



Module 3

Denoising Autoencoders

A **denoising autoencoder** is a neural network model that removes noise from corrupted or noisy data by learning to reconstruct the original data from the noisy version. It is trained to minimize the difference between the original and reconstructed data. Denoising autoencoders can be stacked to form deep networks for improved performance.

The architecture can be adapted to other data types, such as images, audio, and text. Adding noise can be customized, e.g., Salt-and-Pepper, Gaussian noise.

What is Denoising Autoencoders?

Denoising autoencoders are a type of neural network that we can use to learn a representation (encoding) of data in an unsupervised manner. They are trained to reconstruct a clean version of an input signal corrupted by noise. This can be useful for tasks such as image denoising or fraud detection, where the goal is to recover the original signal from a noisy version.

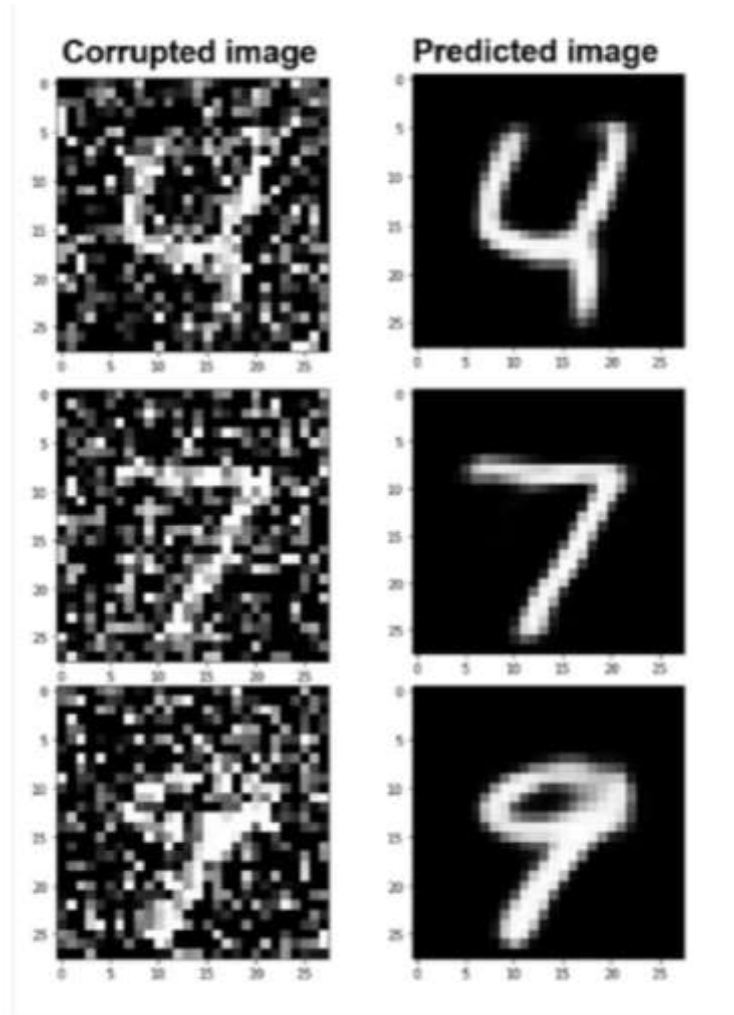
An autoencoder consists of two parts:

- encoder
- decoder.

The **encoder** maps the input data to a lower-dimensional representation or encoding, and the **decoder** maps the encoding back to the original data space. During training, the autoencoder is given a set of clean input examples and the corresponding noisy versions of these examples. The goal is to learn a function that maps the noisy input to the clean output using the encoder-decoder architecture.



Module 3



Architecture of DAE

The architecture of a denoising autoencoder (DAE) is similar to that of a standard autoencoder. It consists of an encoder, which maps the input data to a lower-dimensional representation, or encoding, and a decoder, which maps the encoding back to the original data space.

