AI 101

What is an Autoencoder?

Updated 11 months ago on September 20, 2020
By **Daniel Nelson**

TABLE OF CONTENTS

If you've read about unsupervised learning techniques before, you r
Autoencoders are one of the primary ways that unsupervised learni
autoencoder exactly?

Briefly, autoencoders operate by taking in data, compressing and er
data from the encoding representation. The model is trained until th
as closely as possible. Through this process, an autoencoder can le
that's a quick definition of an autoencoder, it would be beneficial to t
better understanding of how they function. This article will endeavor
architecture of autoencoders and their applications.

What is an Autoencoder?

Autoencoders are neural networks. Neural networks are composed
autoencoder is that the input layers contain exactly as much informa
input layer and output layer has the exact same number of units is t
data. It outputs a copy of the data after analyzing it and reconstructi

The data that moves through an autoencoder isn't just mapped stra
network doesn't just copy the input data. There are three componer
that compresses the data, a component that handles the compresse
portion. When data is fed into an autoencoder, it is encoded and the
network is then trained on the encoded/compressed data and it outp

AI Masterclass:

So why would you want to train a network to just reconstruct the dat
network learns the "essence", or most important features of the inpu
model can be created that can synthesize similar data, with the add
instance, you could train an autoencoder on grainy images and ther
from the image.

Autoencoder Architecture

Let's take a look at the architecture of an autoencoder. We'll discuss
There are variations on this general architecture that we'll discuss ir

