Black box spike and slab variational inference, example with linear models

Laurent de Vito

18. December 2018

Let $\mathcal{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$ be our dataset with $\boldsymbol{x} \in \mathbb{R}^M$. We consider a linear regression model with spike and slab prior.

$$\begin{aligned}
w_m &\sim \mathcal{N}(0, \sigma_w^2), & m = 1, \cdots, M \\
s_m &\sim \text{Bernoulli}(\pi_w), & m = 1, \cdots, M \\
y_n &\sim \mathcal{N}(\sum_{m=1}^M w_m s_m x_{nm}, \sigma^2), & n = 1, \cdots, N
\end{aligned}$$

where x_{nm} designates the m-component of $\boldsymbol{x}_n$. σ^2 is the variance of the i.i.d. Gaussian noise. π_w and σ_w are hyperparameters.

We use black-box variational inference to find an approximation to the posterior over all parameters using optimization. Following Titsias and Lazaro-Gredilla in Spike and Slab Variational Inference for Multi-Task and Multiple Kernel Learning, we propose the following approximation to the posterior:

$$\begin{aligned}
q(\boldsymbol{w}, \boldsymbol{s}) &= \prod_{m=1}^M q(w_m|s_m)\ q(s_m) \\
&= \prod_{m=1}^M \mathcal{N}(s_m\mu_m, s_m\sigma_m^2 + (1 - s_m)\sigma_w^2)\ \pi_m^{s_m}(1 - \pi_m)^{(1-s_m)}
\end{aligned}$$

where μ_m, σ_m^2 and π_m for $m = 1, \cdots, M$ are variational parameters.

The ELBO is given by

$$\mathcal{L}(\phi; \boldsymbol{y}, \boldsymbol{X}) = H(q_\phi(\boldsymbol{w}, \boldsymbol{s}|\boldsymbol{y}, \boldsymbol{X})) + \mathrm{E}_{\boldsymbol{w},\boldsymbol{s}\sim q_\phi(\boldsymbol{w},\boldsymbol{s}|\boldsymbol{y},\boldsymbol{X})}[\log p(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{s}|\boldsymbol{X})]$$

with $\phi = \{(\mu_m, \sigma_m^2, \pi_m)\}_{m=1}^M$. The entropy of independent variables is the sum of the entropies of the independent variables. Hence we have

$$H(q_\phi(\boldsymbol{w}, \boldsymbol{s}|\boldsymbol{y}, \boldsymbol{X})) = \sum_{m=1}^M H(q_\phi(\boldsymbol{w}_m, \boldsymbol{s}_m|\boldsymbol{y}, \boldsymbol{X}))$$

Because of our factorization, we get:

$$
\begin{aligned}
-H(q_\phi(\boldsymbol{w}_m, \boldsymbol{s}_m|\boldsymbol{y}, \boldsymbol{X})) =& \int (1-\pi_m)\mathcal{N}(w_m|0,\sigma_w^2)\log\left[(1-\pi_m)\mathcal{N}(w_m|0,\sigma_w^2)\right]\ dw_m \\
&+ \int \pi_m \mathcal{N}(w_m|\mu_m,\sigma_m^2)\log\left[\pi_m\mathcal{N}(w_m|\mu_m,\sigma_m^2)\right]\ dw_m \\
=&(1-\pi_m)[\log(1-\pi_m) - H(\mathcal{N}(w_m|0,\sigma_w^2))] \\
&+\pi_m[\log\pi_m - H(\mathcal{N}(w_m|\mu_m,\sigma_m^2))] \\
=&(1-\pi_m)[\log(1-\pi_m) - 0.5\log(2\pi e\sigma_w^2))] \\
&+\pi_m[\log\pi_m - 0.5\log(2\pi e\sigma_m^2))] \\
=&(1-\pi_m)\log(1-\pi_m) + \pi_m\log\pi_m \\
&-\frac{1}{2}(1-\pi_m)\log(2\pi e\sigma_w^2) - \frac{1}{2}\pi_m\log(2\pi e\sigma_m^2)
\end{aligned}
$$

An unbiased Monte Carlo approximation to the expectation of $\log p(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{s}|\boldsymbol{X})$ can be computed by first sampling from the Bernoulli variables s_m using the Gumbel-Max trick and then from the Gaussian variables w_m using the reparameterization trick.