Autoencoders

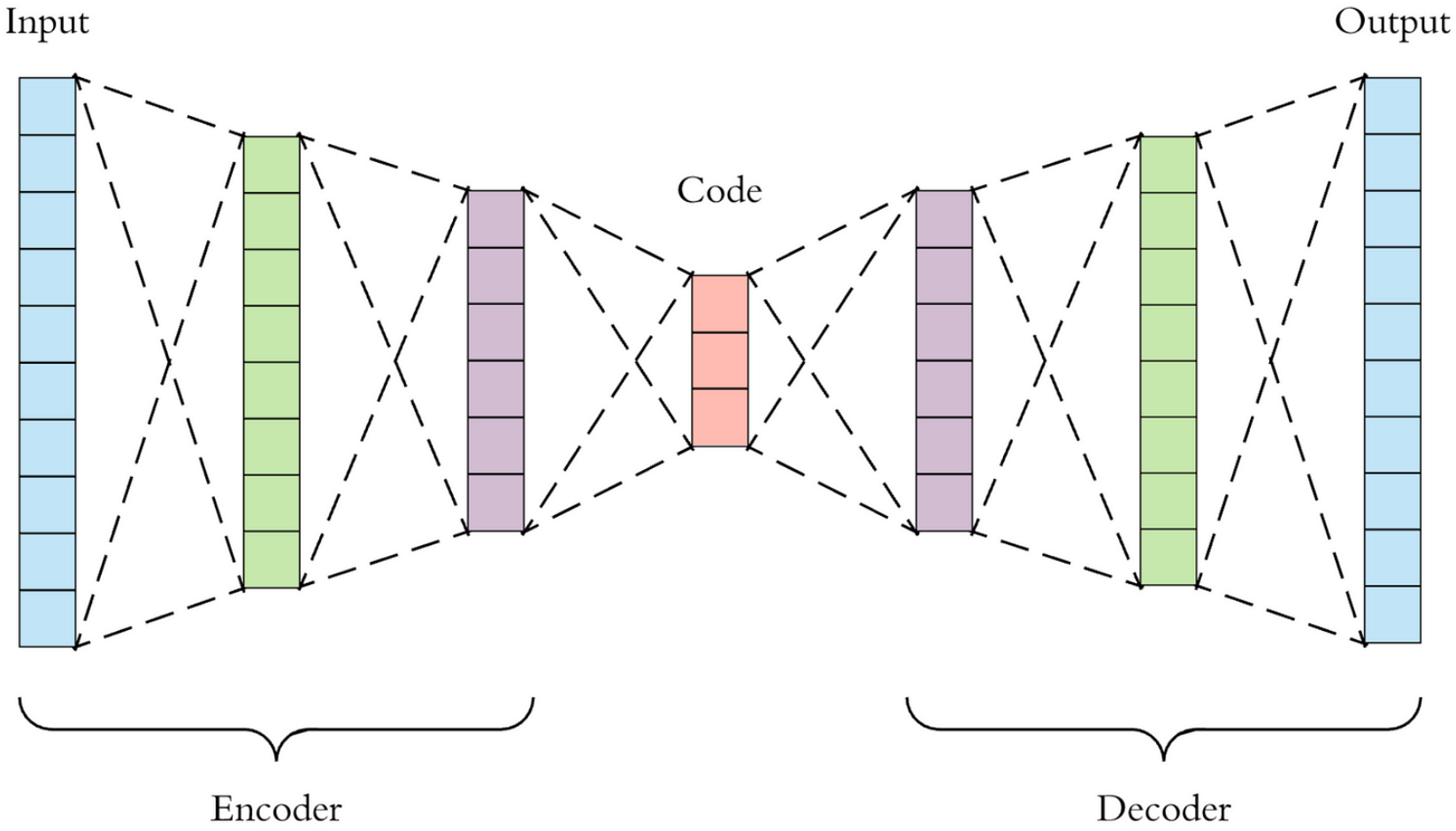
AN INTRODUCTION TO AUTOENCODERS IN DEEP LEARNING

What is an Autoencoder?

- A type of neural network used for unsupervised learning.
- Learns to encode input data into a compressed representation and then reconstruct it.
- Used for dimensionality reduction, denoising, anomaly detection, etc.

Architecture of an Autoencoder

- Encoder: Maps input to a lower-dimensional representation.
- Bottleneck: Compressed representation of input data.
- Decoder: Reconstructs the original input from the compressed representation.


Input
Output
Code
Encoder
Decoder

Types of Autoencoders

- Vanilla Autoencoder: Basic structure.
- Sparse Autoencoder: Introduces sparsity in hidden layers.
- Denoising Autoencoder: Trained to remove noise from input.
- Variational Autoencoder (VAE): Uses probabilistic distribution for latent space.

https://levelup.gitconnected.com/unsupervised-deep-learning-using-auto-encoders-4813bec64d39

Applications of Autoencoders

- Dimensionality Reduction
- Image Denoising
- Anomaly Detection
- Data Compression
- Generative Models (VAEs for synthetic data generation)

How Autoencoders Work?

- Input data is fed into the encoder.
- Encoder compresses data into a latent space representation.
- Decoder reconstructs the original input from latent space.
- The model is trained to minimize reconstruction loss.

Limitations of Autoencoders

- Cannot generalize well to unseen data.
- Struggles with complex data distributions.
- Requires careful tuning of network architecture and loss functions.

Conclusion

- Autoencoders are powerful tools for representation learning.
- Widely used in various applications such as anomaly detection and data compression.
- Advanced variations like VAEs enhance generative modeling capabilities.