Improved Variational Inference with Inverse Autoregressive Flow

Advancements in the framework og normalizing flow have enabled flexible variational inferesnce of complex posteriors over latent variables. The work extends the setup of normalizing flows by proposing Inverse Autoregressive Flow (IAF). IAF makes use of autoregreessive updates in inverse flow transformations which enables scaling of inference to high dimensional latent spaces. Each transformation is based on an autoregressive neural network. Experiments demonstrate the suitability of IAF on MNIST and CIFAR10 which further makes use of a novel ResNet Variational Autoencoder (VAE) architecture consisting of bidirectional inference.

Normalizing flow consist of invertible transformations which enable a simple tensor with known density to be transformed into a ne tensor with complex density function that is still easy to compute. Based on the framework of normalizing flows, the work presents IAF which extends the setup of invertible transformations using autoregressive updates. IAF makes use of Gaussian versions of autoregressive encoders such as MADE and PixelCNN. The encoder yields the mean and variance corresponding to the Gaussian distribution which is transformed using a primary invertible transformation. Following the primary transformation, the flow consists of a chain of invertible transformations consisting of autoregressive neural networks. Corresponding to each step, the neural network yields mean and variance using autoregressive constraints which enforce the jacobians w.r.t latent variables to be lower triangular. The triangular structure of the jacobians further consists of zeros on the diagonal which provision the jacobian between latent variables to have variance values on the diagonal. Following the autoregressive constraints, the latent variable and posterior are updated as per the resulting mean and variance.

IAF demonstrates performance comparable to state-of-the-art methods in computation of the true likelihood on the MNIST benchmark. In the case of CIFAR10, IAF reduces the number of bits per dimension on test set by making use of a novel ResNet VAE architecture. While IAF depicts promise as a suitable scheme for variational inference, it provides room for explaination on a few of its aspects. Firstly, the work does not throw light on the usage of a primary encoding transformation and its necessity in the method. Since the IAF step demonstrates accurate estimation of the posterior, it must be sufficient for yielding suitable transformations. A detailed comparison of the scheme with and without IAF would throw light on the extent of suitability of the proposed method. Moreover, the scheme depicts suitability on only 2 benchmarks consisting of 10 classes each which could be further improved to tackle high dimensional benchmarks such as ImageNet.

Provision of autoregressive updates as invertible flow transformations enables accurate estimation of the posterior. The work demonstrates promise for extending the IAF scheme towards high dimensional latent spaces. The evaluation setup of IAF can be further improved to assess limitations and potential break points of autoregressive flows. Furthermore, the scheme itslef can be improved upon employing novel encoding methods and attention mechanisms which would facilitate extraction of richer representations in latent space.

Normalizing flows provide a suitable framework for variational inference onver complex posterior distributions consisting of latent variables. The work extends the setup of normalizinf flows and proposes IAF which makes use of autoregressive updates enforced in the neural networks. Invertible transformations consist of lower triangular jacobians which enable autoregressive updates allowing the suitability of scheme towards high dimensional latent spaces. In comparison to previous state-of-the-art models, IAF depicts accurate estimation of posterior on MNIST and lower likelihood on

CIFAR10 benchmarks.