Student ID : 20194293

Name : Go, Kyeong Ryeol

[AI 502] Stochastic Gradient VB and the Variational Auto-Encoder

1. Paper Summary

When modeling latent variables in Linear Discriminant Analysis (LDA), EM algorithm can be used which consists of Expectation step and Maximization step. It is based on variational inference technique which introduce artificial distribution to approximate the true distribution. However, when dealing with the continuous latent variable, the denominator term in Bayes formula cannot be computed so that the posterior distribution is intractable. MCMC is a frequently referred sampling method to handle this issue which appears in various generative models such as RBMs, DBMs, DBN. However, MCMC is performed in expensive iterative manner and shows its weakness with high-dimensional latent variables due to mode collapse. Therefore, this author introduces SGVB estimator and AEVB algorithm to perform efficient and approximate inference on continuous latent variable with intractable posterior in large i.i.d dataset. Variational Auto-Encoder (VAE) is a newly proposed generative model where the proposed approach can be directly utilized.

The marginal likelihood $p(x)$ can be decomposed into two terms which are Evidence Lower Bound (ELBO) and KL divergence between variational inference posterior $q_{\phi}(z|x)$ and the intractable true posterior $p_{\theta}(z|x)$. Since KL divergence measures the distance between the probability distributions in positive manner, it can be said that $p(x)$ is lower bounded by ELBO and maximizing ELBO would indirectly maximizes the $p(x)$ so that $q_{\phi}(z|x)$ would approximate $p_{\theta}(z|x)$ when converged. Here, ELBO can be estimated by Stochastic Gradient Variational Bayes (SGVB) estimator which utilizes the reparameterization trick on latent variable. Then, Auto Encoding Variational Bayes (AEVB) algorithm can be applied with SGVB estimator which updates parameters in a mini-batch unit. Therefore, it can be summarized that SGVB plays a role in approximate inference and AEVB plays a role in efficient inference.

VAE is a variant of Autoencoder where the encoder and decoder outputs are the parameters of corresponding probability distributions to model the stochasticity. Mostly, it is assumed that z is supposed to follow the gaussian distribution with 0 mean and identity covariance matrix. This leads to regularization effect on learned posterior $q_{\phi}(z|x)$ so that each of the latent dimension can extract one factor of variation of data. As a result, in the experiment, when visualizing the data manifold, it was observed that similar data were concentrated in particular regions even if the model is trained in unsupervised setting. Moreover, it was shown that comparing to previous learning algorithm like Wake-Sleep or Monte Carlo-EM, AEVB shows considerably faster convergence in MNIST and FreyFace dataset which further verifies its validity for performing efficient and approximate inference.

2. Discussion

Here I would like to offer 3 discussion points. To begin with, what problem may occur when maximizing ELBO? There are two terms in ELBO which are in charge of reconstruction and regularization. Since the regularization term can be computed in relatively small number of parameters, it is vulnerable to posterior collapse. This can be resolved by using some training tricks such as KL annealing. Next, how can we better encourage the disentanglement among latent dimensions? Beta VAE is devised for such issue by introducing an additional hyperparameter beta which popped up when solving the proposed constrained optimization problem with Lagrange Theory. Finally, how can we better encourage the expressiveness of latent representation using auxiliary data? Conditional-VAE (or Improved Conditional-VAE) uses the auxiliary data as an additional input to encoder and decoder network and GPP -VAE is newly proposed novel model that further update parameters from the standard VAE by utilizing gaussian process prior on latent variable. Even if it requires several relaxations for computational ease, the predictive posterior can be utilized which is useful for out-of-sample prediction.