Subset-of-Data Variational Inference for Deep Gaussian-Processes Regression : Supplementary

Ayush Jain[1] **P. K. Srijith**[1] **Mohammad Emtiyaz Khan**[2]

[1]Department of Computer Science and Engineering , Indian Institute of Technology Hyderabad, India
[2]RIKEN Center for AI Project , Tokyo, Japan

1 TRAINING AND INFERENCE IN DEEP GAUSSIAN PROCESS MODELS

The sparse variational inference approach simultaneously addresses intractability and scalability issues in Deep GPs. This introduces variational parameters and inducing points with inducing input $\bar{\mathbf{Z}}^l$ and inducing outputs $\mathbf{U}^l$ for each layer l, all of which are learnt from the variational lower bound.

Figure 1 and (1) provides the graphical model and full likelihood respectively for sparse Deep GP model introduced in Damianou and Lawrence [2013]. While, Fig. 2 and (2) provides them for the sparse deep GP introduced in Salimbeni and Deisenroth [2017]. The later consider the noise inside the kernel and hence drops the explicit noisy representations $\mathbf{Z}^l$ from the model.

$$p(\mathbf{y}|\mathbf{F}^L)\prod_{l=L-1}^{1} p(\mathbf{F}^{l+1}|\mathbf{U}^{l+1},\mathbf{Z}^l,\bar{\mathbf{Z}}^l)p(\mathbf{U}^{l+1}|\bar{\mathbf{Z}}^l)\times p(\mathbf{Z}^l|\mathbf{F}^l)p(\mathbf{F}^1|\mathbf{X},\bar{\mathbf{X}},\mathbf{U}^1)p(\mathbf{U}^1|\bar{\mathbf{X}}) \tag{1}$$

$$p(\mathbf{y}|\mathbf{F}^L)\prod_{l=L-1}^{1} p(\mathbf{F}^{l+1}|\mathbf{U}^{l+1},\mathbf{F}^l,\bar{\mathbf{Z}}^l)p(\mathbf{U}^{l+1}|\bar{\mathbf{Z}}^l)\times p(\mathbf{F}^1|\mathbf{X},\bar{\mathbf{X}},\mathbf{U}^1)p(\mathbf{U}^1|\bar{\mathbf{X}}) \tag{2}$$

Here, $\bar{\mathbf{Z}}^l$ are pseudo inducing inputs at layer l, $\bar{\mathbf{X}}$ pseudo inputs at layer 1, $\mathbf{U}^l = f(\bar{\mathbf{Z}}^l)$ are inducing points which are function values evaluated at corresponding pseudo inputs. For simplicity we will avoid the use of $\bar{\mathbf{Z}}^l$ and $\bar{\mathbf{X}}$ in further discussions.

The marginal likelihood and posterior computation are intractable in these models because $\mathbf{Z}^l$ and $\mathbf{F}^l$ are appearing in non linear manner inside the covariance matrices of distributions $p(\mathbf{F}^{l+1}|\mathbf{U}^l,\mathbf{Z}^l)$ and $p(\mathbf{F}^{l+1}|\mathbf{U}^l,\mathbf{F}^l)$ in (1) and (2) respectively. Both Damianou and Lawrence [2013] and Salimbeni and Deisenroth [2017] obtain tractable lower bound to marginal likelihood using variational inference by introducing an approximate variational posterior distribution. The deep GP inference in Damianou and Lawrence [2013] used a mean field variational approximation while Salimbeni and Deisenroth [2017] used a doubly stochastic variational inference which maintained the conditional structure of deep GP. Equation (3) shows the variational posterior used in Damianou and Lawrence [2013] where $q(\mathbf{U}^l)$ and $q(\mathbf{Z}^l)$ are variational distributions factorised accross the layers. Variational distribution over $\mathbf{Z}^l$ is independent of the input from previous layer which disconnect the link between the layers and result in losing the correlations between layers as shown in Fig. 3.

