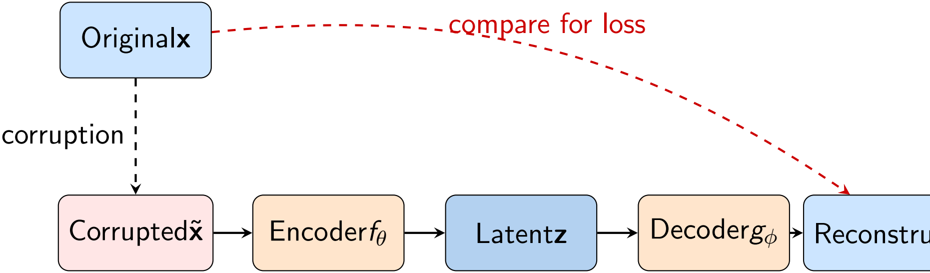
Denoising Autoencoders (Continued)



Frame Title

Common Noise Types

- **Gaussian noise**: $\tilde{x} = x + \mathcal{N}(0, \sigma^2)$
- **Salt & pepper noise**: Random pixels set to min/max
- **Dropout noise**: Random features set to zero
- **Masking**: Blank out regions (used in MAE)

Applications

- Image denoising and restoration
- Signal processing and enhancement
- Missing data imputation
- Self-supervised pretraining
- Robust feature extraction

Linear Autoencoders and PCA - Theoretical Connection

Linear Autoencoder Configuration

A linear autoencoder with the following properties is equivalent to PCA:

- **Linear activations** (no non-linearities)
- **Single hidden layer** (one encoder, one decoder layer)
- **MSE loss function** ($\|x - \hat{x}\|^2$)
- **Undercomplete** representation ($\dim(z) < \dim(x)$)

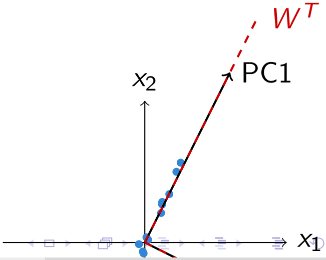
Mathematical Equivalence

For a linear autoencoder:

$$z = Wx \tag{1}$$

$$\hat{x} = W'z = W'Wx \tag{2}$$

At the optimum, the weight matrices:



PCA vs. Autoencoder - Key Differences

PCA Properties

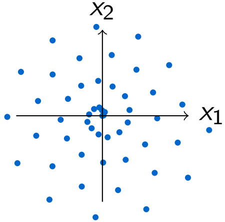
- **Orthogonal components**
  - Directions are perpendicular
  - Uncorrelated features
- **Variance maximization**
  - Directions of maximum variance
  - Components ordered by importance
- **Closed-form solution**
  - Eigendecomposition of covariance
  - Computationally efficient
- **Linear transformations only**
  - Cannot capture non-linear

Autoencoder Properties

- **Not constrained to orthogonality**
  - More flexible representations
  - Can have correlated features
- **Reconstruction optimization**
  - Minimizes reconstruction error
  - No explicit ordering of features
- **Iterative solution**
  - Trained with gradient descent
  - Potentially more computationally intensive
- **Non-linear transformations**
  - Can model complex relationships
  - More powerful for complex data

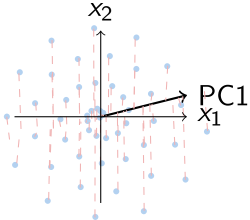
Comparative Visualization - PCA vs. Autoencoder

Original Data



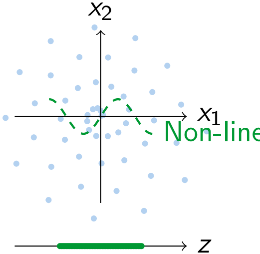
- Non-linear structure (spiral)
- 2D data for visualization
- Complex pattern

PCA Projection (1D)



- Linear projection only
- Cannot capture spiral structure
- Points overlap in projection

Autoencoder Encoding (1D)



1D latent representation

- Non-linear encoding function
- Preserves structure in 1D
- Points remain

The Bottleneck Layer - Information Theory Perspective

Information Bottleneck

- The bottleneck layer creates an **information constraint**
- Forces network to:
  - Discard irrelevant/noisy information
  - Preserve essential structure
  - Learn efficient encodings
- Size determines compression level

Frame Title

Information Theory View

$$I(X;Z) \leq \min(H(X), H(Z)) \quad (5)$$
$$H(Z) \leq \log(|\mathcal{Z}|) \quad (6)$$

where:

- $I(X;Z)$ is mutual information
- $H(X)$, $H(Z)$ are entropies
- $|\mathcal{Z}|$ is size of latent space

Compression vs. Reconstruction Trade-off

- **Too narrow bottleneck:** Poor reconstruction, too much info lost
- **Too wide bottleneck:** Perfect reconstruction but no useful compression
- **Optimal bottleneck:** Preserves essential structure while reducing

Overview

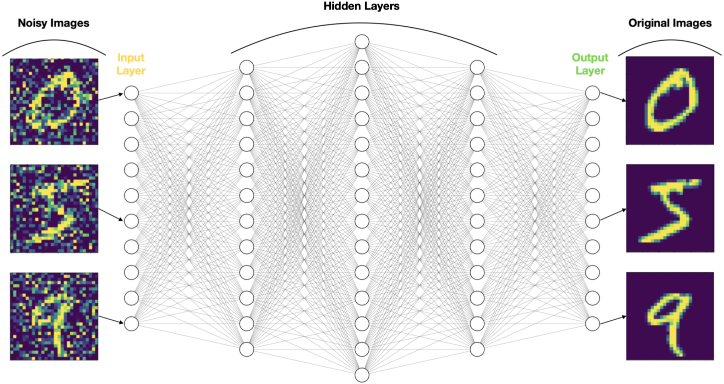


Figure: DAE