Variational Optimization [1]

Xingdong Zuo

IDSIA, Switzerland

June 21, 2018

1 Introduction

Variational optimization (VO) is a technique to optimize a non-differentiable or discrete objective function with a differentiable bound on the optima. Let $x \in \mathcal{D} \subset \mathbb{R}^n$ be a vector in the domain and let $f(x)$ be a non-differentiable function, our goal is to maximize the function, i.e. $f^* = \max_{x\in\mathcal{D}} f(x)$. We cannot use traditional gradient ascent techniques due to non-differentiability of f. In VO, we could replace it with a differentiable surrogate objective by a lower bound on the optimum of $f(x)$

$$f^* = \max_{x\in\mathcal{D}} f(x) \geq \mathbb{E}_{x\sim p(x|\theta)}\left[f(x)\right] \tag{1}$$

where $p(x|\theta)$ is a probability distribution over $\mathcal{D}$ parameterized by θ.

2 Differentiability of the variational bound

We could apply REINFORCE technique to obtain an unbiased gradient estimator, i.e.

$$\nabla_\theta \mathbb{E}_{x\sim p(x|\theta)}\left[f(x)\right] = \mathbb{E}_{x\sim p(x|\theta)}\left[f(x)\nabla_\theta \log p(x|\theta)\right] \tag{2}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)\nabla_\theta \log p(x_i|\theta) \tag{3}$$

3 Concavity of the variational bound

Definition 3.1. *Let $I \subset \mathbb{R}$ be an interval in real line, a function $f : I \to \mathbb{R}$ is concave if $\forall x, y \in I$ and $\forall \alpha \in [0,1]$ such that*

$$f((1-\alpha)x + \alpha y) \geq (1-\alpha)f(x) + \alpha f(y). \tag{4}$$

Alternatively, one can rewrite it as

$$f(x + \alpha(y - x)) \geq f(x) + \alpha(f(y) - f(x)). \tag{5}$$

Definition 3.2. *A probability distribution $p(x|\theta)$ parameterized by θ is expectation affine if*

$$\mathbb{E}_{x\sim p(x|\theta)}\left[f(x)\right] = \mathbb{E}_{z\sim q(z)}\left[f(\alpha(\theta)z + \beta(\theta))\right] \tag{6}$$

where $\alpha(\theta), \beta(\theta)$ are linear functions and $q(z)$ is a probability distribution.

Theorem 3.1. *If a function $f(x)$ is concave and a probability distribution $p(x|\theta)$ parameterized by θ is expectation affine, then the variational lower bound $\mathbb{E}_{x\sim p(x|\theta)}\left[f(x)\right]$ is concave in θ.*

Proof. Let $\lambda \in [0, 1]$ be a scalar. We have

$$\begin{aligned}
\mathbb{E}_{x\sim p(x|\lambda\theta_1+(1-\lambda)\theta_2)}\left[f(x)\right] &= \mathbb{E}_{z\sim q(z)}\left[f(\alpha(\lambda\theta_1 + (1-\lambda)\theta_2)z + \beta(\lambda\theta_1 + (1-\lambda)\theta_2))\right] \\
&= \mathbb{E}_{z\sim q(z)}\left[f(\lambda(\alpha(\theta_1)z + \beta(\theta_1)) + (1-\lambda)(\alpha(\theta_2)z + \beta(\theta_2)))\right] \\
&\geq \mathbb{E}_{z\sim q(z)}\left[\lambda f(\alpha(\theta_1)z + \beta(\theta_1)) + (1-\lambda)f(\alpha(\theta_2)z + \beta(\theta_2))\right] \\
&= \lambda\mathbb{E}_{z\sim q(z)}\left[f(\alpha(\theta_1)z + \beta(\theta_1))\right] + (1-\lambda)\mathbb{E}_{z\sim q(z)}\left[f(\alpha(\theta_2)z + \beta(\theta_2))\right] \\
&= \lambda\mathbb{E}_{x\sim p(x|\theta_1)}\left[f(x)\right] + (1-\lambda)\mathbb{E}_{x\sim p(x|\theta_2)}\left[f(x)\right]
\end{aligned}$$

□

The Theorem 3.1 indicates that if $f(x)$ has unique global optimum, so does its variational lower bound.

4 Estimation of distribution algorithm

Given the optimization problem $f^* = \max_{x\in\mathcal{D}} f(x)$ where $f(x)$ is non-differentiable. We could instead optimize the variational lower bound $\mathbb{E}_{x\sim p(x|\theta)}\left[f(x)\right]$.

The general procedure is shown as following

1. Prior distribution $p_0(x|\theta_0)$
2. Iteratively
   (a) Generate solution set $\{x_n\}$ and evaluate them $\{f(x_n)\}$
   (b) Update distribution: $p(x|\theta_{i+1}) = F\left(p(x|\theta_i), \{x_n\}, \{f(x_n)\}\right)$ for some function F.

References

[1] Joe Staines and David Barber. "Variational optimization". In: *arXiv preprint arXiv:1212.4507* (2012).