

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

Laurent de Vito

18. December 2018

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

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

where x_{nm} designates the m -component of \mathbf{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:

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

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

The ELBO is given by

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

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

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

Because of our factorization, we get:

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

An unbiased Monte Carlo approximation to the expectation of $\log p(\mathbf{y}, \mathbf{w}, \mathbf{s} | \mathbf{X})$ 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**.