$$\begin{aligned} q(\mathbf{F},\mathbf{U},\mathbf{Z}) = \prod_{l=L-1}^{1} & p(\mathbf{F}^{l+1}|\mathbf{U}^{l+1},\mathbf{Z}^l)q(\mathbf{U}^{l+1})\times \\ & q(\mathbf{Z}^l)p(\mathbf{F}^1|\mathbf{U}^1,\mathbf{X})q(\mathbf{U}^1) \end{aligned} \tag{3}$$

The variational posterior used in Salimbeni and Deisenroth [2017] is shown in (4). Here, $q(\mathbf{U}^l)$ is a variational distribution factorised across the layers while $q(\mathbf{F}^l)$ ($q(\mathbf{F}^l) = \bar{q}(\mathbf{F}^l|\mathbf{F}^{l-1}) = \int p(\mathbf{F}^l|\mathbf{U}^l,\mathbf{F}^{l-1})q(\mathbf{U}^l)d\mathbf{U}^l$) maintains the dependence

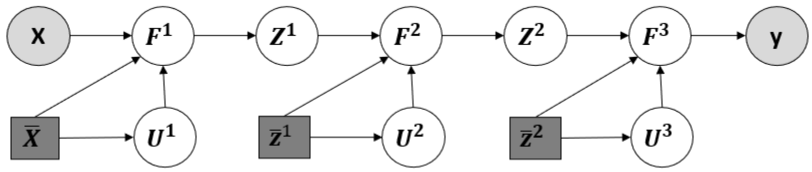


Figure 1: Deep Gaussian Process Model with inducing points as in Damianou and Lawrence [2013]

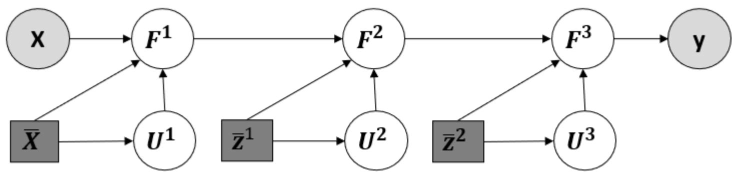


Figure 2: Deep Gaussian Process Model with inducing points in Salimbeni and Deisenroth [2017]

between the layers and full conditional structure as shown in Fig. 4.

$$q(\mathbf{F},\mathbf{U}) = \prod_{l=L-1}^{1} p(\mathbf{F}^{l+1}|\mathbf{U}^{l+1},\mathbf{F}^{l})q(\mathbf{U}^{l+1})\times \\ p(\mathbf{F}^{1}|\mathbf{U}^{1},\mathbf{X})q(\mathbf{U}^{1}) \tag{4}$$

2 PROPOSED INFERENCE METHOD FOR DEEP GAUSSIAN PROCESSES

2.1 EVIDENCE LOWER BOUND

We learn the variational parameters and hyper-parameters by maximizing the evidence lower bound. We derive the evidence lower bound (ELBO) for the proposed approach as follows.

$$ELBO = E_{q\left(\mathbf{F}^L,\left\{\mathbf{F}^l,\mathbf{Z}^l\right\}_{l=1}^{L-1}\right)}\left[\log\left(\frac{p(\mathbf{y},\left\{\mathbf{F}^l,\mathbf{Z}^{l-1}\right\}_{l=1}^{L})}{q\left(\mathbf{F}^L,\left\{\mathbf{F}^l,\mathbf{Z}^l\right\}_{l=1}^{L-1}\right)}\right)\right]$$

$$\text{where } p(\mathbf{y},\left\{\mathbf{F}^l,\mathbf{Z}^{l-1}\right\}_{l=1}^{L}) = \prod_{i=1}^{N} p(y_i|f_i^L)\left(\prod_{l=2}^{L} p(\mathbf{F}^l|\mathbf{Z}^{l-1})p(\mathbf{Z}^{l-1}|\mathbf{F}^{l-1})\right)p(\mathbf{F}^1|\mathbf{X})$$

