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

Result: A network will learn the lower dimensional space that represents the original data

Autoencoder applications

Autoencoders have a wide variety of enterprise applications:

- Dimensionality reduction as preprocessing for classification
- Information retrieval
- Anomaly detection
- Machine translation
- Image-related applications (generation, denoising, processing and compression)
- Drug discovery
- Popularity prediction for social media posts

- Sound and music synthesis
- Recommender systems

Variational Autoencoders

Variational autoencoders also generate a latent representation and then use this representation to generate new samples (i.e. images).

These are some important features of variational autoencoders:

- Data are assumed to be represented by a set of normally-distributed latent factors.
- The encoder generates parameters of these distributions, namely μ and σ.
- Images can be generated by sampling from these distributions.

VAE goals

The main goal of VAEs: generate images using the decoder

The secondary goal is to have similar images be close together in latent space

Loss Function of Variational Autoencoders

The VAE reconstruct the original images from the space of a vector drawn from a standard normal distribution.

The two components of the loss function are:

- A penalty for not reconstructing the image correctly.
- A penalty for generating vectors of parameters μ and σ that are different than 0 and 1, respectively: the parameters of the standard normal distribution.