Encoder:

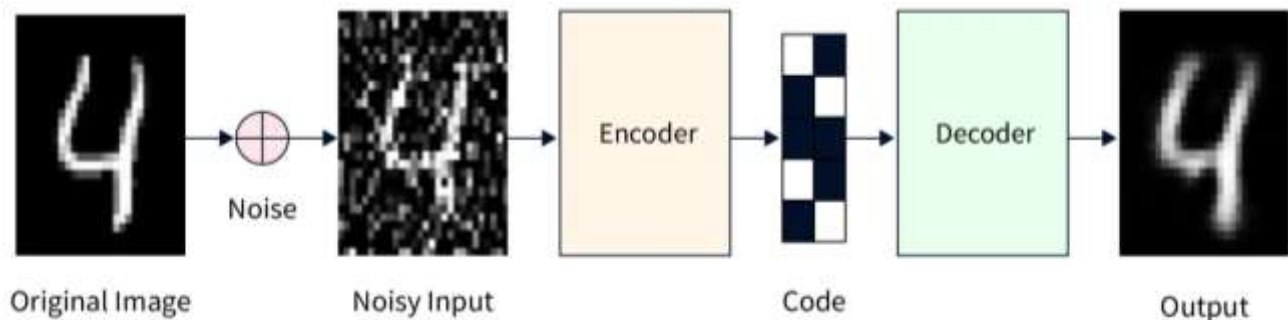
- Encoder is a neural network with one or more hidden layers.
- It takes noisy data as input and outputs an encoding.
- Encoding has fewer dimensions than the input data, so the encoder is a compression function.

Decoder:



Module 3

- Decoder is an expansion function that reconstructs original data from compressed encoding.
- Input to the decoder is the encoding from an encoder, and output is a reconstruction of the original data.
- Decoder is a neural network with one or more hidden layers.



During training, the DAE is given a set of clean input examples and the corresponding noisy versions of these examples. The goal is to learn a function that maps the noisy input to the clean output using the **encoder-decoder architecture**. This is typically done using a reconstruction loss function that measures the difference between the clean input and the reconstructed output. The DAE is trained to minimize this loss by updating the weights of the encoder and decoder through backpropagation.

Application of DAE

Denoising autoencoders (DAEs) have many applications in computer vision, speech processing, and natural language processing. Some examples include:

- **Image denoising:**
DAEs can be used to remove noise from images, such as Gaussian noise or salt-and-pepper noise.
- **Fraud detection:**
DAEs can be used to identify fraudulent transactions by learning to reconstruct normal transactions from noisy versions.
- **Data imputation:**
DAEs can impute missing values in a dataset by learning to reconstruct the missing values from the remaining data.
- **Data compression:**
DAEs can compress data by learning a compact data representation in the encoding space.
- **Anomaly detection:**
DAEs can identify anomalies in a dataset by learning to reconstruct normal data and then flagging difficult inputs as potentially abnormal.



Module 3

Denoising Images Using Denoising Autoencoder

DAEs are trained to reconstruct original data from a noisy version. Input to DAE is a noisy image; output is a reconstructed version of the original image. **Training** is done by minimizing the difference between the original and reconstructed images. Upon completion, the trained DAE can denoise new images by removing noise and reconstructing the original image.

Here is an example of how you could use a denoising autoencoder to denoise images using the Keras digits mnist dataset that contains 60,000 28x28 grayscale images of handwritten digits (0-9) for training and 10,000 images for testing.

Advantages and Limitations

Advantages:

- Denoising autoencoders can learn more robust features compared to standard autoencoders.
- They can improve the performance of other machine learning models by providing cleaner data.
- The approach is unsupervised, which means it does not require labeled data for training.

Limitations:

- The choice of noise and its level can greatly affect the performance of the model and may require careful tuning.
- Like other neural networks, denoising autoencoders can be computationally intensive to train, especially for large datasets or complex architectures.
- While they can remove noise, they may also lose some detail or relevant information in the data if not properly regularized.