Inducing selective posterior-collapse in variational autoencoders by modelling the latent space as a mixture of normal-gamma distributions

Jouffroy Emma, Berthoumieu Yannick, Bach Olivier, Giremus Audrey

Commissariat à l'Energie Atomique et aux energies alternatives (CEA)
Laboratoire de l'Intégration du Matériau au Système (IMS)

Variational autoencoders are deep generative networks widely used for a large area of tasks, such as image or text generation. They are composed of two sub-models. On the one hand, the encoder aims to infer the parameters of the approximate posterior distribution $\mathcal{N}(z; x, \mu(\phi), \sigma(\phi))$ of a low dimensional latent vector z that represents the generative factors of the input data x. On the other hand, the decoder is intended to model the likelihood of the input data $\mathcal{N}(x; z, \mu(\theta), \sigma(\theta))$. The parameters of the model, ϕ, θ are jointly optimized through the maximization of a lower bound of the evidence called the ELBO. A major challenge is to get a disentangled and interpretable latent space in the aim to improve the field of representation learning. However, the vanilla variational autoencoder suffers from many issues, such as entangled latent space and posterior-collapse. These problems are all the more accentuated so as the dimension of the latent space is not well chosen. The goal of our work is to propose a model able to infer a disentangled latent space by taking advantage of a selective posterior-collapse process. Indeed, it can be observed that the variances inferred by the encoder for each latent variable have very different values depending on the information carried by the latter. More precisely, variables that contain a lot of information about the data distribution tend to have a low inferred variance contrary to the others. To leverage this property, we propose a variational autoencoder model which is favored to locate the information in a reduced number of latent variables and not to use the others. In this way, the dimension of the latent space is automatically adjusted to the complexity of the data. In order to do this, the latent variables of the autoencoder are augmented with their inverse variances which are also assumed unknown. Their joint posterior distribution is defined as a mixture of normal-gamma probability density functions $p_i NormalGamma(z_i, \lambda_i; \mu_i, \alpha_1, \beta_2) + (1 - p_i) NormalGamma(z_i, \lambda_i; \mu_i, \alpha_2, \beta_2)$, where for the $i^t exth$ latent variable z_i, λ_i stands for its inverse variance and μ_i is directly inferred by the encoder as well as p_i. The other hyperparameters are defined so that the inverse variances take high values when the encoded variable carries information and are close to 1 otherwise. In this way, we add prior information that fits our assumptions and force the model to encode information in a subset of the latent space by doing a "selective posterior-collapse". To optimize the parameters ϕ, θ, the objective function has to be modified to take into account the model mixture distribution, such as $ELBO_{NG} = E_{q_{\phi(z,\lambda|x)}}[log p_\theta(x|z, \lambda)] - KL[q_\phi(z, \lambda|x)||p(z)p(\lambda)]$ where λ is a vector that gathers all the inverse variances. A reparametrization trick is also proposed for the stochastic vectors λ and z in order to use the algorithm of stochastic gradient descent for the optimization. Our model, the Normal-Gamma VAE (NG-VAE), was tested on datasets with known factors of generation. We set the latent space dimension as highly superior to the number of these factors and validated the selective posterior-collapse process and disentanglement of the latent variables.