RELAX: General relaxations for unbiased gradient estimates in discrete latent-variable models

Anonymous Author(s)
Affiliation
Address
email

Abstract

The recently-developed REBAR method reduces the variance of the REINFORCE gradient estimator by adding a control variate based on the Gumbel-softmax relaxation. The free parameters of REBAR can be tuned by gradient descent to minimize variance. We generalize this method, showing that the particular relaxation used by REBAR can be generalized, and replaced by a neural network to be optimized during learning. We examine the properties of optimal relaxations, and show that this more general procedure enables faster training of discrete latent variable models.

1 Introduction

Contributions

- We show...

2 REINFORCE, Concrete, and REBAR gradient estimator

In this section we briefly review the REINFORCE, Concrete, and REBAR gradient estimators.

2.1 REINFORCE

For simplicity of exposition, consider a simple stochastic model that has a discrete latent variable b with probability $p_\theta(b)$ and loss function $f(b)$.

Note we assume without loss of generality that $f(b)$ is independent of the parameters θ of $p_\theta(b)$. Training the model involves minimizing the expected cost $\mathcal{L}(\theta) = \mathbb{E}_{p_\theta(b)}[f(b)]$. This can be achieved efficiently using gradient optimization, requiring exact or approximate computation of $g = \partial\mathcal{L}/\partial\theta$.

In many cases of interest, an analytic form of the gradient g is not known, and so an estimator $\hat{g}$ is required, such as the Monte Carlo estimator $\hat{g} \approx \sum_{i=1}^{n} \hat{g}(b_i)/n$ where $b_i \sim p_\theta(b)$ and n is the number of samples.

The REINFORCE gradient [5] expands g as $\mathbb{E}_{p_\theta(b)}[f(b)\frac{\partial}{\partial\theta}\log p_\theta(b)]$ using the log-derivative trick, and uses exactly a Monte Carlo estimation scheme. However, the derivative of the log-likelihood w.r.t. its parameters (known as the score function in statistics) has high variance under Monte Carlo estimation, and so variance-reduction techniques such as control variates or Rao-Blackwellization are usually applied to improve the speed of convergence to a good solution.

Submitted to 29th Conference on Neural Information Processing Systems (NIPS 2016). Do not distribute.

2.2 Concrete estimator

Another approach to estimating the gradient of a loss that involves discrete random variables is a continuous relaxation. [2] and [1] developed in parallel a reparameterization of a categorical distribution using the Gumbel-Max trick for sampling from a discrete distribution.

The reparameterization is intuitive to understand when considering each point in discrete distribution as one-hot vector. Since we want the gradient to be defined at all points, we can extend the distribution over one-hot vectors to allow it to take values in its convex hull, so that the support is maintained of the discrete points in the support of the distribution. The convex hull is the d-1 dimensional probability simplex in $\mathbb{R}^k$, where k is the dimension of the discrete distribution.

Let $G_{1:k} = -\log -\log(U_{i:k})$ be samples from the Gumbel distribution, and learnable parameters $(\alpha_1, \ldots, \alpha_k)$ be interpreted as some unnormalized parameterization of the discrete distribution under consideration. Then, consider the following sampling procedure: for each k, find the k that maximizes $\log \alpha_k - G_k$, and then set $D_k = 1$ and $D_{i \neq k} = 0$. The Gumbel-Max trick states that sampling from the discrete distribution is equivalent to taking this argmax, that is, $p(D_k = 1) = \alpha_k / \sum_{i=1}^{n} \alpha_i$.

Since taking an argmax is still a discontinuous operation, [2] and [1] proposed further relaxing the argmax operator through the softmax function with an additional temperature parameter λ:

$$x_k = \frac{\exp\{(\log \alpha_k + G_k)/\lambda\}}{\sum_{i=1}^{n} \exp\{(\log \alpha_i + G_i)/\lambda\}} \tag{1}$$

This relaxation allows values within the simplex, but in the low temperature limit, it becomes exactly the discrete argmax. One limitation of the concrete distribution is that it is a biased estimator except in limiting temperature. In other words, a small amount of bias is present for a non-zero temperature.

2.3 REBAR estimator

The REBAR gradient estimator develops a lower-variance gradient estimator that outperforms one based on a Concrete relaxation. REBAR relies on a carefully designed control variate, so we begin with a review of this theory.

