



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

Denoising Autoencoders

Denoising autoencoders are a specific type of neural network that enables unsupervised learning of data representations or encodings. Their primary objective is to reconstruct the original version of the input signal corrupted by noise.

An autoencoder consists of two main components:

- Encoder: This component maps the input data into a low-dimensional representation or encoding.
- Decoder: This component returns the encoding to the original data space.

During the training phase, present the autoencoder with a set of clean input examples along with their corresponding noisy versions. The objective is to learn a task using an encoder-decoder architecture that efficiently transforms noisy input into clean output.

Encoder:

- The encoder creates a neural network equipped with one or more hidden layers.
- Its purpose is to receive noisy input data and generate an encoding, which represents a low-dimensional representation of the data.
- Understand an encoder as a compression function because the encoding has fewer parameters than the input data.

Decoder:

- Decoder acts as an expansion function, which is responsible for reconstructing the original data from the compressed encoding.
- It takes as input the encoding generated by the encoder and reconstructs the original data.
- Like encoders, decoders are implemented as neural networks featuring one or more hidden layers.

During the training phase, present the denoising autoencoder (DAE) with a collection of clean input examples along with their respective noisy counterparts. The objective is to acquire a function that maps a noisy input to a relatively clean output using an encoder-decoder architecture. To achieve this, a reconstruction loss function is typically employed to evaluate the disparity between the clean input and

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

the reconstructed output. A DAE is trained by minimizing this loss through the use of backpropagation, which involves updating the weights of both encoder and decoder components.

The training process involves minimizing the discrepancy between the original and reconstructed images. Once the DAE has completed its training, it can be employed to denoise new images by removing unwanted noise and reconstructing the original image.

