Horseshoe for random features

Beau Coker

1 Model

Network:

$$f(\mathbf{x}_n; \nu, \{\tau_d\}, \{\beta_k\}, \{\mathbf{w}_k\}, b) = \sum_{k=1}^{K} \beta_k \Phi(\mathbf{x}_n; \nu, \tau_d, \mathbf{w}_k, b) \tag{1}$$

$$\Phi(\mathbf{x}_n; s_d, \mathbf{w}_k, b) = \sqrt{2} \cos\left(\sum_{d=1}^{D} \nu \tau_d w_{k,d} x_{n,d}\right) \tag{2}$$

In matrix notation:

$$f(X; S, \boldsymbol{\beta}) = \Phi(X; S, W, b)\boldsymbol{\beta} \tag{3}$$

$$\Phi(X; S, W, b) = \sqrt{2} \cos(XSW + \mathbf{b}) \tag{4}$$

$$\tag{5}$$

where $S = \mathrm{diag}(\nu\tau_d, \ldots, \nu\tau_D)$ and $W = [\mathbf{w}_1^T, \ldots, \mathbf{w}_K^T]^T$.

Model:

$$\beta_k \sim \mathcal{N}(0, \sigma_\beta^2), \quad k = 1, \ldots, K \tag{6}$$

$$\tau_d \sim C^+(0, b_\tau), \quad d = 1, \ldots, D \tag{7}$$

$$\nu \sim C^+(0, b_\nu) \tag{8}$$

$$y_n \,|\, \mathbf{x}_n, \{\beta_k\}, \nu, \{\tau_d\}, \{\mathbf{w}_k\}, b \sim \mathcal{N}(f(\mathbf{x}_n; \nu, \{\tau_d\}, \{\beta_k\}), \sigma^2), \quad n = 1, \ldots, N \tag{9}$$

To improve inference, we add auxiliary variables $\{\lambda_d\}_{d=1}^D$ and ϑ:

$$p(\tau_d, \lambda_d) = p(\tau_d \,|\, \lambda_d) p(\lambda_d) = \mathrm{InvGamma}\left(\tau_d; \frac{1}{2}, \frac{1}{\lambda_d}\right) \mathrm{InvGamma}\left(\lambda_d; \frac{1}{2}, \frac{1}{b_\tau}\right) \tag{10}$$

$$p(\nu, \vartheta) = p(\nu \,|\, \vartheta) p(\vartheta) = \mathrm{InvGamma}\left(\nu; \frac{1}{2}, \frac{1}{\vartheta}\right) \mathrm{InvGamma}\left(\vartheta; \frac{1}{2}, \frac{1}{b_\nu}\right) \tag{11}$$

2 Variational approximation

We approximate the posterior distribution of $\{\tau_d\}$, ν, $\{\lambda_d\}$, and ϑ with a fully factorized variational distribution. We use a log normal distribution for ν and each τ_d, and an inverse Gamma distribution for ϑ and each λ_k.

3 Inference

For $s = 1, \ldots$

- Sample $\{\beta_k^{(s)}\}$ from the posterior conditional on ν and $\{\tau_d^{(s)}\}$ (conjugate).
- Update the variational parameters for $\{\tau_d\}$, ν, $\{\lambda_d\}$, and ϑ by optimizing the ELBO conditional on $\{\beta_k^{(s)}\}$ for a fixed number of steps.