MODEL D: LEARNING V_J

V. CHEN

1. First algorithm. This algorithm reparameterizes the problem in terms of the random vectors v and ξ. v_j modifies the scale of influence of q_j, the jth eigenvector of the graph Laplacian, on the classifying function u.

$$T(v,\xi) = \sum_{i=0}^{M} v_i \xi_i q_i = u \tag{1}$$

Here, M is fixed. We take $\xi \sim \mathsf{N}(0, I)$. For the prior of v, the first step is to slightly "cheat" by using a working solution to this problem: apply one of the algorithms that learns τ, α and get estimates for $\mathbb{E}u_j^2$ where $u = \sum_{i=0}^{M} u_j q_j$. Then, pick the prior for v to be

$$v_j \sim \mathsf{U}\left((1-a)(\mathbb{E}u_j^2)^{1/2}, (1+a)(\mathbb{E}u_j^2)^{1/2}\right)$$

where a is a fixed scalar.
Finally, we derive an expression for the posterior with Bayes' theorem.

$$\begin{aligned}
\mathbb{P}(v,\xi|y) &\propto \mathbb{P}(y|v,\xi)\mathbb{P}(v,\xi)\\
&\propto \exp\left(-\Phi(T(v,\xi))\right)\mathbb{P}(v)\mathbb{P}(\xi)\\
&\propto \exp\left(-\Phi(T(v,\xi)) + \log(\pi_0(v)) - \frac{1}{2}\langle \xi, \xi\rangle\right)
\end{aligned}$$

Let $h(v,\xi)$ denote the joint posterior on v and ξ. Then,

$$h(v,\xi) \propto \exp\left(-\Phi(T(v,\xi)) + \log(\pi_0(v)) - \frac{1}{2}\langle \xi, \xi\rangle\right). \tag{2}$$

2. Second algorithm. We can also take a Gaussian prior on v_j. With fixed τ_0, α_0, take the prior

$$v_j \sim \mathsf{N}\left(\frac{1}{(\lambda_j + \tau_0^2)^{\alpha_0/2}}, \sigma_j^2\right)$$

so that in all, $v \sim \mathsf{N}(\mu, D)$ with $\mu_j = \frac{1}{(\lambda_j+\tau_0^2)^{\alpha_0/2}}$ and $D = \mathrm{diag}(\sigma_j^2)$. Again, we derive an expression for the joint posterior $h(v,\xi)$:

$$\begin{aligned}
\mathbb{P}(v,\xi|y) &\propto \mathbb{P}(y|v,\xi)\mathbb{P}(v,\xi)\\
&\propto \exp\left(-\Phi(T(v,\xi))\right)\mathbb{P}(v)\mathbb{P}(\xi)\\
&\propto \exp\left(-\Phi(T(v,\xi)) - \frac{1}{2}\langle v-\mu, D^{-1}(v-\mu)\rangle - \frac{1}{2}\langle \xi, \xi\rangle\right).
\end{aligned}$$

And we obtain:

$$h(v,\xi) = \exp\left(-\Phi(T(v,\xi)) - \frac{1}{2}\langle v-\mu, D^{-1}(v-\mu)\rangle - \frac{1}{2}\langle \xi, \xi\rangle\right). \tag{3}$$

REFERENCES

Algorithm 1 Non-centered parameterization, hierarchical with v

Choose $v^{(0)}, \xi^{(0)} \in \mathbb{R}^N, \beta \in (0,1], \epsilon > 0$.
for $k = 0$ to S **do**
  Propose $\hat{\xi}^{(k)} = (1-\beta^2)^{\frac{1}{2}} \xi^{(k)} + \beta \zeta^{(k)}$, $\zeta^{(k)} \sim \mathsf{N}(0, I)$
  Make transition $\xi^{(k)} \to \hat{\xi}^{(k)}$ with probability

$$A(\xi^{(k)} \to \hat{\xi}^{(k)}) = \min\left\{1, \exp\left(\Phi(T(v^{(k)}, \xi^{(k)})) - \Phi(T(v^{(k)}, \hat{\xi}^{(k)}))\right)\right\}$$

  Propose $\hat{v}^{(k)} = v^{(k)} + \epsilon \rho^{(k)}, \rho^{(k)} \sim \mathsf{N}(0, I)$
  if $\hat{v}_j^{(k)} \notin \left[(1-a)(\mathbb{E}u_j^2)^{1/2}, (1+a)(\mathbb{E}u_j^2)^{1/2}\right]$ for any j **then**
    Reject and set $v^{(k+1)} = v^{(k)}$.
  else
    Make transition $v^{(k)} \to \hat{v}^{(k)}$ with probability

$$\begin{aligned} A(v^{(k)} \to \hat{v}^{(k)}) &= \min\left\{1, \frac{h(\hat{v}^{(k)}, \xi^{(k+1)})}{h(v^{(k)}, \xi^{(k+1)})}\right\} \\ &= \min\left\{1, \exp\left(\Phi(T(v^{(k)}, \xi^{(k+1)})) - \Phi(T(\hat{v}^{(k)}, \xi^{(k+1)}))\right)\right\} \end{aligned}$$

  end if
end for
return $\{T(v^{(k)}, \xi^{(k)}), v^{(k)}, \xi^{(k)}\}$

Algorithm 2 Non-centered parameterization, hierarchical with v, Gaussian prior on v

Choose $v^{(0)}, \xi^{(0)} \in \mathbb{R}^N, \beta_1, \beta_2 \in (0,1]$.
for $k = 0$ to S **do**
  Propose $\hat{\xi}^{(k)} = (1-\beta_1^2)^{\frac{1}{2}} \xi^{(k)} + \beta_1 \zeta^{(k)}$, $\zeta^{(k)} \sim \mathsf{N}(0, I)$
  Make transition $\xi^{(k)} \to \hat{\xi}^{(k)}$ with probability

$$A(\xi^{(k)} \to \hat{\xi}^{(k)}) = \min\left\{1, \exp\left(\Phi(T(v^{(k)}, \xi^{(k)})) - \Phi(T(v^{(k)}, \hat{\xi}^{(k)}))\right)\right\}$$

  Propose $\hat{v}^{(k)} = \mu + (1-\beta_2^2)^{\frac{1}{2}}(v-\mu) + \beta_2 \rho^{(k)}, \rho^{(k)} \sim \mathsf{N}(0, D)$
  Make transition $v^{(k)} \to \hat{v}^{(k)}$ with probability

$$A(v^{(k)} \to \hat{v}^{(k)}) = \min\left\{1, \exp\left(\Phi(T(v^{(k)}, \xi^{(k+1)})) - \Phi(T(\hat{v}^{(k)}, \xi^{(k+1)}))\right)\right\}$$

end for
return $\{T(v^{(k)}, \xi^{(k)}), v^{(k)}, \xi^{(k)}\}$