Variational Information Maximizing Exploration

Xinghu Yao

October 25, 2018

1 Optimize Variational Posterior

1.1 Objective Function

Minimizing $D_{\mathbb{KL}}[q(\theta;\phi)\|p(\theta|D)]$ can be done as follows

$$\begin{aligned}\phi^{'} &= \underset{\phi}{\operatorname{argmin}}\, D_{\mathbb{KL}}[q(\theta;\phi)\|p(\theta|D)] \\ &= \underset{\phi}{\operatorname{argmin}}\, -\mathrm{ELBO} \\ &= \underset{\phi}{\operatorname{argmin}} \int_{\theta\sim q(\theta;\phi)} q(\theta;\phi) log \frac{q(\theta;\phi)}{p(\theta)p(D|\theta)} \\ &= \underset{\phi}{\operatorname{argmin}}\, D_{\mathbb{KL}}[q(\theta;\phi)\|p(\theta)] - \mathbb{E}_{\theta\sim q(\theta;\phi)}[logp(D|\theta)] \end{aligned} \tag{1}$$

For simplicity we shall denote it as:

$$F(D;\theta) = D_{\mathbb{KL}}[q(\theta;\phi)\|p(\theta)] - \mathbb{E}_{\theta\sim q(\theta;\phi)}[logp(D|\theta)] \tag{2}$$

We can approximate this exact cost as:

$$F(D;\theta) \simeq \frac{1}{N}\sum_{i=1}^{N} logq(\theta^{(i)};\phi) - logp(\theta^{(i)}) - logp(D|\theta^{(i)}) \tag{3}$$

Let us now look at each of these single terms individually.

1.2 Likelihood

We use the softmax function to define our likelihood $p(D_i|\theta)$. So, we have the following log-likelihood function

$$\log p(D_i|\theta) = \log \frac{e^{\theta_i}}{\sum_{j=1}^{n} e^{\theta_j}} \tag{4}$$

Thus the log-likelihood term can be seen as locally linear and the Hessian matrix can be ignored to speed the computation.

1.3 Gaussian prior

A popular and simple prior is the Gaussian distribution. The prior over the entire weight vector simply corresponds to the product of the individual Gaussians:

$$p(\theta) = \prod_{i=1}^{\Theta} \mathcal{N}(\theta_i|\mu_i,\sigma_i) \tag{5}$$

And the log-prior $\log p(w)$ can be expressed as follows:

$$\log p(\theta) = \sum_{i=1}^{\Theta} \log \mathcal{N}(\theta_i|\mu_i,\sigma_i) \tag{6}$$

1.4 Variational Posterior

The variational posterior on the weights is centered on the mean vector μ and has variance σ^2:

$$q(\theta;\phi) = \prod_{i}^{\Theta} \log\mathcal{N}(\theta_i|\mu_i,\sigma_i^2) \tag{7}$$

And the log-posterior $\log q(\theta;\phi)$ is given by

$$\log q(\theta;\phi) = \sum_{i}^{\Theta} \log\mathcal{N}(\theta_i|\mu_i,\sigma_i^2) \tag{8}$$

1.5 Optimizer

Suppose that the variational posterior is a diagonal Gaussian distribution, then a sample of the weights θ can be obtained by sampling a unit Gaussian, shifting it by a mean μ and a standard deviation σ. We parameterise the standard deviation pointwise as $\sigma = \log(1+\exp(\rho))$ and so ρ is always non-negative. The variational posterior parameters are $\theta = (\mu, \rho)$. Thus the transform from a sample of parameter-free noise and the variational posterior parameters that yields a posterior sample of the weight θ is: $\theta = t(\phi,\epsilon) = \mu + \log(1+\exp(\rho)) \circ \epsilon$ where $\circ$ is pointwise multiplication. Each step of optimization proceeds as follows:
1. Sample $\epsilon \sim \mathcal{N}(0, I)$.
2. Let $\phi = \mu + \log(1+\exp(\rho)) \circ \epsilon$.
3. Let $\theta = (\mu, \rho)$.
4. Let $f(\theta,\phi) = \log q(\theta;\phi) - \log p(\theta) p(D|\theta)$
5. Calculate the gradient with respect to the mean μ and standard derivation parameter ρ

$$\Delta_\mu = \frac{\partial f(\theta;\phi)}{\partial\theta} + \frac{\partial f(\theta;\phi)}{\partial\mu}, \qquad \Delta_\rho = \frac{\partial f(\theta;\phi)}{\partial\theta}\frac{\epsilon}{1+\exp(-\rho)} + \frac{\partial f(\theta;\phi)}{\partial\rho} \tag{9}$$

6. Update the variational parameters

$$\mu \leftarrow \mu - \alpha\Delta_\mu, \qquad \rho \leftarrow \rho - \alpha\Delta_\rho \tag{10}$$

Thus we can get the variational posterior.

