ARSM: Augment-REINFORCE-Swap-Merge Estimator for Gradient Backpropagation Through Categorical Variables

Mingzhang Yin*, Yuguang Yue*, Mingyuan Zhou

The University of Texas at Austin
Department of Statistics and Data Sciences
IROM Department, McCombs School of Business

International Conference on Machine Learning
Long Beach, CA, June 13, 2019

Categorical latent variable optimization

- **Goal**: Maximize the expectation with respect to categorical variables

$$\mathcal{E}(\phi) = \int f(\boldsymbol{z}) q_\phi(\boldsymbol{z}) d\boldsymbol{z} = \mathbb{E}_{\boldsymbol{z} \sim q_\phi(\boldsymbol{z})}[f(\boldsymbol{z})]$$

- **Notations:**
  1. $f(\boldsymbol{z})$ is the reward function for categorical $\boldsymbol{z}$
  2. $\boldsymbol{z} = (z_1, \dots, z_K) \in \{1, 2, \dots, C\}^K$ is a K-dimensional C-way multivariate categorical vector
  3. $q_\phi(\boldsymbol{z}) = \prod_{k=1}^K \text{Categorical}(z_k; \sigma(\phi_k))$ is the categorical distribution whose parameters $\phi \in \mathbb{R}^{KC}$ needs to be optimized
- **Challenge:** It is difficult to estimate

$$\nabla_\phi \mathcal{E}(\phi)$$

especially for large K and C.

Derivation of ARSM

- **Augment:** the categorical variable $z \sim \text{Cat}(\sigma(\phi))$ can be equivalently generated as
$$z = \underset{i \in \{1,\dots,C\}}{\arg\min}\ \pi_i e^{-\phi_i}, \ \boldsymbol{\pi} \sim \text{Dir}(\mathbf{1}_C).$$
Thus $\mathcal{E}(\phi) = \mathbb{E}_{\boldsymbol{z} \sim q_\phi(\boldsymbol{z})}[f(\boldsymbol{z})] = \mathbb{E}_{\boldsymbol{\pi} \sim \text{Dir}(\mathbf{1}_C)}[f(\arg\min_i \pi_i e^{-\phi_i})]$.
- **REINFORCE:**
$$\nabla_\phi \mathcal{E}(\phi) = \mathbb{E}_{\boldsymbol{\pi} \sim \text{Dir}(\mathbf{1}_C)}[f(\arg\min_i \pi_i e^{-\phi_i})(1 - C\boldsymbol{\pi})]$$
- **Swap:** Swapping the i^{th} and j^{th} elements of $\boldsymbol{\pi}$ would not change the expectation, which is a property used to provide self-controlled variance reduction (without any tuning parameters).
- **Merge**: Sharing random numbers between differently expressed but equivalent expectations leads to $\nabla_{\phi_c} \mathcal{E}(\phi) = \mathbb{E}_{\pi \sim \text{Dir}(\mathbf{1}_C)}[g_{\text{ARSM}}(\boldsymbol{\pi})_c]$
$$g_{\text{ARSM}}(\boldsymbol{\pi})_c := \frac{1}{C} \sum_{j=1}^{C} \left[f(z^{c \leftrightharpoons j}) - \frac{1}{C} \sum_{m=1}^{C} f(z^{m \leftrightharpoons j}) \right] (1 - C\pi_j)$$

An illustration example

Optimize $\phi \in \mathbb{R}^C$ to maximize $\mathbb{E}_{z\sim \text{Cat}(\sigma(\phi))}[f(z)], \quad f(z) := 0.5 + z/(CR)$

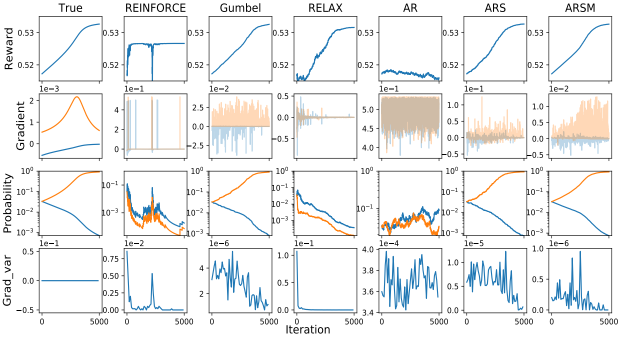

Figure: The optimal solution is $\sigma(\phi) = (0, \ldots, 1)$. The reward is computed analytically by $\mathbb{E}_{z\sim \text{Cat}(\sigma(\phi))}[f(z)]$ with maximum as 0.533.

VAEs with one or two categorical hidden layers (20-dimensional 10-way categorical)

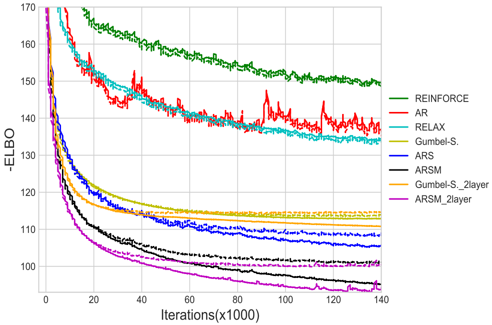

Figure: Plots of negative ELBOs (nats) on binarized MNIST against training iterations. The solid and dash lines correspond to the training and testing respectively.

Reinforcement Learning (a sequence of categorical actions)

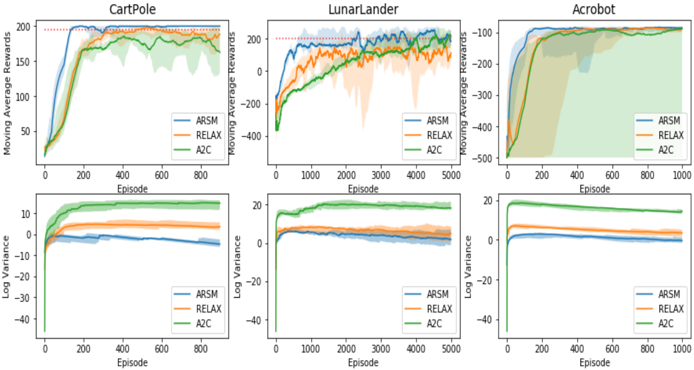

Figure: Moving average reward and log-variance of gradient estimator. In each plot, the solid lines are the median value of ten independent runs. The opaque bars are 10th and 90th percentiles.

Thank you!

Welcome to our poster at Pacific Ballroom #85