

Implicit Autoencoders

Deterministic autoencoders often learn to copy inputs which prevents the extraction of meaningful representations in the form of implicit distributions. The work introduces the Implicit Autoencoder (IAE) consisting of generative and recognition parameterized by implicit distributions. Based on the maximum likelihood learning rule, implicit distributions facilitate the learning of more expressive posterior and conditional likelihood distributions. This is indicative of the latent code capturing only the high-level information in the data, while the low-level information is abstracted by the implicit distribution. IAEs demonstrate improved performance in disentangling content and style information, clustering, semi-supervised learning and learning of expressive variational distributions.

The framework of Generative Adversarial Networks (GANs) presents a generic method for learning implicit distributions. IAEs adopt this framework in order to learn implicit distributions which provision learning of expressive posterior and conditional likelihood distributions. This leads to a tighter variational bound and results in abstraction of high and low levels of information between prior and conditional likelihood respectively. Similar to conventional AEs, IAE consists of an encoder $q(z|x)$ which encodes the data sample $x \sim p_d(x)$ to a latent code \hat{z} upon corruption by a noise vector ϵ . The latent code \hat{z} along with a noise vector n are then decoded using the decoder $p(x|z)$ resulting in the reconstruction \hat{x} . In addition to AEs, the IAE utilizes adversarial reconstruction and regularization costs which provide a tighter approximation in comparison to VAE and AAE costs. Additionally, the costs are independent of data factorization which is indicative of the fact that IAEs motivate stochastic representations. Optimization of the IAE objective is carried out using GANs wherein each of the adversarial objectives is minimized using a separate GAN architecture.

IAEs present apt learning of high and low level representations captured by implicit distributions on the MNIST, SVHN and CelebA datasets. IAEs are also found suitable for clustering and semi-supervised learning wherein a categorical latent code allows disentangling of discrete and continuous variations. While GANs learn an interpolating mapping between samples, IAEs accurately group data points in the mixture of Gaussian distribution. Application of IAEs further extend to style and domain transfer by constructing CycleIAEs. While the CycleGAN only learns unimodal mappings between distributions, CycleIAE does away with the $L1$ reconstruction error and learns cross-domain mappings. However, this occurs at the cost of using a deterministic encoder and a stochastic decoder. In another experiment for translating female faces to male faces, the method employs a stochastic encoder with a deterministic decoder. A more suitable scheme for evaluating consistency across domains would be to assess CycleIAEs with both stochastic encoder and decoder. Lastly, the work extends IAEs to Flipped IAE (FIAE) which reconstructs the latent codes based on posteriors $p(x|z)$ and $q(z|x)$. FIAE provides accurate variational approximation using non-factorized distributions in comparison to InfoGAN which presents mode-averaging across different posterior clusters.

IAEs present a novel method for abstracting high and low level information by doing away with factorized distributions. To this end, the work introduces two new directions for future research. Firstly, IAEs can be extended towards complex and high-dimensional latent spaces wherein the latent code consists of greater noise per dimension. Secondly, future work could shed more light into the upper bound of FIAE and how it can be tightened to extend variational approximations towards high-dimensional true posterior distributions.