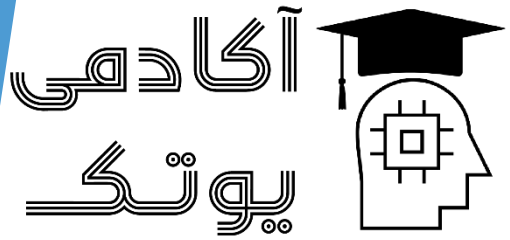


Machine Learning Course

Vahid Reza Khazaie

Unsupervised Learning

- **Dimensionality Reduction**
- **Clustering**
- **Autoencoding**
- **Anomaly Detection**
- **…**

Autoencoder

- Autoencoders are a type of **unsupervised neural network** (i.e., no class labels or labeled data) that seek to:
  - Accept an input set of data (i.e., the *input*).
  - Internally *compress* the input data into a **latent-space representation** (i.e., a single vector that *compresses* and *quantifies* the input).
  - **Reconstruct the input data** from this latent representation (i.e., the output).

Autoencoder

- **Encoder:** Accepts the input data and compresses it into the latent-space.
- **Decoder:** The decoder is responsible for accepting the latent-space representation and then reconstructing the original input.

X = input image

E -> encoder

D -> decoder

$\hat{X}$ = reconstructed image

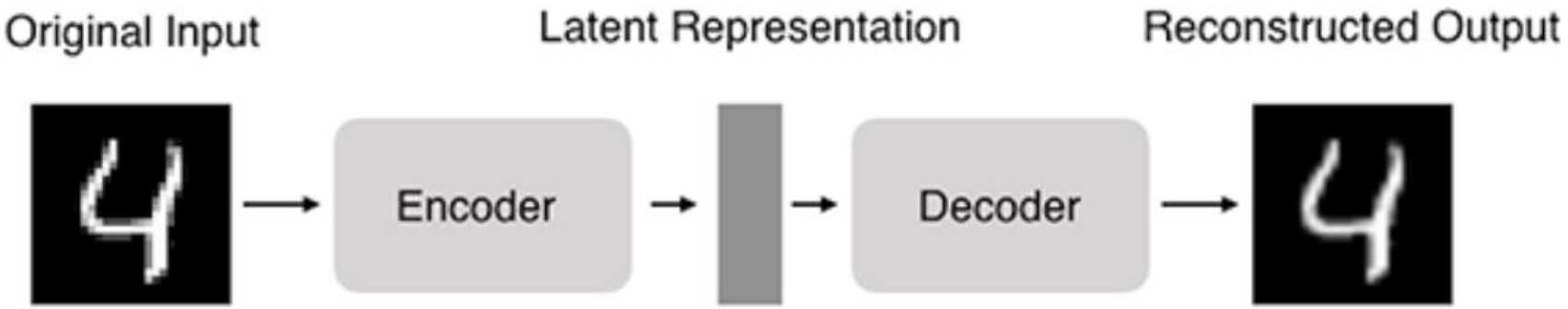
$\hat{X}$ = D(E(X))

Autoencoder

- We input a digit to the autoencoder.
- The encoder subnetwork creates a latent representation of the digit. This latent representation is ***substantially smaller*** (in terms of dimensionality) than the input.
- The decoder subnetwork then reconstructs the original digit from the latent representation.

Autoencoder

- **An autoencoder is a network that *reconstructs its input?!***

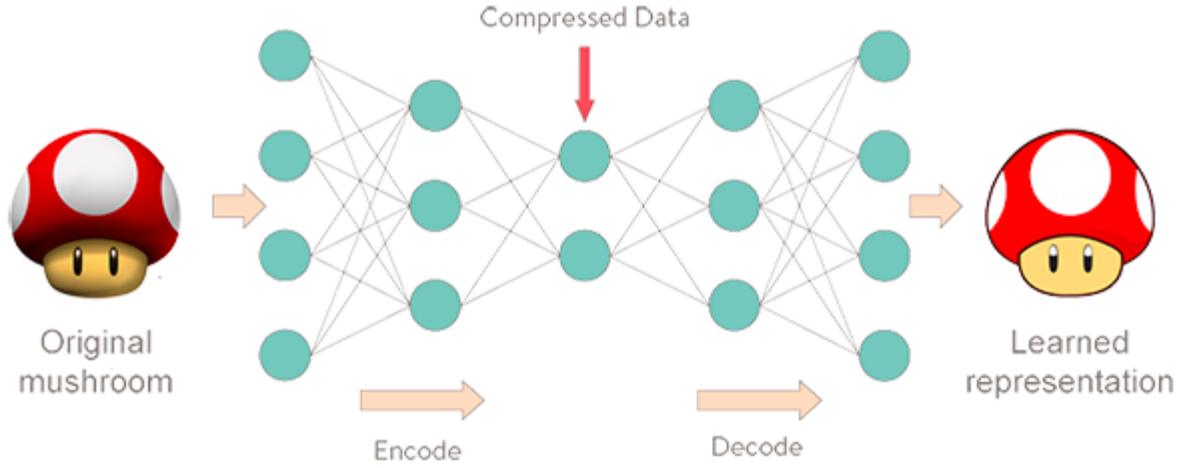


Autoencoder

- During the training process, our goal is to train a network that can learn how to reconstruct our input data. So what's the point?
- **The true value of the autoencoder lives inside that latent-space representation.**
- Keep in mind that autoencoders *compress* our input data and, more to the point, when we train autoencoders, what we *really* care about is the **encoder.**

