How is Autoencoder different from PCA

Last Updated : 22 Feb, 2022

In this article, we are going to see how is Autoencoder different from Principal Component Analysis (PCA).

Role of Dimensionality Reduction in ML

We often come into the curse of dimensionality issues in machine learning projects, when the amount of data records is not a significant component of the number of features. This often causes issues since it necessitates training a large number of parameters with a limited data set, which may easily lead to overfitting and poor generalization. High dimensionality also entails lengthy training periods. To solve these challenges, dimensionality reduction methods are often utilized. Despite its location in high-dimensional space, feature space often possesses a low-dimensional structure.

PCA and auto-encoders are two popular methods for lowering the dimensionality of the feature space.
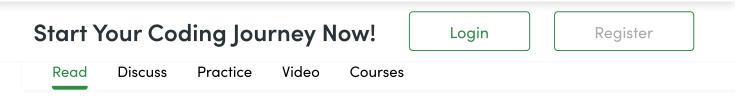
Principal Component Analysis (PCA)

PCA simply projects the data into another space by learning a linear transformation with projection vectors specified by the data's variance. Dimensionality reduction may be achieved by limiting the dimensionality to a small number of components that account for the majority of the variation in the data set.
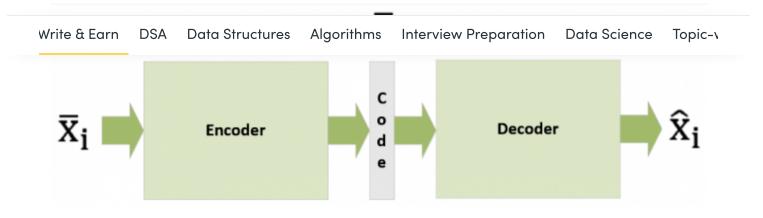
Autoencoders

Autoencoders are neural networks that stack numerous non-linear transformations to reduce input into a low-dimensional latent space (layers). They use an encoder-decoder system. The encoder converts the input into latent space, while the decoder reconstructs it. For accurate input reconstruction, they are trained through backpropagation. Autoencoders may be used to reduce dimensionality when the latent space has fewer dimensions than the input. Because they can rebuild the input, these low-dimensional latent variables should store the most relevant properties, according to intuition.

Simple Illustration of a generic autoencoder

PCA vs Autoencoder

- Although PCA is fundamentally a linear transformation, auto-encoders may describe complicated non-linear processes.

(However, regularization and proper planning might help to prevent this).

How to select the models?

Aside from processing computational resources, the choice of approach is influenced by the features of the feature space itself. If the features have a non-linear connection, the autoencoder may compress the data more efficiently into a low-dimensional latent space by utilizing its capacity to represent complicated non-linear processes.

Researchers created a two-dimensional feature space with linear and non-linear relationships between them (x and y are two features) (with some added noise). After projecting the input into latent space, we can compare the capabilities of autoenocoders and PCA to properly reconstruct the input. PCA is a linear transformation with a well-defined inverse transform, and the reconstructed input comes from the autoencoder's decoder output. For both PCA and autoencoders, we employ a one-dimensional latent space.

Autoencoded latent space may be employed for more accurate reconstruction if there is a nonlinear connection (or curvature) in the feature space. PCA, on the other hand, only keeps the projection onto the first principal component and discards any information that is perpendicular to it.

Conclusion:

There must be underlying low-dimensional structure in the feature space for dimensionality reduction to be successful. To put it another way, the characteristics should be related to one another. Autoencoders may encode more information with fewer dimensions if the low dim structure has non-linearity or curvature. As a result, in certain cases, they are a superior dimensionality reduction strategy.