Adversarial optimization

Maxim Borisyak

National Research University Higher School of Economics (HSE)

August 9, 2018

Adversarial

Example

Consider a task of tuning unknown parameters of the PYTHIA event generator to a particular set of data.

An approach

Event generator tuning using Bayesian optimization

Philip Ilten, Mike Williams, and Yunjie Yang

Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA 02139

- make two histogram for each parameter: data_i and MC_i;
- use Bayesian Optimization on the objective function:

$$\chi^2 = \sum_{i=1}^{n_{bins}} \frac{(\mathrm{data}_i - \mathrm{MC}_i)^2}{\sigma^2_{\mathrm{data},i} + \sigma^2_{\mathrm{MC},i}}$$

Adversarial

Notation:

- parameters θ of the PYTHIA define a distribution $p_\theta = p(\cdot \mid \theta)$ on events;
- p_{data}: real distribution;
- consider both distributions as intractable and can only be sampled from.

Adversarial objective can be used instead *:

$$\theta^* = \arg\min_\theta \ \text{Jensen-Shannon}(p_\theta, p_{\text{data}}) = \arg\max_\theta \min_f \left[\text{cross-entropy}_f(p_\theta, p_{\text{data}})\right]$$

* Any other statistical distance (e.g. Wasserstein) can be also used.

Why adversarial objective

- sufficiently powerful discriminator does not create 'fake' minima:

$$\text{Jensen-Shannon}(p_\theta, p_{\text{data}}) = 0 \iff p_\theta = p_{\text{data}}$$

- prior knowledge can be expressed via the choice of discriminator, e.g.:
    - architecture and regularization for neural networks;
    - feature engineering for tree-based algorithms.

Optimization

Black-box

Differences from GAN:

- non-differentiable generator;
- thus, **black-box optimization**;
- hence, discriminator can be non-differentiable as well (e.g. tree-based).

Bayesian Optimization example

initialize Bayesian Optimization

while patience is not ran out **do**

  $\theta \leftarrow \mathrm{askBO}()$

  $X^{\theta}_{\mathrm{train}}, X^{\theta}_{\mathrm{test}} \leftarrow \mathrm{sample}(\theta)$

  $f \leftarrow$ train discriminator on $X^{\theta}_{\mathrm{train}}$ and $X^{\mathrm{real}}_{\mathrm{train}}$

  $\mathcal{L} \leftarrow \frac{1}{2 \cdot m} \left[\sum_{i=1}^{m} \log f(X^{\theta,i}_{\mathrm{test}}) + \sum_{i=1}^{m} \log(1 - f(X^{\mathrm{real},i}_{\mathrm{test}})) \right]$

  $\mathrm{tellBO}(\theta, \log 2 - \mathcal{L})$

end while

Adversarial Variational Optimization

Motivation

Adversarial Variational Optimization of Non-Differentiable Simulators

Gilles Louppe G.LOUPPE@ULIEGE.BE
University of Liège, Belgium

Kyle Cranmer KYLE.CRANMER@NYU.EDU
New York University, USA

Exact match between generator and real data may not exist:

- search for solution as mixture of generators defined by $q(\theta \mid \psi)$:

$$x \sim p(x \mid \theta),\ \theta \sim q(\theta \mid \psi)$$

or

$$x \sim \phi(x \mid \psi)$$

Variational

The formal problem statement:

$$\mathcal{L} = \frac{1}{2}\left[\mathop{\mathbb{E}}_{x \sim \phi(x|\psi)} \log f(X) + \mathop{\mathbb{E}}_{X \sim p_{\mathrm{real}}} \log(1 - f(X))\right];$$
$$\psi^* = \arg\max_{\psi} \min_{f} \mathcal{L};$$

- x is now sampled from a compound distribution;
- optimization is done by distribution parameters ψ (and not by generator parameters θ).

Gradient estimation

$$\nabla_\psi \mathcal{L} = \nabla_\psi \frac{1}{2}\left[\underset{x \sim \phi(x|\psi)}{\mathbb{E}} \log f(X)\right] =$$

$$\frac{1}{2}\nabla_\psi \int_\theta \int_x d\theta\, dx\; p(x \mid \theta)\, q(\theta \mid \psi) \log f(X) =$$

$$\frac{1}{2} \int_\theta \int_x d\theta\, dx\; p(x \mid \theta) \nabla_\psi q(\theta \mid \psi) \log f(X) =$$

$$\frac{1}{2} \int_\theta \int_x d\theta\, dx\; p(x \mid \theta)\, q(\theta \mid \psi) \nabla_\psi \log q(\theta \mid \psi) \log f(X) =$$

$$\frac{1}{2} \underset{x \sim \phi(x|\psi)}{\mathbb{E}} \log f(X) \cdot \nabla_\psi \log q(\theta \mid \psi)$$

The math works almost exactly as in Variational Optimization:

- discriminator is trained to distinguish samples from $\phi(\cdot \mid \psi)$ not from individual generators;
- conventional VO applied to adversarial objective would converge to the single best generator.

Adversarial Variational Optimization

initialize $q(\cdot \mid \psi)$

while not bored **do**

  sample X_{train} from $\phi(x \mid \psi)$

  $f \leftarrow$ train discriminator on X_{train} and $X_{\text{train}}^{\text{real}}$

  $X_{\text{test}} \leftarrow$ sample from $\phi(x \mid \psi)$

  $\nabla_\psi \mathcal{L} \leftarrow \frac{1}{m} \sum_{i=1}^{m} \log f(X_{\text{test}}^{i}) \cdot \nabla_\psi \log q(\theta \mid \psi)$

  $\theta \leftarrow \text{Adam}(\nabla_\psi \mathcal{L})$

end while

Summary

Summary

- adversarial objective can be utilized for non-differentiable generators;
    - which allows to tune MC models to real data;
- it is possible to find a solution as a mixture of generators.

References

- Ilten, P., Williams, M. and Yang, Y., 2017. Event generator tuning using Bayesian optimization. Journal of Instrumentation, 12(04), p.P04028.
- Louppe, G. and Cranmer, K., 2017. Adversarial Variational Optimization of Non-Differentiable Simulators. arXiv preprint arXiv:1707.07113.