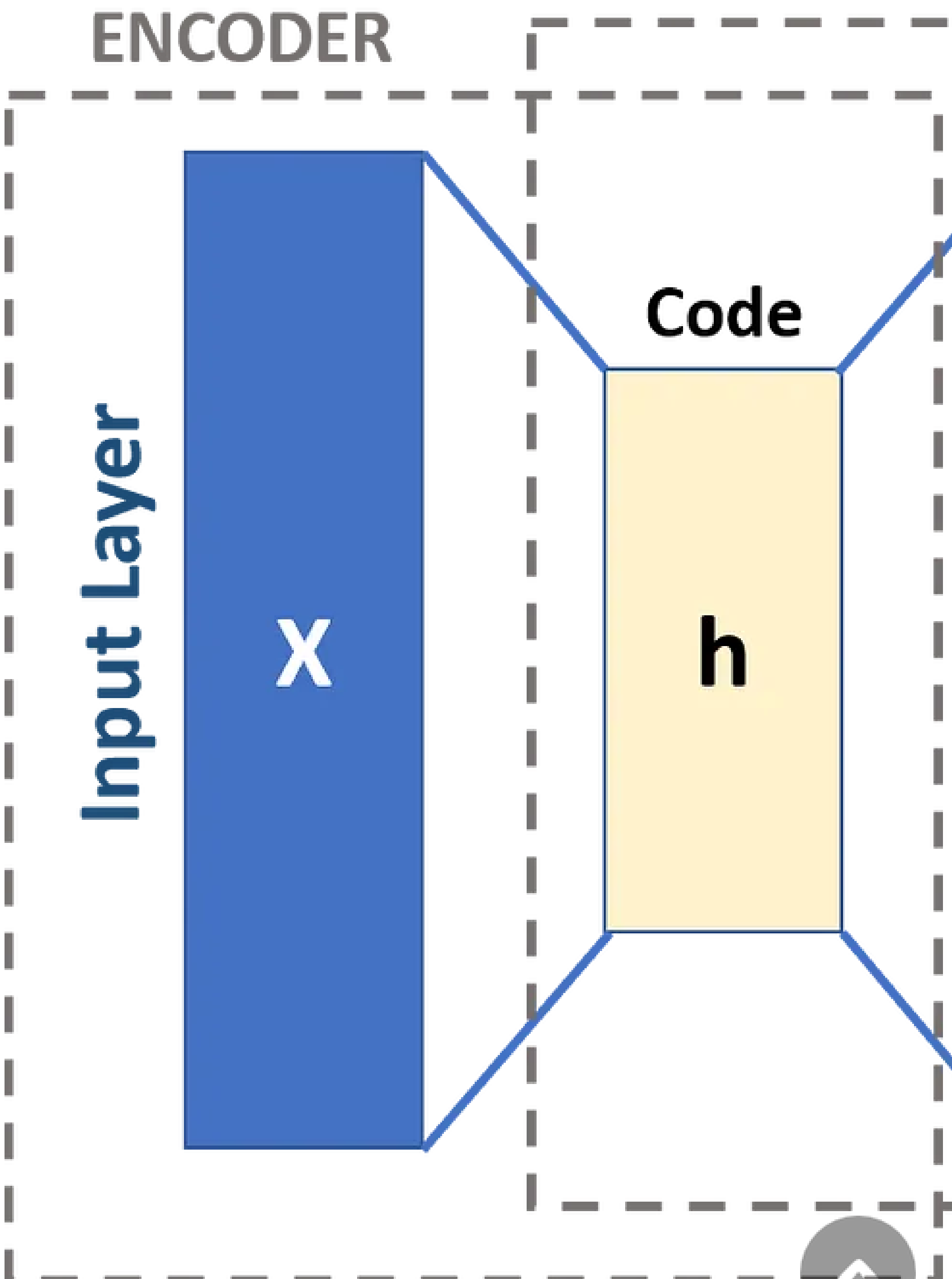


Photo: Michela Massi via Wikimedia Commons,(https://commons.wikimedia.org/wiki/File:Auto

AI Masterclass:

— Terminology (A to D)

Autoencoder

Backpropagation

Bayes Theorem

Big Data

Computer Vision

Confusion Matrix

Convolutional Neural Networks

Cybersecurity

Data Science

Decision Tree

Deepfakes

Deep Learning

Deep Reinforcement Learning

Dimensionality Reduction

+ Terminology (E to K)

+ Terminology (L to Q)

+ Terminology (R to Z)

As previously mentioned an autoencoder can essentially be divided
a bottleneck, and the decoder.

The encoder portion of the autoencoder is typically a feedforward, d
encoding layers is to take the input data and compress it into a later
representation of the data that has reduced dimensionality.

The code layers, or the bottleneck, deal with the compressed repres
carefully designed to determine the most relevant portions of the ob
features of the data that are most important for data reconstruction.
the data need to be preserved and which can be discarded. The bot
considerations: representation size (how compact the representatio
bottleneck performs element-wise activation on the weights and bia
sometimes called a latent representation or latent variables.

The decoder layer is what is responsible for taking the compressed
with the same dimensions as the original, unaltered data. The conve
representation that was created by the encoder.

The most basic architecture of an autoencoder is a feed-forward arc
layer perceptron used in multilayer perceptrons. Much like regular fe
trained through the use of backpropagation.

Attributes of An Autoencoder

There are various types of autoencoders, but they all have certain p

Autoencoders learn automatically. They don't require labels, and if g
autoencoder to reach high performance on a specific kind of input d

Autoencoders are data-specific. This means that they can only com
autoencoder has already been trained on. Autoencoders are also lo
be degraded in comparison to the input data.

When designing an autoencoder, machine learning engineers need
hyperparameters: code size, layer number, nodes per layer, and los

The code size decides how many nodes begin the middle portion of
data more. In a deep autoencoder, while the number of layers can

appropriate, the number of nodes in a layer should decrease as the
true in the decoder, meaning the number of nodes per layer should
final layer. Finally, the loss function of an autoencoder is typically eit
Binary cross-entropy is appropriate for instances where the input va

Autoencoder Types

As mentioned above, variations on the classic autoencoder architec
autoencoder architectures.