$$\begin{aligned}\text{and } q\left(\mathbf{F}^L,\left\{\mathbf{F}^l,\mathbf{Z}^l\right\}_{l=1}^{L-1}\right) &= p(\mathbf{y}_S|\mathbf{F}_S^L)\left(\prod_{l=2}^{L} p(\mathbf{F}_{\bar{S}}^l|\mathbf{F}_S^l,\mathbf{Z}^{l-1})q(\mathbf{F}_S^l)p(\mathbf{Z}^{l-1}|\mathbf{F}^{l-1})\right)p(\mathbf{F}_{\bar{S}}^1|\mathbf{F}_S^1,X)q(\mathbf{F}_S^1)\frac{1}{\mathcal{Z}} \\ &= \prod_{n\in S} p(y_n|f_n^L)\left(\prod_{l=2}^{L}\prod_{d=1}^{D^l} p(\mathbf{f}_{\bar{S},d}^l|\mathbf{f}_{S,d}^l,\mathbf{Z}^{l-1})q(\mathbf{f}_{S,d}^l)\prod_{n=1}^{N}\mathcal{N}(Z_{n,:}^{l-1};\mathbf{f}_{n,:}^{l-1},\sigma^2 I)\right) \\ &\prod_{d=1}^{D^1} p(\mathbf{f}_{\bar{S},d}^1|\mathbf{f}_{S,d}^1,\mathbf{X})q(\mathbf{f}_{S,d}^1)\frac{1}{\mathcal{Z}}\end{aligned}$$

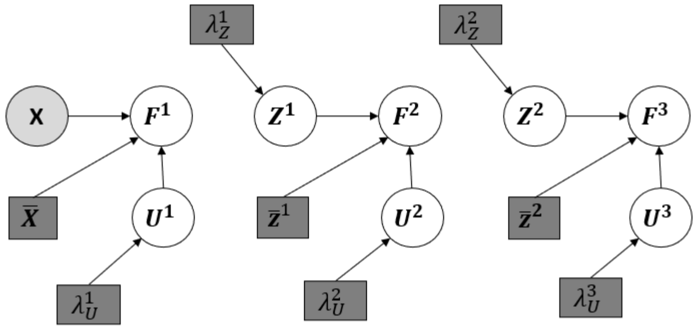


Figure 3: Approximate Inference For Deep Gaussian Process in Damianou and Lawrence [2013]

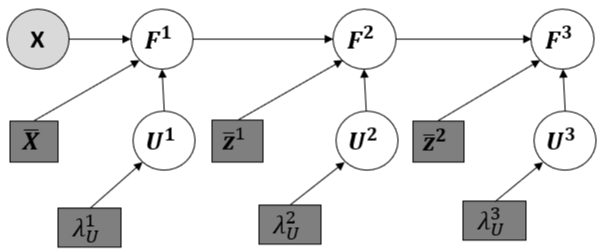


Figure 4: Doubly Stochastic Variational Inference For Deep Gaussian Process Salimbeni and Deisenroth [2017]

In the ELBO derivation, a few terms inside the logarithm cancel due to the factorisation assumed in the variational distribution

