Lecture 28: Variational Inference

Professor Ilias Bilionis

Overview of variational inference

PREDICTIVE
SCIENCE LABORATORY

Journal of Machine Learning Research 18 (2017) 1-45 Submitted 3/16; Revised 8/16; Published 1/17

Automatic Differentiation Variational Inference

Alp Kucukelbir ALP@CS.COLUMBIA.EDU
Data Science Institute, Department of Computer Science
Columbia University
New York, NY 10027, USA

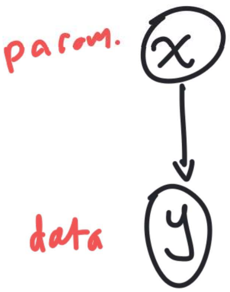
Dustin Tran DUSTIN@CS.COLUMBIA.EDU
Department of Computer Science
Columbia University
New York, NY 10027, USA

Rajesh Ranganath RAJESHR@CS.PRINCETON.EDU
Department of Computer Science
Princeton University
Princeton, NJ 08540, USA

Andrew Gelman GELMAN@STAT.COLUMBIA.EDU
Data Science Institute, Departments of Political Science and Statistics
Columbia University
New York, NY 10027, USA

David M. Blei DAVID.BLEI@COLUMBIA.EDU
Data Science Institute, Departments of Computer Science and Statistics
Columbia University
New York, NY 10027, USA

Variational Inference



<u>Prior</u> : $x \sim p(x)$

<u>Likelihood</u> : $y|x \sim p(y|x)$

<u>Posterior</u> : $\boxed{p(x|y) = \frac{p(y|x)\,p(x)}{Z} = \frac{p(y|x)\,p(x)}{p(y)}}$

$$Z \equiv p(y) \equiv \int p(y|x)\,p(x)\,dx$$

parameters

1) <u>Picks a candidate posterior</u> : $q(x;\phi)$

<u>guide</u>

parametric pdf

2) Find ϕ^* so that $q(x;\phi)$ matches $p(x|y)$.

$$\phi^* = \arg\min \mathrm{KL}\left(q(x;\phi)\,\|\,p(x|y)\right)$$

Kullback Leibler divergence

information loss incurred when replacing p w/ q.

What is the Kullback-Leibler divergence

$$KL\left(q(x;\varphi)\,\|\,p(x|y)\right) \equiv \int q(x;\varphi) \ln \frac{q(x;\varphi)}{p(x|y)}\,dx$$

$$= \mathbb{E}_{q(x;\varphi)}\left[\ln \frac{q(x;\varphi)}{p(x|y)}\right]$$

$$= \mathbb{E}_{q(x;\varphi)}\left[\ln q(x;\varphi) - \ln p(x|y)\right]$$

$$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$$

$$= \mathbb{E}_{q(x;\varphi)}\left[\ln q(x;\varphi) - \ln p(y|x) - \ln p(x) + \ln p(y)\right]$$

$$= \underbrace{\mathbb{E}_{q(x;\varphi)}\left[\ln \frac{q(x;\varphi)}{p(x)}\right]}_{\text{- entropy of } q(x;\varphi) \text{ relatively to prior } p(x)} - \underbrace{\mathbb{E}_{q(x;\varphi)}\left[\ln p(y|x)\right]}_{\text{max likelihood fits}} \cancel{+ \ln p(y)}$$

PREDICTIVE SCIENCE LABORATORY

Properties of the KL-divergence

not distance

1) $KL(q(x;\varphi) \| p(x|y)) \geq 0$

2) $KL(q(x;\varphi) \| p(x|y)) = 0 \Rightarrow q(x;\varphi) = p(x|y)$

3) $KL(p(x|y) \| q(x;\varphi)) \neq KL(q(x;\varphi) \| p(x|y))$

PREDICTIVE
SCIENCE LABORATORY

The evidence lower bound

$$KL(q(x;\varphi)\|p(x|y)) = \mathbb{E}_{q(x;\varphi)}[\ln q(x;\varphi)] - \mathbb{E}_{q(x;\varphi)}[\ln \underbrace{p(y|x)p(x)}_{p(x,y)}] + \ln p(y)$$

$$\geq 0$$

$$\Rightarrow \underbrace{\ln p(y)}_{\text{Evidence}} \geq \underbrace{-\mathbb{E}_{q(x;\varphi)}[\ln q(x;\varphi)] + \mathbb{E}_{q(x;\varphi)}[\ln p(x,y)]}_{\substack{\text{Evidence Lower Bound} \\ ELBO(\varphi)}}$$

$$\min_{\varphi} KL(q(x;\varphi)\|p(x|y)) \Longleftrightarrow \boxed{\max_{\varphi} ELBO(\varphi)}$$