Sparse

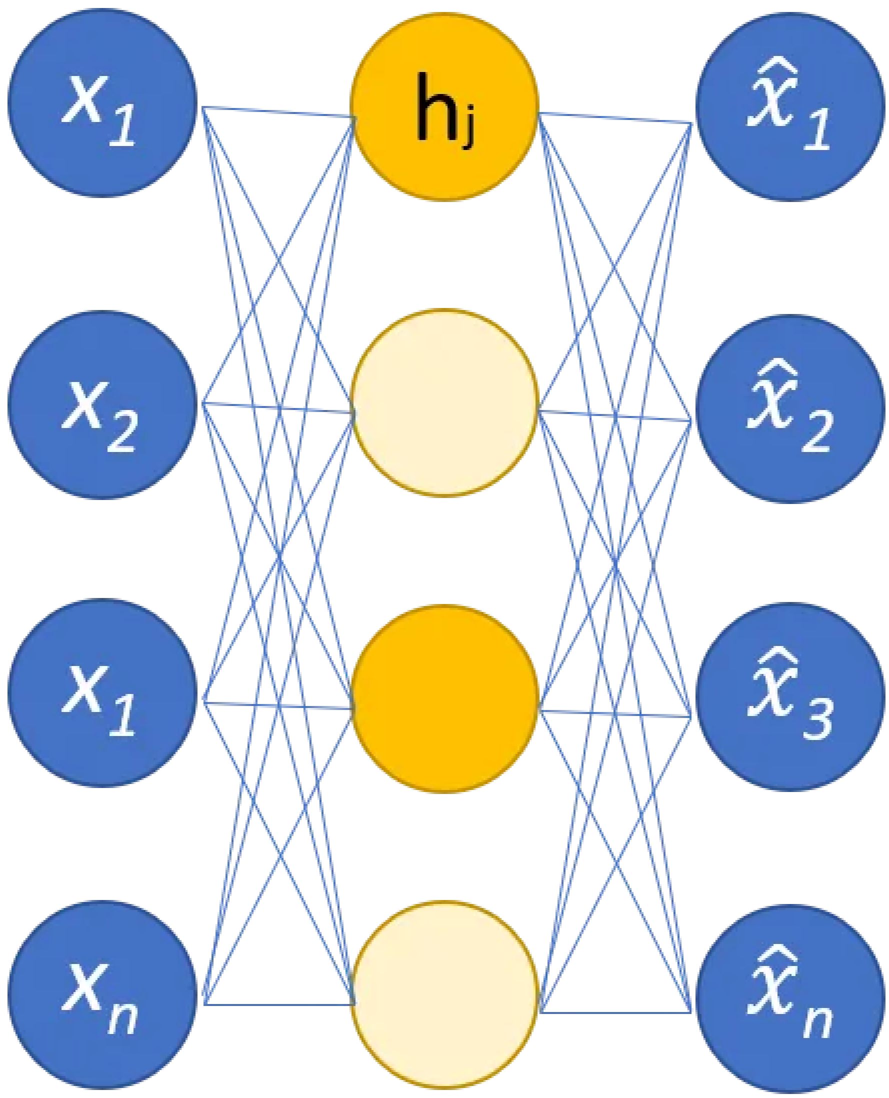


Photo: Michela Massi via Wikimedia Commons, CC BY SA 4.0
(https://commons.wikimedia.org/wiki/File:Autoencoder_sparso.png)

While autoencoders typically have a bottleneck that compresses the
autoencoders are an alternative to that typical operational format. In
the same size as the encoder and decoder layers. Instead, the activ
it up so the loss function better captures the statistical features of in

AI Masterclass:

hidden layers of a sparse autoencoder have more units than a tradi
them are active at any given time. The most impactful activation fun
this constraint helps the network determine just the most salient fea

Contractive

Contractive autoencoders are designed to be resilient against small
representation of the data. This is accomplished by applying a pena
technique is based on the Frobenius norm of the Jacobian matrix fo
regularization technique is that the model is forced to construct an e
encodings.

Convolutional

Convolutional autoencoders encode input data by splitting the data
subsections into simple signals that are summed together to create
convolution neural networks, a convolutional autoencoder specializ
filter that is moved across the entire image section by section. The e
be used to reconstruct the image, reflect the image, or modify the im
learned by the network, they can be used on any sufficiently similar

