Stochastic SuSiE for general likelihoods 2

Karl Tayeb

2022-02-19

Contents

Sample many times, fit one regression

Before we described an approach where we fit many SERs, then use a first order approximation to justify averaging the natural parameters. Here we describe an approach for optimizing $\hat{\mathcal{L}}_{\mathrm{SuSiE},M}$ more directly using a single SER.

Recall,

$$\hat{\mathcal{L}}_{\mathrm{SuSiE},M}(q_l) = \mathbb{E}_{q_l}\left[\frac{1}{M}\sum \log p(y|b_l, \gamma_l, b_{-l}^{(m)}, \gamma_{-l}^{(m)}) + \log p(b_l, \gamma_l)\right] + H(q_l)$$

This can be thought of us a weighted regression problem with an augmented data set– there are now M copies of the data, each with their own predicted offset determined by the sample $b_{-l}^{(m)}, \gamma_{-l}^{(m)}$. But each observation is given a weight $1/M$.

Online learning

We can imagine adaptively choosing M for each of the variables separately. After each sample we can update our regression. With this approach we can quickly drop the features that are unlikely to contribute substantially to the SER. We can focus more computational effort on estimating the posterior for features with stronger association. Our approximation will be acurate where it counts, without wasting computation where it contributes very little to inference (e.g. a feature that will have $q(\gamma_l = j) \approx 0)$)