Summary 18: Auto-Encoding Variational Bayes

Susan Cherry

This paper introduces a stochastic variational inference and learning algorithm. First, it shows that a re-parameterization of the variational lower bound gives a lower bound estimator that can be optimized using stochastic gradient descent. Then it shows that for i.i.d. datasets with continuous latent variables per datapoint, posterior inference can be made efficient by fitting an approximate inference model using the lower bound estimator. The Variational Bayes Method is an alternative to the mean-field approach that is commonly used. The mean field approach requires analytical solutions of the expectation with respect to the approximate posterior.

The paper begins by discussing the strategy used for deriving the lower bound estimator. The authors begin by discussing the problem scenario. They want an algorithm that works efficiently in the case of both intractability and large datasets. To do this, they propose to use efficient approximate ML and MAP estimation for the parameters θ, efficient approximate posterior inference of the latent variables, and efficient approximate marginal inference of the variables x. They specify the variational bound and then introduce a practical estimator of the lower bound and its derivatives with respect to the parameters. On page 4, they outline the Algorithm for the Minibatch version of the Auto-Encoding VB algorithm. They specify the three basic approaches for choosing a differentiable transformation $g_\phi(.)$ and auxiliary variable ϵ. The first is a tractable inverse CDF, the second was analogous to the Gaussian example, and the third was composition of different transformations of auxiliary variables.

Next, the give the Variational Auto-Encoder as an example of where they use a neural net for $g_\phi(.)$ and where the parameters ϕ and θ are optimized jointly with the AEVB algorithm. Then they move to experiments. They train generative models of the images from the MNIST and Frey Face datasets. They compare the variational lower bound of the learning algorithms and the estimated marginal likelihood. They use the generative model and variational approximation described in earlier sections. The Frey Face data are continuous, so they used a decoder with Gaussian output. This was identical to the encoder except that the means were contained in the interval (0,1) using a sigmoid activation function. Figure 2 shows the comparisons of lower bounds. They find that superfluous latent variables don't result in overfitting. Figure 3 shows results for the Marginal Likelihood. They compare the convergence speed for the three algorithms for small and large training set sizes.

In summary, this paper introduces a new estimator of the variational lower bound. It describes Stochastic Gradient VB which can be used for efficient approximate inference with continuous latent variables. It can be optimized using stochastic gradient methods. For i.i.d. datasets and continuous latent variables, they describe the AEVB, which learns an approximate inference model using the Stochastic Gradient VB estimator described earlier. They conclude by providing experimental results that provide additional support for the theoretical advantages of SGVB. They finally end the paper by suggesting avenues for future work.