Student ID : 20194293

Name : Go, Kyeong Ryeol

[AI 502] β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework

1. Paper Summary

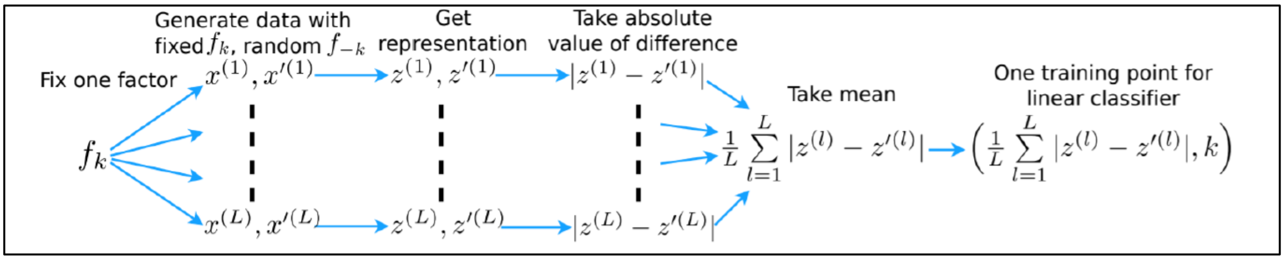
Learning good representation of high dimensional data is beneficial in terms of computational cost and generalization. When working on tasks like transfer learning, domain adaptation, and knowledge distillation, low dimensional and compressive representation would boost the learning and improve the overall model performance. The concept of disentanglement is suggested which implies that one latent dimension indicates one factor of generative factors while relatively invariant to others. Previously, the priori knowledge about the number and/or nature of these factors is required, but usually they are not feasible in the real-world problem.

Recently, a scalable unsupervised approach for disentangled factor learning has been developed by InfoGAN which extends Generative Adversarial Network (GAN) by further maximizing the mutual information between a subset of the generating noise variables and the output of the recognition network. However, since it's based on GAN framework, the model suffers from unstable training process, reduced sample diversity and lack of principled inference framework. Therefore, the author proposed 'β-VAE' which augments the original VAE framework with a single hyperparameter β that modulates the learning constraints. The validity of this model is verified in a qualitative manner through latent traverse and in a quantitative manner through the newly proposed disentanglement metric.

$$\max_{\phi,\theta} \mathbb{E}_{x\sim D}\left[\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]\right] \quad \textit{subject to} \;\; D_{KL}(q_\phi(z|x)||p_\theta(z)) < \varepsilon$$

The paper deals with the above constrained optimization problem. Here, the objective function is identical to Stochastic Gradients Variational Bayes (SGVB) estimator and the constraint is encouraging the learned posterior to be similar to the isotropic gaussian prior. To reformulate it to unconstrained optimization problem, Lagrange Theory can be applied and the author defines β to be identical to one plus Lagrange multiplier so that β must be set to greater than equal to 1. By regarding it as a hyperparameter, the disentanglement among feature is encouraged by setting its value to be high. However, since it results in empowering the regularization effect, the reconstruction quality may be degraded which requires elaborate tuning of β.

The disentanglement metric is devised to measure both independence and interpretability of the inferred latent variables. The following is the figure describing the way to compute the metric; In summary, the accuracy of linear classifier which predict the index of the generative factor that was kept fixed for a given batch is the proposed metric.



2. Discussion

How can we encourage disentanglement while the reconstruction quality is remained the same (or improve)? Total correlation, which is one of the generalizations of mutual information that was referred in InfoGAN, can be used when modifying ELBO with information theoretic perspective. Factor-VAE and Beta-TCVAE are recently proposed variants with this idea.