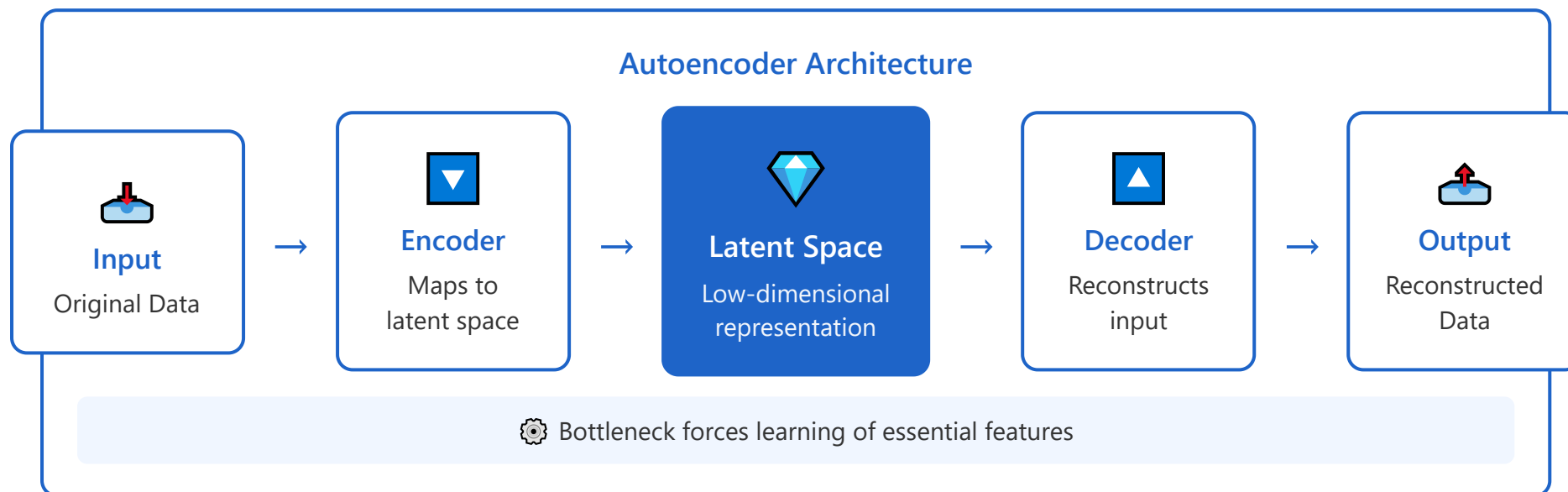


Autoencoders and Latent Representations

Unsupervised learning of compressed data representations



Historical Context

1980s: Hinton and Rumelhart pioneered dimensionality reduction using backpropagation

2006: Hinton's Deep Belief Networks marked the dawn of the deep learning era

2013: Kingma and Welling proposed the Variational Autoencoder (VAE)

Present: Evolved into a core technology for generative models, anomaly detection, and representation learning

Structural Details

Encoder: $x \rightarrow z = f_{\text{enc}}(x; \theta_{\text{enc}})$

- Compresses input dimension n to latent dimension d ($d \ll n$)
- Composed of multiple neural network layers (e.g., FC layers, CNNs)

Decoder: $z \rightarrow \hat{x} = f_{\text{dec}}(z; \theta_{\text{dec}})$

- Reconstructs original input from latent representation z
- Can be symmetric or independent from encoder architecture

Loss Function: $L = \|x - \hat{x}\|^2$ (reconstruction error)



Vector Operation Example

Simple Example: 784-dimensional MNIST image \rightarrow 2-dimensional latent space

Input vector: $x \in \mathbb{R}^{784}$ (28×28 image flattened)
e.g., $x = [0.1, 0.0, 0.8, \dots, 0.3]$

Encoder operations:

$h_1 = \text{ReLU}(W_1 x + b_1)$ # $W_1 \in \mathbb{R}^{(128 \times 784)}$
 $h_2 = \text{ReLU}(W_2 h_1 + b_2)$ # $W_2 \in \mathbb{R}^{(64 \times 128)}$
 $z = W_3 h_2 + b_3$ # $W_3 \in \mathbb{R}^{(2 \times 64)}$, $z \in \mathbb{R}^2$
 $\rightarrow z = [1.2, -0.5]$ # 2D latent vector

Decoder operations:

$h_3 = \text{ReLU}(W_4 z + b_4)$ # $W_4 \in \mathbb{R}^{(64 \times 2)}$
 $h_4 = \text{ReLU}(W_5 h_3 + b_5)$ # $W_5 \in \mathbb{R}^{(128 \times 64)}$
 $\hat{x} = \sigma(W_6 h_4 + b_6)$ # $W_6 \in \mathbb{R}^{(784 \times 128)}$
 $\rightarrow \hat{x} = [0.09, 0.02, 0.79, \dots, 0.31]$

Loss: $\text{MSE} = (1/784) \sum (x_i - \hat{x}_i)^2 = 0.003$



Variational Autoencoders (VAE)

Learn probabilistic latent distributions for generation and sampling

- Models latent space as probability distribution in the form $z \sim N(\mu, \sigma^2)$
- Learns regularized latent space with additional KL divergence



Dimensionality
Reduction



Denoising



Generation