$$
\begin{aligned}
ELBO &= E_q\left[\log\left(\frac{\prod_{i=1}^{N} p(y_i|\mathbf{f}_i^L)\left(\prod_{l=2}^{L} p(\mathbf{F}^l|\mathbf{Z}^{l-1})p(\mathbf{Z}^{l-1}|\mathbf{F}^{l-1})\right)p(\mathbf{F}^1|\mathbf{X})\mathcal{Z}}{p(\mathbf{y}_S|\mathbf{F}_S^L)\left(\prod_{l=2}^{L} p(\mathbf{F}_{\bar{S}}^l|\mathbf{F}_S^l,\mathbf{Z}^{l-1})q(\mathbf{F}_S^l)p(\mathbf{Z}^{l-1}|\mathbf{F}^{l-1})\right)p(\mathbf{F}_{\bar{S}}^1|\mathbf{F}_S^1,X)q(\mathbf{F}_S^1)}\right)\right] \\
&= E_{q\left(\mathbf{F}^L,\left\{\mathbf{F}^l\mathbf{Z}^l\right\}_{l=1}^{L-1}\right)}\left[\log\left(\frac{p(\mathbf{y}_{\bar{S}}|\mathbf{F}_{\bar{S}}^L)\left(\prod_{l=2}^{L} p(\mathbf{F}_S^l|\mathbf{Z}_S^{l-1})\right)p(\mathbf{F}_S^1|\mathbf{X}_S)\mathcal{Z}}{\prod_{l=1}^{L} q(\mathbf{F}_S^l)}\right)\right] \\
&= E_{q(\mathbf{F}_{\bar{S}}^L)}[log(p(\mathbf{y}_{\bar{S}}|\mathbf{F}_{\bar{S}}^L))] + log(\mathcal{Z}) - \sum_{l=2}^{L-1}\sum_{d=1}^{D^l} E_q(\mathbf{Z}_S^{l-1})\left[KL(q(\mathbf{f}_{S,d}^l)||p(\mathbf{f}_{S,d}^l|\mathbf{Z}_S^{l-1}))\right] \\
&- \sum_{d=1}^{D^1} KL(q(\mathbf{f}_{S,d}^1)||p(\mathbf{f}_{S.d}^1|\mathbf{X}_S)) + E_{q(\mathbf{Z}_S^{L-1})}\left[E_{\hat{q}(\mathbf{F}_S^L)}\left[log\left(\frac{p(\mathbf{F}_S^L|\mathbf{Z}_S^{L-1})}{q(\mathbf{F}_S^L)}\right)\right]\right] \\
&= E_{q(\mathbf{F}_{\bar{S}}^L)}[log(p(\mathbf{y}_{\bar{S}}|\mathbf{F}_{\bar{S}}^L))] + log(\mathcal{Z}) - \sum_{l=2}^{L-1}\sum_{d=1}^{D^l} E_q(\mathbf{Z}_S^{l-1})\left[KL(q(\mathbf{f}_{S,d}^l)||p(\mathbf{f}_{S,d}^l|\mathbf{Z}_S^{l-1}))\right] \\
&- \sum_{d=1}^{D^1} KL(q(\mathbf{f}_{S,d}^1)||p(\mathbf{f}_{S.d}^1|\mathbf{X}_S)) + E_{q(\mathbf{Z}_S^{L-1})}\left[E_{\hat{q}(\mathbf{F}_S^L)}\left[log\left(\frac{p(\mathbf{F}_S^L|\mathbf{Z}_S^{L-1})\hat{q}(\mathbf{F}_S^L)}{\hat{q}(\mathbf{F}_S^L)q(\mathbf{F}_S^L)}\right)\right]\right]
\end{aligned}
$$

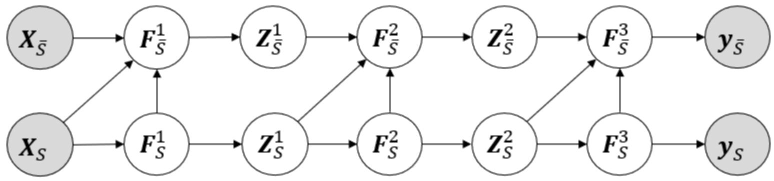


Figure 5: Proposed Subset-of-Data Representation of Original Deep Gaussian Process