2 Compute KL Divergence To Get Intrinsic Reward

The next step is to compute the KL divergence in the total reward function, which can be computed through the following minimization

$$\phi' = \underset{\phi}{\operatorname{argmin}} \left[\overbrace{\underbrace{D_{\mathbb{KL}}[q(\theta;\phi)\|q(\theta;\phi_{t-1})]}_{l_{\mathbb{KL}}(q(\theta;\phi))} - \mathbb{E}_{\theta\sim q(\cdot;\phi)}[\log p(s_t|\xi_{t-1},a_t;\theta)]}^{l(q(\theta;\phi),s_t)} \right] \tag{11}$$

The difference with the original loss is that we only update based on the latest sample. This means that instead of using the prior $p(\theta)$, we use the previous approximated posterior $q(\theta)$ for the KL term in the objective function $D_{\mathbb{KL}}[q(\theta;\phi)\|p(\theta)]$ becomes $D_{\mathbb{KL}}[q(\theta;\phi_t)\|q(\theta;\phi_{t-1})]$. Once $\phi^{'}$ has been obtained, we can use it to compute the intrinsic reward.
To optimize Eq. (11) efficiently, we only take a single second-order step. This way, the gradient is rescaled according to the curvature of the KL divergence at the origin. As such, we compute $D_{\mathbb{KL}}[q(\theta;\phi+\lambda\Delta\phi)\|q(\theta;\phi)]$, with the update step $\Delta\phi$ defined as

$$\Delta\phi = H^{-1}(l)\nabla_\phi l(q(\theta;\phi), s_t), \tag{12}$$

in which $H(l)$ is the Hessian of $l(q(\theta;\phi), s_t)$. Since we assume that the variational approximation is a fully factorized Gaussian, the KL divergence from posterior to prior has a particularly simple form

$$D_{\mathbb{KL}}[q(\theta;\phi)\|q(\theta;\phi')] = \frac{1}{2}\sum_{i=1}^{|\Theta|}\left(\left(\frac{\sigma_i}{\sigma_i^{'}}\right)^2 + 2\log\sigma_i^{'} - 2\log\sigma_i + \frac{(\mu_u^{'} - \mu_i)}{\sigma_i^{'2}}\right) - \frac{|\Theta|}{2}. \tag{13}$$

Because this KL divergence is approximately quadratic in tis parameters and the log-likelihood term can be seen as local linear compared to this highly curved KL term, we approximate H by only calculate it for the term KL term $l_{\mathbb{KL}}(q(\theta;\phi))$. This can be computed very efficiently in case of a fully factorized Gaussian distribution, as this approximation becomes a diagonal matrix. Looking at Eq. (13), we can calculate the following Hessian at the origin. The μ and ρ entries are defined as

$$\frac{\partial^2 l_{\mathbb{KL}}}{\partial \mu_i^2} = \frac{1}{log^2(1+e^{\rho_i})} \qquad \text{and} \qquad \frac{\partial^2 l_{\mathbb{KL}}}{\partial \rho_i^2} = \frac{2e^{2\rho_i}}{(1+e^{\rho_i})^2}\frac{1}{\log^2(1+e^{\rho_i})}, \tag{14}$$

while all other entries are zero. Furthermore, it is also possible to approximate the KL divergence through a second-order Taylor expansion as $\frac{1}{2}\Delta\phi H\Delta\phi = \frac{1}{2}(H^{-1}\nabla)^T H(H^{(}-1)\nabla)$, since both the value and gradient of the Kl divergence are zero at the origin. This gives us

$$D_{\mathbb{KL}}[q(\theta;\phi+\lambda\Delta\phi)\|q(\theta;\phi)] \simeq \frac{1}{2}\lambda^2 \nabla_\phi l^T H^{-1}(l_{\mathbb{KL}})\nabla_\phi l \tag{15}$$

Note that $H^{-1}(l_{\mathbb{KL}})$ is diagonal, so this expression can be computed efficiently.

3 conclusion

To minimize Eq.(2), we use Monte Carlo method approximate this exact cost as Eq.(3). Vy using Eq.(10), we can get the optimization solution of Eq.(3).
The next problem is how to compute the KL divergence in the total reward function. We optimizing Eq.(11) to get the KL divergence. And the Newton's update step can be written as: $\Delta\phi = H^{-1}(l)\nabla_\phi l(q(\theta;\phi), s_t)$. If we use the softmax function for discrete random variables and logistic function for continuous random variables, then the log-likelihood term in Eq.(11) can can be seen as locally linear compared to highly curved KL term of Gaussian. Thus, we use Eq.(15) to compute the KL divergence $D_{\mathbb{KL}}[q(\theta;\phi+\lambda\Delta\phi)\|q(\theta;\phi)]$.