Denoising

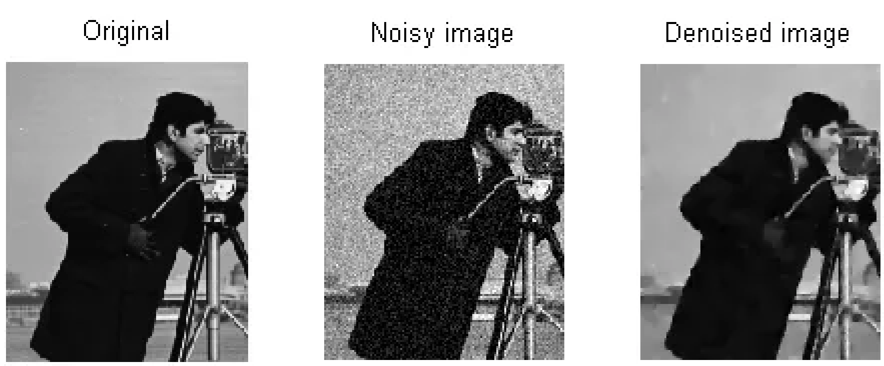


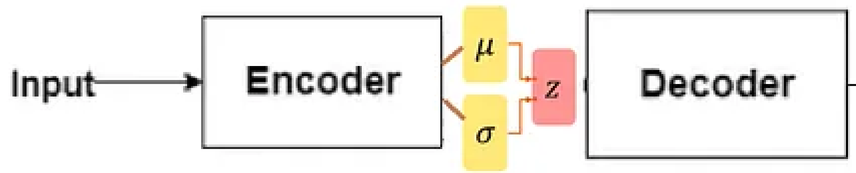
Photo: MAL via Wikimedia Commons, CC BY SA 3.0
(https://en.wikipedia.org/wiki/File:ROF_Denoising_Example.png)

Denoising autoencoders introduce noise into the encoding, resulting
original input data. This corrupted version of the data is used to train
output values with the original input and not the corrupted input. The
reproduce the original, non-corrupted version of the image. By comp
the network learns which features of the data are most important an
other words, in order for a model to denoise the corrupted images, i
the image data.

AI Masterclass:

Variational

Variational autoencoders operate by making assumptions about hov
A variational autoencoder produces a probability distribution for the
latent attributes. When training, the encoder creates latent distributi



learns the features or images as Gaussian distributions instead of d
generate new images. The Gaussian distribution is sampled to crea
network, which renders an image based on this vector of samples. E
the training images and assigns them some probability that they wil
used to reverse engineer an image, generating new images that res

When training the network, the encoded data is analyzed and the re
the mean and standard deviation of the images. A distribution is cre
different latent states. The decoder then takes random samples fro
to reconstruct the initial inputs to the network.

Autoencoder Applications

Autoencoders can be used for a wide variety of applications, but the
reduction, data denoising, feature extraction, image generation, seq
recommendation systems.

Data denoising is the use of autoencoders to strip grain/noise from i
repair other types of image damage, like blurry images or images m
help high capacity networks learn useful features of images, meanin
training of other types of neural networks. This is also true of using a
autoencoders can be used to identify features of other training datas

In terms of image generation, autoencoders can be used to generat
which has applications in designing face recognition systems or aut

AI Masterclass:

— Terminology (A to D)

Autoencoder
Backpropagation
Bayes Theorem
Big Data
Computer Vision
Confusion Matrix
Convolutional Neural Networks
Cybersecurity
Data Science
Decision Tree
Deepfakes
Deep Learning
Deep Reinforcement Learning
Dimensionality Reduction

+ Terminology (E to K)
+ Terminology (L to Q)
+ Terminology (R to Z)

Sequence to sequence prediction models can be used to determine autoencoder can be used to generate the next even in a sequence. to generate videos. Finally, deep autoencoders can be used to crea patterns relating to user interest, with the encoder analyzing user er recommendations that fit the established patterns.

RELATED TOPICS: #AUTOENCODER #UNSUPERVISED LEARNING

DON'T MISS
What Is Synthetic Data?

UP NEXT
What i

Daniel Nelson

Blogger and programmer with specialties in Machine Learning and Deep Learning social good.

YOU MAY LIKE

Microsoft Proposes GODIVA, a Text-To-Video Machine Learning Framework

AI Researchers Develop Method to Existing Drugs to Fight Covid-19

Supervised vs Unsupervised Learning