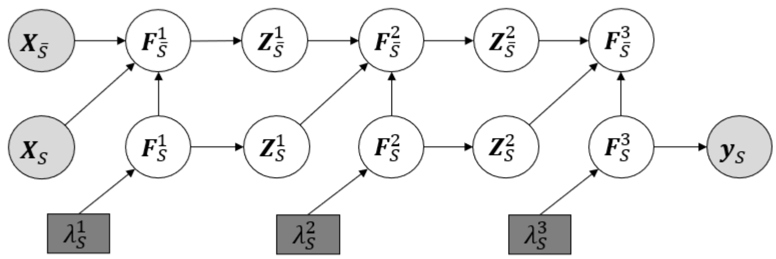


Figure 6: Proposed Subset-of-Data Variational Inference For Deep Gaussian Process

$$
\begin{aligned}
&= E_{q(\mathbf{F}^L_{\bar{S}})}[\log(p(\mathbf{y}_{\bar{S}}|\mathbf{F}^L_{\bar{S}}))] + \log(\mathcal{Z}) - \sum_{l=2}^{L-1}\sum_{d=1}^{D^l} E_{q(\mathbf{Z}^{l-1}_S)}\left[KL(q(\mathbf{f}^l_{S,d})\|p(\mathbf{f}^l_{S,d}|\mathbf{Z}^{l-1}_S))\right] \\
&- \sum_{d=1}^{D^1} KL(q(\mathbf{f}^1_{S,d})\|p(\mathbf{f}^1_{S.d}|\mathbf{X}_S)) - E_{q(\mathbf{Z}^{L-1}_S)}[KL(\hat{q}(\mathbf{F}^L_S)\|p(\mathbf{F}^L_S|\mathbf{Z}^{L-1}_S))] \\
&+ KL(\hat{q}(\mathbf{F}^L_S)\|q(\mathbf{F}^L_S)) \\
&= E_{q(\mathbf{F}^L_{\bar{S}})}[\log(p(\mathbf{y}_{\bar{S}}|\mathbf{F}^L_{\bar{S}}))] + \log(\mathcal{Z}) - \sum_{l=2}^{L-1}\sum_{d=1}^{D^l} E_{q(\mathbf{Z}^{l-1}_S)}\left[KL(q(\mathbf{f}^l_{S,d})\|p(\mathbf{f}^l_{S,d}|\mathbf{Z}^{l-1}_S))\right] \\
&- \sum_{d=1}^{D^1} KL(q(\mathbf{f}^1_{S,d})\|p(\mathbf{f}^1_{S.d}|\mathbf{X}_S)) - E_{q(\mathbf{Z}^{L-1}_S)}[KL(\hat{q}(\mathbf{F}^L_S)\|p(\mathbf{F}^L_S|\mathbf{Z}^{L-1}_S))] \\
&+ E_{\hat{q}(\mathbf{F}^L_S)}[log(p(\mathbf{y}_S|\mathbf{F}^L_S))] - \log(\mathcal{Z}) \\
&= E_{q(\mathbf{F}^L_{\bar{S}})}[\log(p(\mathbf{y}_{\bar{S}}|\mathbf{F}^L_{\bar{S}}))] + E_{\hat{q}(\mathbf{F}^L_S)}[log(p(\mathbf{y}_S|\mathbf{F}^L_S))] - \sum_{d=1}^{D^1} KL(q(\mathbf{f}^1_{S,d})\|p(\mathbf{f}^1_{S.d}|\mathbf{X}_S)) \\
&- \sum_{l=2}^{L-1}\sum_{d=1}^{D^l} E_{q(\mathbf{Z}^{l-1}_S)}\left[KL(q(\mathbf{f}^l_{S,d})\|p(\mathbf{f}^l_{S,d}|\mathbf{Z}^{l-1}_S))\right] - E_{q(\mathbf{Z}^{L-1}_S)}[KL(\hat{q}(\mathbf{F}^L_S)\|p(\mathbf{F}^L_S|\mathbf{Z}^{L-1}_S))]
\end{aligned}
$$