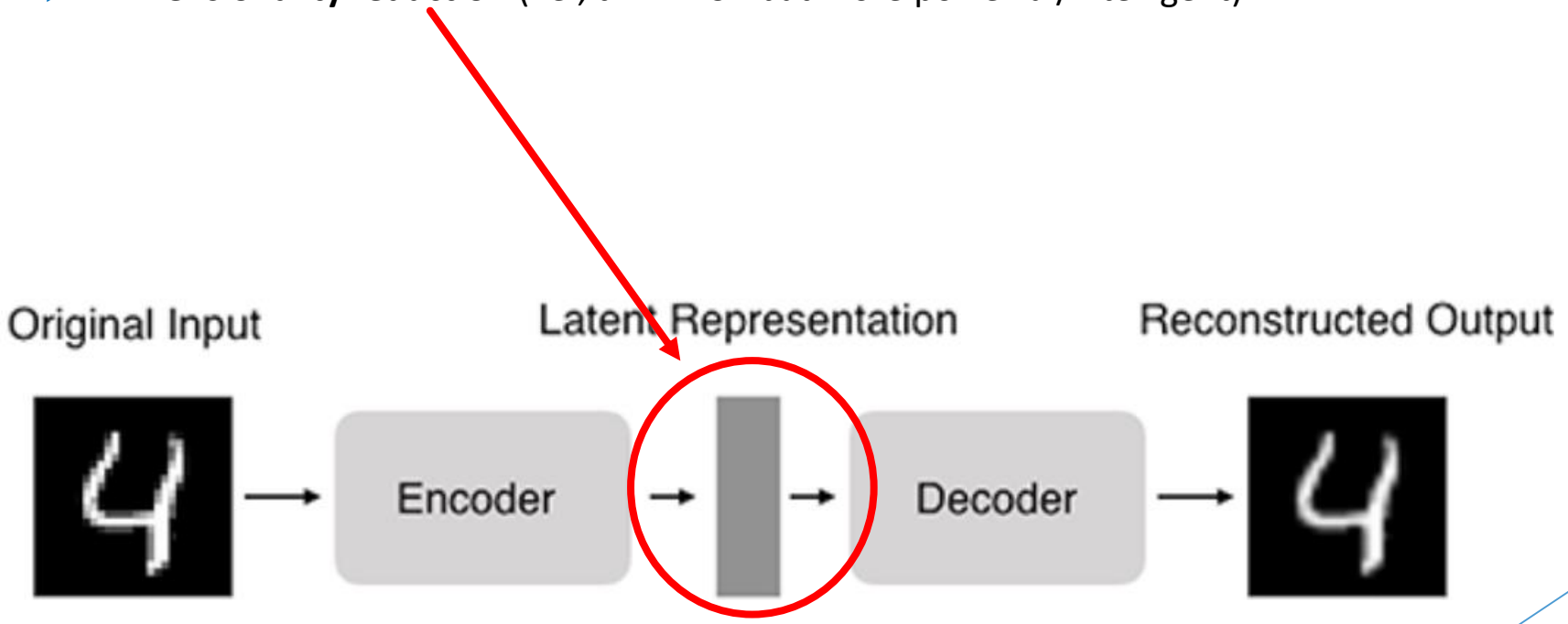
Autoencoder

- **Loss is difference between the reconstructed and the original image**



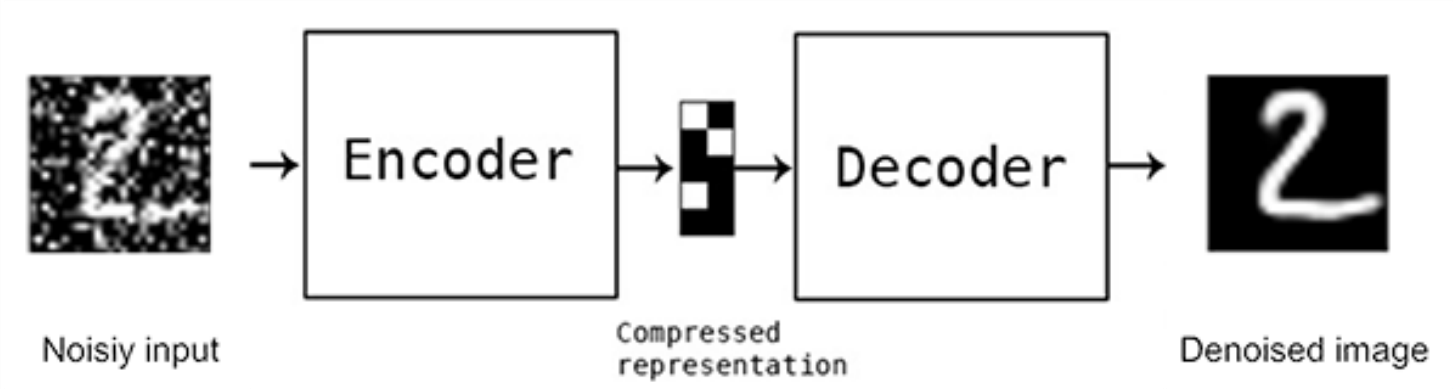
Autoencoder

- **Applications of autoencoders**
  - **Dimensionality reduction** (i.e., think PCA but more powerful/intelligent).



Autoencoder

- **Applications of autoencoders**
  - **Denoising** (ex., removing noise and preprocessing images to improve OCR accuracy).



Autoencoder

- **Applications of autoencoders**
  - **Anomaly/outlier detection** (ex., detecting mislabeled data points in a dataset or detecting when an input data point falls well outside our typical data distribution).