2.3.1 Control Variates for Variance Reduction

For an unbiased estimator $f(b)$, a control variate is a function $\tilde{f}$ with a known or estimatable mean $\mathbb{E}[\tilde{f}]$. Since the mean of the function can be subtracted from the expectation, $f(b) - \eta(\tilde{f} - \mathbb{E}[\tilde{f}])$ remains an unbiased estimator. A control variate is scaled by a constant η. Note that we can write $\text{Var}(f(b) + \eta(\tilde{f} - \mathbb{E}[\tilde{f}]))$ as

$$\text{Var}(f(b) + \eta\tilde{f}) = \text{Var}(X) + \eta^2 \text{Var}(\tilde{f}) + 2\eta \text{Cov}(f(b), \tilde{f}), \tag{2}$$

which, evaluating the first derivative w.r.t. η and solving for zero yields:

$$\eta = -\frac{Cov(f(b), \tilde{f})}{Var(\tilde{f})}. \tag{3}$$

The variance reduction effect of a control variate is induced by the high correlation of the control variate with the original estimator. The intuition behind this is as follows. When $f(b)$ and $\tilde{f}$ are positively correlated, then this means $\tilde{(}f)$ is large when $f(b)$ is large. So, if in some minibatch $\tilde{f}$ is greater than its known mean, then with high probability $f(b)$ is also greater than its true mean. That means this minibatch would contribute variance to the overall estimation algorithm, since the values we're estimating of $f(b)$ are with high probability greater than the true mean (the true gradient, in the case of REBAR). With positive correlation, $\text{Cov}(f(b), \tilde{f})) > 0$, and so η is negative. Then, the effect of the control variate in such a minibatch is to reduce the REINFORCE estimate by subtracting the quantity $\eta(\tilde{f} - \mathbb{E}[\tilde{f}]$.

Hence, a suitably scaled control variate with high correlation to the original estimator dampens or amplifies the value of the overall estimator towards the true mean per minibatch. The same explanation applies to the case where the correlation is negative by swapping the terms "lesser" for "greater" and "reducing" by "increasing". This is why high correlation of $f(b)$ with $\tilde{f}$ is desirable:

it renders the control variate effective in drawing down the effect of high variability in a stochastic estimator by a magnitude commensurate with the difference of a highly correlated function and its known mean.

2.3.2 Reducing Gradient Variance through a Low-Variance Control Variate

Following [4], the following overview focuses on a single discrete Bernoulli random variable. The core of the REBAR estimator is a REINFORCE-style estimate of a non-differentiable reparameterization of the discrete latent variable as $b = H(z)$, where H is the hard-threshold function an

$$z := g(u, \theta) := \log \frac{\theta}{1-\theta} + \log \frac{u}{1-u}, u \sim \text{Unif}[0, 1] \tag{4}$$

While z is a reparameterization that renders the parameters of b learnable by gradient-based methods, H introduces a new discontinuity in the loss. Instead of relaxing the hard threshold as in [2], [4] uses a REINFORCE estimator for the gradient reparameterized with $H(z)$:

$$\frac{\partial}{\partial\theta}\mathbb{E}_{p(b)}[f(b)] = \frac{\partial}{\partial\theta}\mathbb{E}_{p(u)}[f(H(z))] = \mathbb{E}_{p(u)}[f(H(z))\frac{\partial}{\partial\theta}\log p(z)] \tag{5}$$

This allows the parameters θ to be learned using gradient information, but the loss is still non-differentiable due to the hard threshold. Since this uses a REINFORCE estimator, it also has high variance.

Hence, [4] develop a control variate. A natural continuous relaxation of the hard threshold function is the sigmoid function, leading [4] to choose a $H(z) \approx \sigma_\lambda(z) := \sigma(z/\lambda) = (1 + \exp(-z/\lambda))^{-1}$. This relaxation leads to the following control variate summed with its expectation:

$$\frac{\partial}{\partial\theta}\mathbb{E}_{p(z)}[f(\sigma_\lambda(z))] = \mathbb{E}_{p(z)}[f(\sigma_\lambda(z))\frac{\partial}{\partial\theta}\log p(z)]. \tag{6}$$

Unfortunately, simply applying this control variate was found to be ineffective. The author's key insight was to derive a low-variance form of this control variate that takes advantage of a conditional marginalization of the reparameterized z given a particular choice of discrete b. This introduces a second reparameterization $\tilde{z}$ of $p(z|b)$, which depends on another sample $v \sim \text{Unif}[0, 1]$. See [4] for details of the derivation.

The control variate has the following form:

$$f(\sigma_\lambda(\tilde{z}))\frac{\partial}{\partial\theta}\log p(H(z)) \tag{7}$$

and noting that

$$\mathbb{E}_{p(u,v)}[f(\sigma_\lambda(\tilde{z}))\frac{\partial}{\partial\theta}\log p(H(z))] = \mathbb{E}_{p(u,v)}[\frac{\partial}{\partial\theta}f(\sigma_\lambda(\tilde{z})) - \frac{\partial}{\partial\theta}f(\sigma_\lambda(z))] \tag{8}$$

gives us the REBAR gradient estimator:

$$\frac{\partial}{\partial\theta}\mathbb{E}_{p(b)}[f(b)] = \mathbb{E}_{p(u,v)}[f(\sigma_\lambda(\tilde{z}))\frac{\partial}{\partial\theta}\log p(H(z)) - \eta f(\sigma_\lambda(\tilde{z}))\frac{\partial}{\partial\theta}\log p(H(z)) + \eta\frac{\partial}{\partial\theta}f(\sigma_\lambda(z)) - \eta\frac{\partial}{\partial\theta}f(\sigma_\lambda(\tilde{z}))] \tag{9}$$

where η is trained to minimize the variance of the estimator.

