Autoencoders

Annada Behera

December 13, 2021

What is a autoencoder?

- An autoencoder[1] is a artificial neural network that learns the efficient codings of unlabeled data.
- It works both as a discriminator and a generator.

[1] `https://en.wikipedia.org/wiki/Autoencoder`

Definition

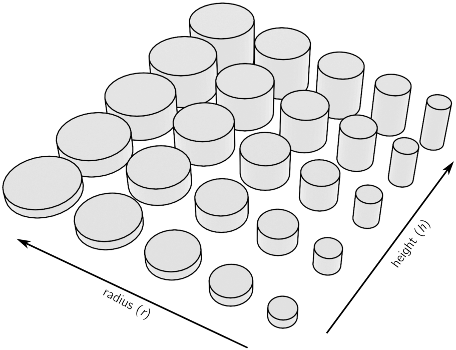
- Discriminator $p(y|x)$ estimates the probability of a label y given an observation x.
- Generator $p(x|y)$ is the probability of observing x for a optional label y.

Generator Framework

- We have a set of observation X.
- We assume that the observation was generated by some unknown distribution P_d.
- Our aim is to generate a model P_m that is as close to P_d as possible.
- We are happy with P_m if
  - The observations sampled from P_m appears to be drawn from P_d.
  - The observations sampled from P_m is considerably different from X.

An example

- Given a set of observations X in a very high dimensional sample space, they can be instead represented in a compact low dimensional space.
- Learning the low dimensional representation is called **representation learning**.
- The low dimensional representation is called **latent space**.



Definition

- The sample space is the set of all observation x can take, i.e, X.
- The probability density function, $p(x)$ is a function that maps a point in the sample space to a number between 0 and 1.
- An encoder $Enc(x)$ takes a high-dimensional input vector and "compresses" the data to a low dimensional output.
- A decoder, Dec "decompresses" the data from low dimension to a high dimension.

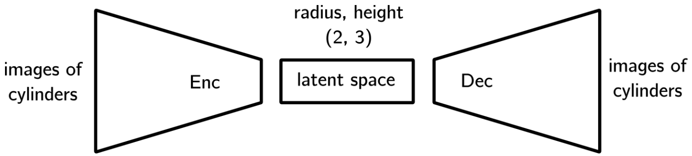
$$Enc : \mathbb{R}^m \to \mathbb{R}^n, \; Dec : \mathbb{R}^n \to \mathbb{R}^m \qquad m >> n \tag{1}$$

Definition

An autoencoder $AE : \mathbb{R}^n \to \mathbb{R}^n$ is defined as,

$$AE_\theta(x) = Dec_\theta(Enc_\theta(x)) \tag{2}$$

The *Enc* and *Dec* functions are both neural networks in our cases, but it is not necessary that they should be one.
θ is the weights of the neural networks.



The AE loss

The loss function is defined as, $\mathcal{L}(x, AE(x))$ is generally uses the RMSE loss, defined as,

$$RMSE(x, AE(x)) = \sqrt{\mathbb{E}[(x - AE(x))^2]} \tag{3}$$

and the KL-divergence loss,

$$D_{kl}(P_d, P_m) = \sum_{x \in X} P_d(x) \log \left(\frac{P_d(x)}{P_m(x)} \right) \tag{4}$$