Black box variational inference as a solution for the smoothing problem: a comparison of state of the art machine learning methods.

September 27, 2017

The smoothing problem is a common problem where the main objective is to estimate an unknown state of a time series system with help of some measurements. In other words, we have a model given by

$$\begin{cases} p_\theta(\mathbf{z}_t|\mathbf{z}_{t-1}) \\ p_\theta(\mathbf{x}_t|\mathbf{z}_t) \end{cases} \implies$$

$$p_\theta(\mathbf{x},\mathbf{z}) = p_\theta(\mathbf{z}_0)\Big[\prod_{t=1}^{T} p_\theta(\mathbf{z}_t|\mathbf{z}_{t-1})\Big]\Big[\prod_{t=1}^{T} p_\theta(\mathbf{x}_t|\mathbf{z}_t)\Big],$$

and we are trying to find a posterior

$$p_\theta(\mathbf{z}|\mathbf{x}) = \frac{p_\theta(\mathbf{x},\mathbf{z})}{p_\theta(\mathbf{x})}$$

where $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_t, ..., \mathbf{z}_T)$ is a vector containing the unknown states up untill time t, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t, ..., \mathbf{x}_T)$ is a similar vector containing the measurements and θ are the parameters used in the model. For linear systems with Gaussian noise the optimal estimation is known as the Kalman filter. For systems that aren't linear with Gaussian noise there are different methods for finding an estimate of the state $\mathbf{x}$. One new method is by using black box variational inference. The general idea is to approximate the posterior $p_\theta(\mathbf{z}|\mathbf{x})$ with $q_\phi(\mathbf{z}|\mathbf{x})$ and optimizing the evidence lower bound, or ELBO

$$\mathcal{L}(\theta;\phi;\mathbf{x}) = \mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\big[-\log(q_\phi(\mathbf{z}|\mathbf{x})) + \log(p_\theta(\mathbf{x},\mathbf{z}))\big].$$

One convenient choice of q is a multivariate Gaussian distribution,

$$q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}\big(\mu_\phi(\mathbf{x}), \Sigma_\phi(\mathbf{x})\big).$$

It is possible to then estimate the parameters $\mu_\theta(\mathbf{x}_t)$ and $\Sigma_\theta(\mathbf{x}_t)$ using neural networks, one for $\mu_\theta(\mathbf{x})$ and several for different parts of $\Sigma_\theta(\mathbf{x})$. [1]

In this project the aim is to implement the method presented by Archer et al. [1] and compare it to an alternative method by doing simulation studies for the different methods. To do this project some background knowledge in machine learning (such as the course Statistical Machine Learning) and an interest in the theoretical parts of machine learning is needed.

References

[1] Evan Archer, Il Memming Park, Lars Buesing, John Cunningham, and Liam Paninski. Black box variational inference for space state models. *ICLR*, 2016.