Autoencoders

Autoencoders are a type of neural network that are used for unsupervised learning. They are called "autoencoders" because they are designed to encode (or compress) data and then decode (or decompress) it back to its original form.

Autoencoders consist of an encoder and a decoder. The encoder takes the input data and compresses it into a lower-dimensional representation, often referred to as a "latent space". The decoder then takes this compressed representation and reconstructs the original data as closely as possible.

Autoencoders are useful in a variety of applications, such as image and audio compression, anomaly detection, and data denoising. For example, in image compression, an autoencoder could be used to compress an image into a smaller file size while maintaining as much visual detail as possible. In anomaly detection, an autoencoder could be trained on a dataset of normal data, and then used to identify any data points that do not fit within that learned pattern.

To train an autoencoder, we typically use a loss function that measures the difference between the input and the output of the autoencoder. One commonly used loss function is mean squared error, which measures the average squared difference between the input and output.