Day 10: Autoencoders

Introduction to Autoencoders

What

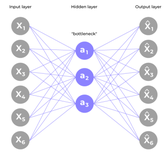
- Special type of neural network - Designed for Unsupervised learning
  - Primarily - dimensionality reduction and data compression, feature reduction
- Essential principle
  - Learning to compress and effectively represent input data without specific labels
  - Reduces the complexity of data, which reduces irrelevant data and improves performance

Working

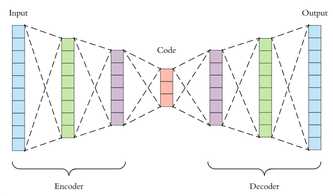
- Using a two-fold structure that consists of an encoder and a decoder
- Encoder:
  - Transforms the input data into a reduced-dimensional representation (a.k.a. latent space)
- Decoder:
  - rebuilds the initial input
- This process also defines essential features to gain meaningful patters

Architecture

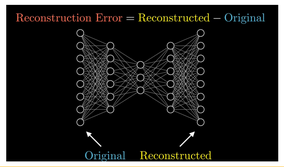
- An encoder, decoder, and bottleneck layer



- Encoder:
  - Input layer takes raw input data
  - Hidden layers progressively reduce the dimensionality of the input
    - Captures important features and patterns
- Decoder:
  - Bottleneck layer: takes the encoded representation and expands it back to the dimensionality of the original input
  - Hidden layers progressively increase the dimensionality and aim to reconstruct the original input
  - Output layer produces the reconstructed output, which ideally should be as close as possible to the input data
  - These results backpropagate the neural network in the form of the loss function



- Loss function used during training is the reconstruction loss

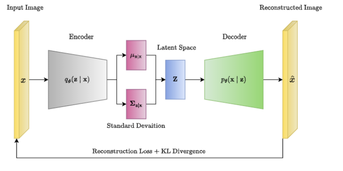


  - Measures the difference between the input and the reconstructed output
- During training the autoencoder learns to minimize the reconstruction loss
  - Which forces the network to capture the most important features in the bottleneck layer

Introduction to Variational Autoencoders (VAEs)

- It is a more advanced version of the Autoencoder in a way that it provides a statistical manner for describing the samples of the dataset in latent space
- In a VAE: the encoder outputs a probability distribution in the bottleneck layer instead of a single output value

Architecture of Variational Autoencoder



- The encoder network takes raw input data and transforms it into a probability distribution within the latent space
- Decoder takes the sampled point from the latent distribution and reconstructs it back into data space
- The loss function in VAE consists of two terms: Reconstruction Loss & Regularization term
  - Reconstruction Loss: Compels the model to accurately reconstruct the input
  - Regularization Term: Makes the latent space stick to the chosen distribution
- These parameters are iteratively adjusted to make the VAE learn to encode input data into meaningful latent space representation

Hands-On Project: Variational Autoencoders for Image Compression