STCS 6701: Foundations of Graphical Models: Reading 10

Maxime TCHIBOZO (MT3390)

November 2020

1 Covariances, Robustness, and Variational Bayes (2018) – Giordano et al.

While Variational Inference (a.k.a Variational Bayes) usually provides accurate estimates of posterior parameter means, it underestimates posterior and marginal covariances. VI is not naturally suited to approximating posterior moments, and this paper proposes an approach based on sensitivity analysis (sensitivity analysis: evaluating how target variables change with input variables – closely related to optimization problems) for covariance estimation.

The local sensitivity $\mathbf{S}_{\alpha_0}$ of $\mathbb{E}_{p_\alpha}[g(\theta)]$ is defined as: $\frac{d\mathbb{E}_{p_\alpha}[g(\theta)]}{d\alpha}\Big|_{\alpha_0}$. We can then use a Taylor-style expansion at a point α_0 of the parameter support to approximate the global sensitivity: $\mathbb{E}_{p_\alpha}[g(\theta)] \approx \mathbb{E}_{p_{\alpha_0}}[g(\theta)] + \mathbf{S}_{\alpha_0}^T(\alpha - \alpha_0)$.

While $\mathbb{E}_{p_\alpha}[g(\theta)]$ cannot be usually be differentiated, we can use the above approximation and differentiate S_{α_0}, which under certain regularity assumptions is equal to a covariance which can be evaluated using MCMC samples of a known distribution.

We recall the VI approximation measure $q_\alpha(\theta) = argmin_q\{KL\big(q(\theta;\eta)\big|\big|p_\alpha(\theta)\big)\} = q(\theta, \eta^*(\alpha))$ and define:

$$H_{\eta,\eta} = \frac{\partial^2 KL(q(\theta;\eta)||p_0(\theta))}{\partial\eta\partial\eta^T}\Big|_{\eta=\eta_0^*}, f_{\alpha\eta} = \frac{\partial^2\mathbb{E}_{q(\theta;\eta)}[\rho(\theta,\alpha)]}{\partial\alpha\partial\eta^T}\Big|_{\eta=\eta_0^*,\alpha=\alpha_0}, g_\eta = \frac{\partial\mathbb{E}_{q(\theta;\eta)}[g(\theta)]}{\partial\eta^T}\Big|_{\eta=\eta_0^*}.$$

Under continuity, differentiability and smoothness assumptions:

$\frac{d\mathbb{E}_{q_\alpha}[g(\theta)]}{d\alpha T}\Big|_{\alpha_0} = g_\eta H_{\eta\eta}^{-1} f_{\alpha\eta} T$ and $\frac{d\mathbb{E}_{q_\alpha}[g(\theta)]}{d\alpha^T} \approx \frac{d\mathbb{E}_{p_\alpha}[g(\theta)]}{d\alpha^T}$.

The above conditions combined allow for a definition of the Linear Response Varitional Bayes (LRVB) covariance: $Cov_{q_0}^{LR}(g(\theta)) = g_\eta H_{\eta\eta}^{-1} g_\eta^T$ and $Cov_{q_0}^{LR}(g(\theta)) \approx Cov_{p_0}(g(\theta))$.

This means that $Cov_{q_0}^{LR}(g(\theta))$ is a much better approximation of $Cov_{p_0}(g(\theta))$ than $Cov_{q_0}(g(\theta))$, in particular where there are at least some moments of p_0 that q_0 does not approximate well. In other words, the LRVB approximation compensates for regions in the support which the Mean-Field Exponential Family assumption is too strong and gives poor approximations of p. In past research, this problem was addressed by further increasing the complexity of the variational family, but this hinders the computational efficiency which makes VI so attractive. Another approach is to use the Laplace Approximation to approximate p_0:

$$q_{Lap}(\theta) \sim \mathcal{N}\left(\hat{\theta}_{Lap}, Cov_{q_{Lap}}^{Lap}\right) \text{ with } \hat{\theta}_{Lap} = argmax_{\theta} p_0(\theta) \text{ and } Cov_{q_{Lap}}^{Lap} = -\frac{\partial^2 p_0(\theta)}{\partial\theta\partial\theta^T}\Big|_{\hat{\theta}_{Lap}}.$$

The Laplace Approximation covariance estimate can be seen as an instance of the LRVB covariance approximation under specific constraints on the variational family q.