The special form of $\tilde{z}$ yields a lower-variance gradient estimate because a number of the random variables are conditionally marginalized out of the estimator. Two features of this control variate make it particularly effective: its high correlation with the REINFORCE gradient, and a low-variance, reparameterized form of certain terms in the estimator.

3 The generalized REBAR estimator

The REBAR estimator uses a control variate that evaluates the original loss function at relaxed inputs, reparameterized both unconditionally (denoted z, and conditionally, denoted $\tilde{z}$). The central result of this paper is that learning the function in the control variate leads to even better convergence

properties. Specifically, we generalize the conditional marginalization and control variate of REBAR to the following form:

$$\mathbb{E}_{p(u,v)}[Q(\tilde{z};\phi)\frac{\partial}{\partial\theta}\log p(H(z))] = \mathbb{E}_{p(u,v)}[\frac{\partial}{\partial\theta}Q(\tilde{z};\phi) - \frac{\partial}{\partial\theta}Q(z;\phi)], \tag{10}$$

where Q is an MLP with parameters ϕ.

The generalized REBAR estimator replaces the loss function evaluations in the control variate with an adaptive Q function which is trained via gradient decent to minimize the variance of the estimator. As shown in [4], this can be easily computed. Denoting the RELAX estimator as $r(\phi)$ we obtain:

$$\frac{\partial}{\partial\phi}\text{Var}(r(\phi)) = \frac{\partial}{\partial\phi}\mathbb{E}[r(\phi)^2] + \frac{\partial}{\partial\phi}\mathbb{E}[r(\phi)]^2 = \frac{\partial}{\partial\phi}\mathbb{E}[r(\phi)^2] = \mathbb{E}[\frac{\partial}{\partial\phi}r(\phi)^2] \tag{11}$$

Where the second equality comes from the fact that for all ϕ, the RELAX estimator is unbiased and therefore $\frac{\partial}{\partial\phi}\mathbb{E}[r(\phi)]^2 = 0$.

** TODO: flesh out more or present toy examples

4 Properties of Optimal Relaxations

In [4] the concrete distribution [2] is used in the control variate due to its similarity to the Bernoulli distribution and assumed correlation with the target function (**TODO EXPAND THIS).

NEXT: PROVE THAT CONCRETE IS NOT THE OPTIMAL RELAXATION TO USE (SHOULD BE EASY TO PROVE BUT WE NEED TO DO IT)

NEXT: OPTIMAL RELAXATION FOR TOY PROBLEM NEXT: ELABORATE ON DEPENDENCE ON THETA

**TODO NEW EXPERIMENT: HALFWAY THROUGH A VAE TRAINING STOP, AND OPTIMIZE THE VARINACE OF Q AND OF REBAR WRT TEMP AND ETA AND COMPARE GRADIENTS

5 Understanding the REBAR estimator through the Generalized Reparameterization Gradient

REBAR and the generalization in this paper uses a mixture of score function and reparameterization gradients. A recent paper by [3] unifies these two gradient estimators as the generalized reparameterization gradient (GRG). This framework can help disentangle the various components of generalized REBAR.

5.1 Combining score function and reparameterization gradients

Generalizing the reparameterization trick

Write sample from distribution $s(\epsilon)$ as $\epsilon = \mathcal{T}^{-1}(\mathbf{z};\nu)$ for some invertible transform $\mathcal{T}$ with variational parameters ν. write out transformed density example: normal with standard normal s example: inverse CDF of Gaussian with uniform s write out expected gradient under transformation show decomposition of expected gradient into reparameterization and correction terms

Applying GRG to REBAR

show mapping of terms note denser derivation in REBAR appendix

Interpreting REBAR through GRG

REBAR innovation as further decomposition the correction term into secondary reparameterization components note this is a recursive application of the principles of GRG observe that the GRG suggests this recursive application to components of an estimator propose that other estimators could be similarly recursively decomposed?

[TODO: cite certigrad arxiv 1706.08605]

6 Experiments

6.1 Toy example

References

[1] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.

[2] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*, 2016.

[3] Francisco R Ruiz, Michalis Titsias RC AUEB, and David Blei. The generalized reparameterization gradient. In *Advances in Neural Information Processing Systems*, pages 460–468, 2016.

[4] George Tucker, Andriy Mnih, Chris J Maddison, and Jascha Sohl-Dickstein. Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models. *arXiv preprint arXiv:1703.07370*, 2017.

[5] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.

7 Appendix A: Control Variates