SVKM'S NMIMS Deemed to be UNIVERSITY

Auto-Encoders

M.Tech. Data Science, Second Year, NMIMS

By,

Bilal Hungund, Data Scientist, Halliburton

Auto-Encoders

→ It is a special type of neural network architectures in which the output is same as the input.

→ It is trained in an unsupervised manner in order to learn extremely low representations of an input data.

→ These low level features are deformed back to its actual data.

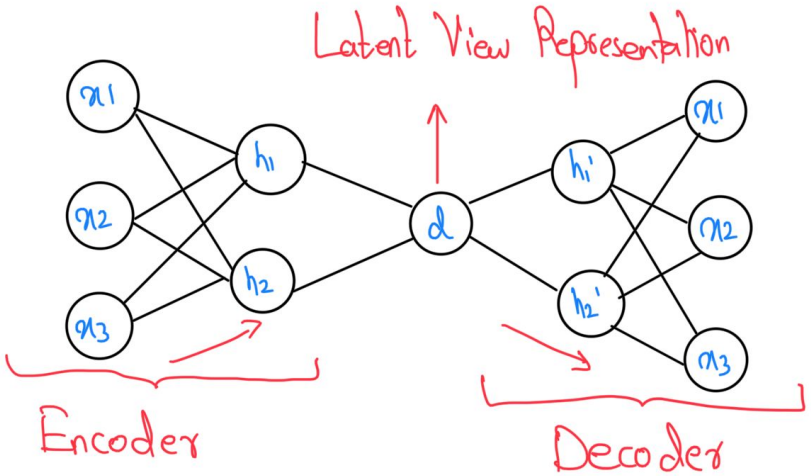
→ It is a regression task where the network is asked to predict its input (model identity function).

Auto - Encoders architecture

Encoding architecture: It comprises series of layers with decreasing number of nodes and ultimately reduces to a latent view representation

Latent View Representation: It represents the lowest level space in which the inputs are reduced and information preserved.

Decoding architecture: It is the mirror image of the encoder but in which number of nodes in every layer increases and ultimately outputs the similar (almost) input.



Encoders $z = f(Wx + b)$

Decoders $\hat{x} = f'(W'z + b)$

Loss $L(x, \hat{x}) = |x - \hat{x}|$

Use cases

- Image Reconstruction
- Image Enhancement
- Image Compression
- Image Denoising
- Feature Extraction
- Binary Classification

References

- https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73
- https://medium.com/analytics-vidhya/mathematical-prerequisites-for-understanding-autoencoders-and-variational-autoencoders-vaes-8f854025390e