Midterm Report - Variational Inference

Shubham Gupta
Department of Computer Science and Automation
Indian Institute of Science
Bangalore, India 560012
shubham.gupta@csa.iisc.ernet.in

Abstract

Inference in probabilistic models is one of the fundamental tasks. Efficient approximate inference has been the goal of statistics community, since one can show that the problem of probabilistic inference is NP-hard in general. Many methods exist for this task and Variational Inference is one such method. The aim of this project is to explore Variational Inference. The methodology we adopt is that of learning by doing where we take toy examples on synthetically generated data to demonstrate concepts. In the end we also give a full size real life application of Variational Inference, that of training a Variational Autoencoder.

1 Introduction

The GitHub repository which contains the code and other resources for this project can be found at:

https://github.com/sh-gupta/VariationalInference

The aim of this project is to explore the idea of Variational Inference. Most of the work done in this project will be primarily based on/extend the work of Blei et al. [1]. In the end we will employ Variational Inference to solve a real life problem at scale - that of generative modeling in images using Variational Autoencoder [2].

The philosophy is to first demonstrate the concepts using simple examples based on synthetically generated datasets. We will derive the theoretical results from scratch as done in [1]. To understand the theoretical results better we will obtain visualizations based on toy examples. Next we will focus on some applications of Variational Inference and take the example of Variational Autoencoder as a case study to see Variational Inference in action on a full scale real life problem as done in [2].

2 Brief Summary of Blei et al. (2016)

The paper serves as an introduction to Variational Inference and has been written in the form of a tutorial. The authors remark that although Variational Inference has been around for a long time [3] and has been applied for solving many problems successfully [2][4], the theoretical understanding of Variational Inference is still lacking. The authors intend that this paper will catalyze the research in Variational Inference.

Variational Inference is a method for performing approximate inference in graphical models. The core idea is to convert the problem of inference into an optimization problem. There are two main benefits involved here:

1. Optimization is a well studied problem and hence we have very efficient tools to do it
2. It is relatively easy to scale up to large datasets

The paper derives the optimization problem associated with Variational Inference from scratch. The concept of Evidence Lower Bound Objective (ELBO) is introduced and it is shown how the maximization of ELBO corresponds to the maximization of log probability of data.

The paper specifically focuses on Mean-Field approach for Variational Inference and outlines the Coordinate Ascent Variational Inference (CAVI) algorithm for solving the optimization problem. Most of the concepts in the paper are illustrated on Gaussian Mixture Models although we will also look into other models like simple Gaussian Model in this project.

A major research area in the domain of Variational Inference is proving bounds and guarantees related to convergence. As opposed to the traditional Markov Chain Monte Carlo (MCMC) methods, which have a lot of theoretical results, we do not have a sound theoretical understanding of Variational Inference. The paper compares Variational Inference with MCMC methods.

Since Variational Inference is an emerging field of study, there is still a lot left to be explored. For example, currently KL-Divergence is the most widely used measure of the quality of approximation for Variational Inference. One can explore other methods and alternative ways to define the optimization problem. The paper outlines many different open questions in the field.

Since the paper is written by the author of Latent Dirichlet Allocation (LDA) model, it takes LDA as a case study to demonstrate the power of Variational Inference on a real life large scale problem. LDA is concerned with finding topics from unsupervised text data. In this project we will briefly look at the LDA model although for the purpose of a detailed case study we will implement a Variational Autoencoder.

3 Brief Summary of Kingma et al. (2013)

This is the paper which rejuvenated the interest of Deep Learning community in Variational Inference. It applies Variational Inference to train a generative model for images. It uses MNIST [5] as training data and trains a special type of autoencoder called Variational Autoencoder to generate handwritten digits similar to those in MNIST.

The paper shows how Variational Inference can be applied in practice to a large scale model. There are a lot of practical issues involved here. For example, one has to employ what is called a re-parametrization trick to make the cost function differentiable in order to support end-to-end training.

We use this paper as our final case study showing the application of Variational Inference on a full scale problem on real data. We will implement the Variational Autoencoder and present the results for MNIST dataset as done in the paper. Further we will also try to obtain results on the CIFAR-10 dataset [6].

4 Dataset Description

While for most of the examples we will use synthetically generated datasets as described in the iPython Notebook given in the code repository, for the final case study, we will use the following two standard datasets.

4.1 MNIST

MNIST is a dataset of handwritten digits. Each image is a grayscale image of size 28×28. A single digit (0-9), appropriately centered and scaled, appears in each of the image. The digit is drawn on a white background using black ink. MNIST is one of simplest and cleanest computer vision related dataset. It contains $60,000$ training images along with the class labels and $10,000$ test images. There are equal number of images for each digit in both training and test set.

4.2 CIFAR-10

CIFAR-10 is a dataset of $60,000$ colored images belonging to 10 different classes along with class labels. Each images is of size 32×32. The classes include airplanes, cat, frog, ship etc. The images have been taken in natural setting. There are 6000 images per class.

5 Proposed Experiments

We propose to do the following experiments:

1. **Toy Example 1**: Application of Variational Inference to estimate the parameters of a Gaussian Distribution
2. **Toy Example 2**: Application of Variational Inference to infer the parameters for a Gaussian Mixture Model
3. **Intermediate Example 1**: Approximating expectation using a single sample in Toy Example 1
4. **Intermediate Example 2**: Comparison between Variational Inference and MCMC using synthetic data
5. **LDA**: A brief overview of LDA, no implementation
6. **Case Study**: Implementation of Variational Autoencoder and demonstration of performance on MNIST and CIFAR-10

Visualizations will be produced for each experiment as per the need. Toy Example 1 also includes comparison with the Maximum Likelihood approach. Other experiments may be added. Apart from these experiments some theoretical issues will also be explored.

6 Tentative Timeline

1. **Toy Example 1**: Done (see iPython Notebook on GitHub for results)
2. **Toy Example 2**: By March 26, 2017
3. **Intermediate Example 1**: By March 31, 2017
4. **Intermediate Example 2**: By April 5, 2017
5. **LDA**: By April 7, 2017
6. **Case Study**: By April 10, 2017
7. **Writing Final Report**: By April 15, 2017

Theoretical work will be done in between the experiments wherever necessary.

References

[1] Blei, D.M., Kucukelbir, A. & McAuliffe, J.D. (2016). Variational Inference: A Review for Statisticians. 2016arXiv160100670B.

[2] Kingma, D.P & Welling, M. (2013). Auto-Encoding Variational Bayes. 2013arXiv1312.6114K.

[3] Jordan, M. I., Ghahramani, Z., Jaakkola, T., & Saul, L. (1999). Introduction to variational methods for graphical models. Machine Learning, 37:183233.

[4] Blei, D., Ng, A., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3:9931022.

[5] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86(11):2278-2324

[6] Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images