On Variational Bounds of Mutual Information

A number of machine learning problems make use of Mutual Information (MI) as a metric for estimating the relationship between variables. However, bounding MI is a challenging task. Variational bounds allow tractabe objectives which are scalable to high dimensional tasks. The work presents a unified framework for variational bounds which consists of existing lower and upper bounds on MI. The bounds exhibit high bias or variance when MI is large. To that end, the framework is improved by providing a continuum on lower bounds which trades off bias and variance. Suitability of bounds is assessed on high dimensional control problems and representation learning.

Variational bounds provide tractability of objectives and improved approximations in the case f parameterized distributions. The work reviews existing bounds in literature by accumulating variational objectives in a unified framework. The framework consists of bounds which maximize or minimize MI pertaining to the auxilary objective. In order to improve the performance of bounds with large values of MI, a continuum of multi-sampple bounds is proposed which reduces variance by sample dependence. The multi-sample unnormalized setup consists of samples from $p(x_1)p(y|x_1)$ and access to $K-1$ additional samples which are used to estimate MI $I(X;Y)$. The bounds proposed in the work make use of an interpolation scheme which is upper bounded by $\log(\frac{K}{\alpha})$ where α is a hyperparameter which allows in the trade off between bias and variance. The multi-sample InfoNCE bound is obtained as a lower bound the MI $I(X;Y)$. Validation of the bound is carried out on 2 correlated Gaussian problems with the multi-sample bound demonstrating lower variance.

The multi-sample InfoNCE bound trades off variance with bias for the correlated Gaussian problem. Additionally, the utility of bounds for representation learning is highlighte din the case of position and scale on the dSpirtes dataset. However, the bound presents two shortcomings. Firstly, InfoNCE objective demonstrates saturation of estimates at $\log(batch - size)$. This hinders the scalability of the bound to problems consisting of larger batch sizes. Secondly, the I_α metric outperforms other bounds in terms of bias and variance and results in a reduced MSE in the InfoMAX objective. This presents InfoMAX as a suitable candidate for multi-sample inpterpolation. However, the work does not throw light on this aspect.

Multi-sample interpolated bounds allow the objective function to be controlled as per the nature of problem setting. However, their application is limited to optimal MI and batch sizes. The interpolation avenue presents two possible directions for future work. Firstly, the multi-sample framework can be extended to I_α in the InfoMAX setting which may provide insights into improved bias and variance as per the structure of the problem. Secondly, the InfoNCE lower bound on MI can be sharpened using generalization bounds which would aid in improved estimation but may hinder tractability in the case of parameteric settings.

Variational bounds provide tractable information theoretic objectives wherein the goal is to learn a conditional distribution. The work presents a unified framework for variational bounds based on pre-existing bounds. The framework consists of upper and lower bounds on MI. High variance of unnormalized bounds poses a hindrance in MI estimation. To that end, the work proposes a continuum of multi-sample unnormalized bounds which trade off bias and variance. The proposed objectives make use of nonlinear interpolation samples which are bounded by the number of additional samples and tractable in the log of batch size. Empirical evaluation of the multi-sample approach depicts suitable trade off between bias and variance on correlated Gaussians. On the other hand, suitability of variational bounds is validated in decoder-free representation learning.