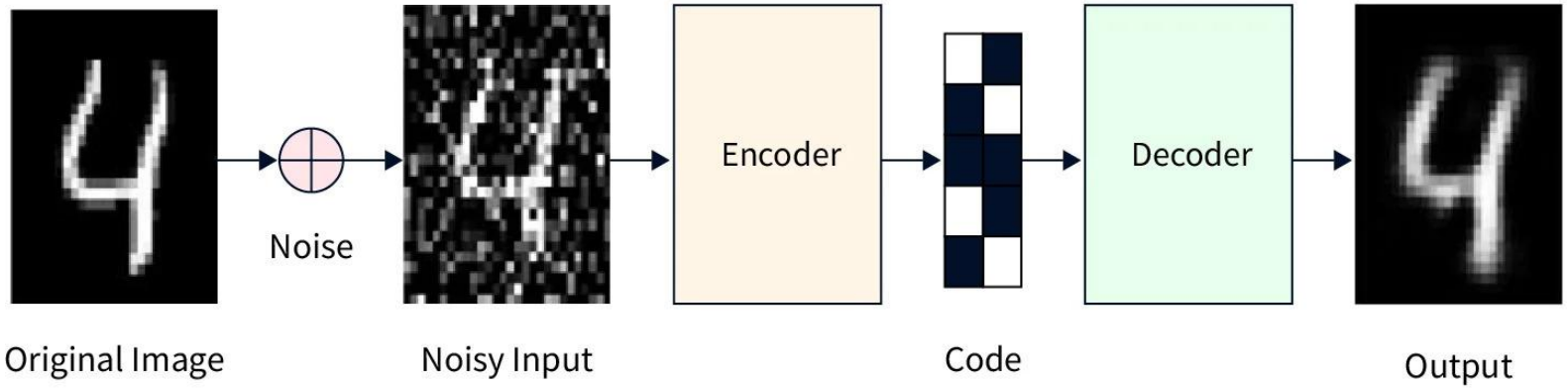
- Prevents overfitting
- Salient features are easier to pick out
- Faster computation (fewer active neurons)


Input layer
encoder
Hidden layers
decoder
Output layer
x_1
x_2
x_3
x_4
x_5
x_6
a_1
a_2
a_3
a_4
a_5
a_6
a_1
a_2
a_3
a_4
a_5
a_6
a_1
a_2
a_3
a_4
a_5
a_6
$\hat{x}_1$
$\hat{x}_2$
$\hat{x}_3$
$\hat{x}_4$
$\hat{x}_5$
$\hat{x}_6$


Original digits
Autoencoder without sparsity
Autoencoder with sparsity

Denoising Autoencoders are used for reducing noise from noisy data


Corrupted image
Predicted image


Noise
Encoder
Decoder
Original Image
Noisy Input
Code
Output

Applications of Autoencoders

- Dimensionality Reduction
- Feature Extraction
- Image Denoising
- Image/audio compression
- Image search
- Anomaly detection
- Missing value imputation