☰ Item Navigation

Summary/Review

Autoencoders

Autoencoders are a neural network architecture that forces the learning of a lower dimensional representation of data, commonly images.

Autoencoders are a type of unsupervised deep learning model that use hidden layers to decompose and then recreate their input. They have several applications:

- Dimensionality reduction
- Preprocessing for classification
- Identifying 'essential' elements of the input data, and filtering out noise

One of the main motivations is find whether two pictures are similar.

Autoencoders and PCA

Autoencoders can be used in cases that are suited for Principal Component Analysis (PCA).

Autoencoders also help to deal with some of these PCA limitations: PCA has learned features that are linear combinations of original features.

Autoencoders can detect complex, nonlinear relationship between original features and best lower dimensional representation.

Autoencoding process

The process for autoencoding can be summarized as:

1. Feed image through encoder network
2. Generate the lower dimension embedding
3. Feed embedding through decoder network
4. Generate reconstructed version of the original data
5. Compare the result of the generated vs the original image