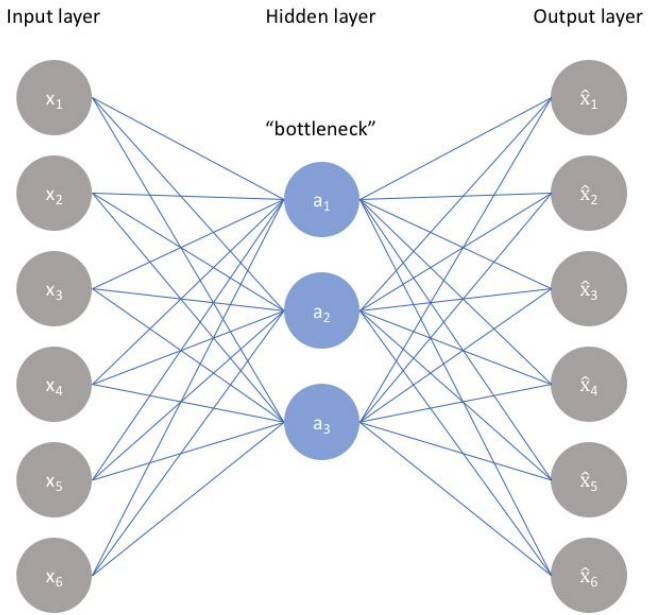
Practices in visual computing 2

Lab2: Autoencoders

Simon Fraser University
Spring 2025

Autoencoder

An autoencoder is a neural network designed to learn an efficient representation (encoding) of data.

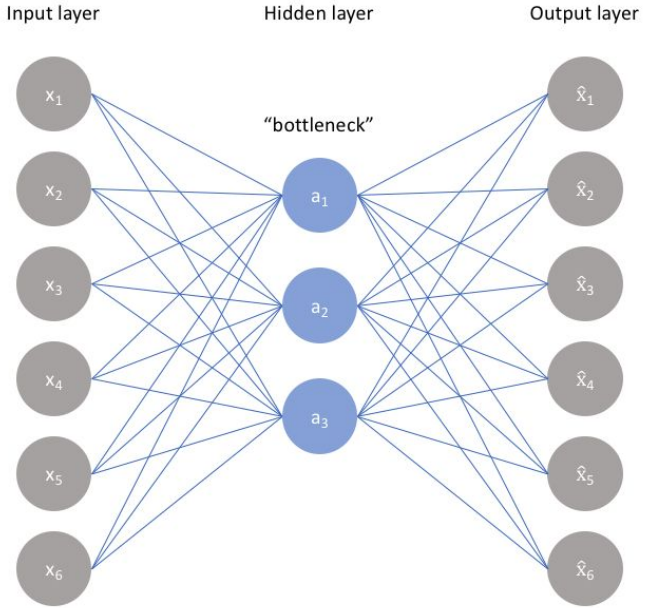


Autoencoder

Encoder: Maps the input data x to a latent representation z.

Latent Space: A compressed or reduced representation of the data.

Decoder: Reconstructs the input data from z.



Loss Function

Reconstruction Loss (e.g., Mean Squared Error, Cross-Entropy)

Objective: Minimize the difference between original input x and reconstruction x′.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Different Types of Autoencoders

Undercomplete Autoencoder: Latent dimension < input dimension.

Overcomplete Autoencoder: Latent dimension ≥ input dimension, typically with regularization.

Sparse Autoencoder: Encourages a sparse distribution in the latent space (few active neurons).

Denoising Autoencoder: Trains to reconstruct clean data from noisy inputs.

Variational Autoencoder (VAE): Learns a probabilistic latent space; used in generative modeling.

Applications of Autoencoders

Dimensionality Reduction: Like PCA, but nonlinear.

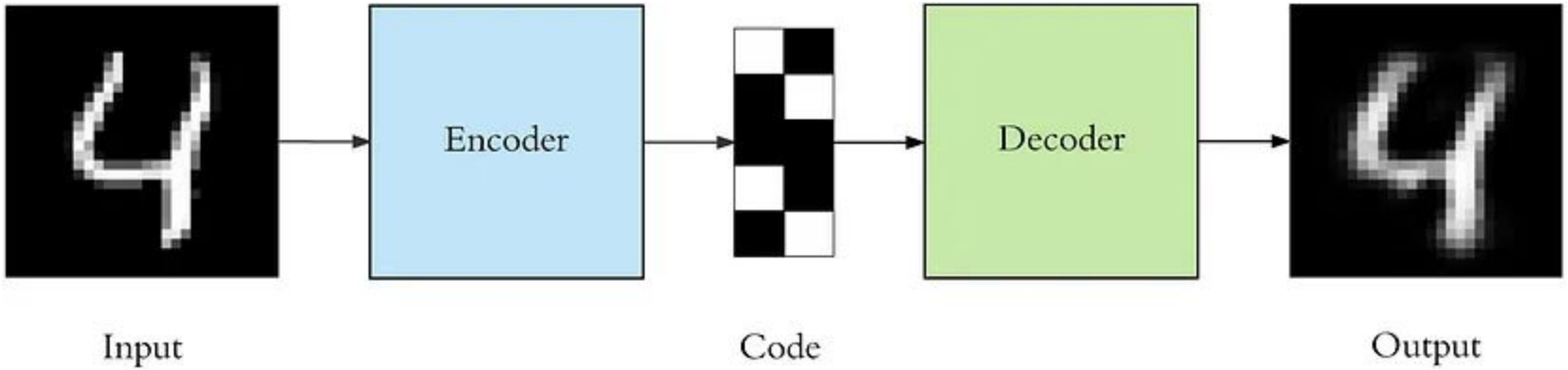
Denoising: Clean images, signals from noisy inputs.

Anomaly Detection: Learn normal patterns; high reconstruction error signals an anomaly.

Feature Extraction: Compressed latent representations can be used in downstream tasks.

Generative Modeling (VAE): Generate new, unseen samples.

Mnist Autoencoder



Implementation Details

Model: 2 linear layers with ReLu and Sigmoid

Epochs: 20

Optimizer: Adam

Learning Rate: 0.001

Reference

https://www.geeksforgeeks.org/auto-encoders/

https://github.com/udacity/deep-learning-v2-pytorch/blob/master/autoencoder/linear-autoencoder/Simple_Autoencoder_Exercise.ipynb