where

$$
\begin{aligned}
p(\mathbf{f}^l_{\bar{S},d}|\mathbf{f}^l_{S,d},\mathbf{Z}^{l-1}) = \mathcal{N}(\mathbf{f}^l_{\bar{S},d};\mathbf{m}^l_{\bar{S},d},\mathbf{V}^l_{\bar{S}}) \quad ; \quad \mathbf{m}^l_{\bar{S},d} = \mathbf{K}^l_{\bar{S},S}{\mathbf{K}^l_{S,S}}^{-1}\mathbf{f}^l_{S,d} \quad ; \quad \mathbf{V}^l_{\bar{S}} = \mathbf{K}^l_{\bar{S},\bar{S}} - \mathbf{K}^l_{\bar{S},S}{\mathbf{K}^l_{S,S}}^{-1}\mathbf{K}^l_{S,\bar{S}} \\
q(\mathbf{f}^l_{S,d}) = \mathcal{N}(\mathbf{f}^l_{S,d};\mu^l_{S,d},\mathbf{\Sigma}^l_{S,d}) \quad ; \quad p(\mathbf{y}_S|\mathbf{F}^L_S) = \mathcal{N}(\mathbf{y}_S;\mathbf{F}^L_S,\sigma^2\mathbf{I}) \\
\mathcal{Z} = p(\mathcal{D}_{\mathcal{S}}) = \int p(\mathbf{y}_S|\mathbf{F}^L_S)q(\mathbf{F}^L_S)d\mathbf{F}^L_S = \mathcal{N}(\mathbf{y}_S|\mu^L_S,\sigma^2\mathbf{I}+\mathbf{\Sigma}^L_S)
\end{aligned}
$$

2.2 COMPUTATION OF MARGINAL DISTRIBUTION $q(\mathbf{F}^L_{\bar{S}})$

$$q\left(\mathbf{F}^L, \left\{\mathbf{F}^l, \mathbf{Z}^l\right\}_{l=1}^{L-1}\right) = p(\mathbf{y}_S|\mathbf{F}^L_S)\left(\prod_{l=2}^{L} p(\mathbf{F}^l_{\bar{S}}|\mathbf{F}^l_S, \mathbf{Z}^{l-1})q(\mathbf{F}^l_S)p(\mathbf{Z}^{l-1}|\mathbf{F}^{l-1})\right)p(\mathbf{F}^1_{\bar{S}}|\mathbf{F}^1_S, X)q(\mathbf{F}^1_S)\frac{1}{\mathcal{Z}}$$

$$\mathcal{Z} * q(\mathbf{F}^L_{\bar{S}}) = \int p(\mathbf{y}_S|\mathbf{F}^L_S)\left(\prod_{l=2}^{L} p(\mathbf{F}^l_{\bar{s}}|\mathbf{F}^l_S, \mathbf{Z}^{l-1})q(\mathbf{F}^l_S)p(\mathbf{Z}^{l-1}|\mathbf{F}^{l-1})\right)p(\mathbf{F}^1_{\bar{S}}|\mathbf{F}^1_S, \mathbf{X})q(\mathbf{F}^1_S)$$
$$d\mathbf{F}^L_S d\left\{\mathbf{F}^l, \mathbf{Z}^l\right\}_{l=1}^{L-1}$$

$$\mathcal{Z} * q(\mathbf{F}^L_{\bar{S}}) = \int (p(\mathbf{y}_S|\mathbf{F}^L_S)q(\mathbf{F}^L_S))p(\mathbf{F}^L_{\bar{S}}|\mathbf{F}^L_S, \mathbf{Z}^{L-1})d\mathbf{F}^L_S\left(\prod_{l=2}^{L-1} q(\mathbf{Z}^l|\mathbf{Z}^{l-1}\right)q(\mathbf{Z}^1|\mathbf{X})d\left\{\mathbf{Z}^l\right\}_{l=1}^{L-1}$$

Using Bayes theorem to rewrite first term in integral

$$\cancel{\mathcal{Z}} * q(\mathbf{F}^L_{\bar{S}}) = \int (\cancel{\mathcal{Z}} q(\mathbf{F}^L_S|\mathbf{y}_S))p(\mathbf{F}^L_{\bar{S}}|\mathbf{F}^L_S, \mathbf{Z}^{L-1})d\mathbf{F}^L_S\left(\prod_{l=2}^{L-1} q(\mathbf{Z}^l|\mathbf{Z}^{l-1}\right)q(\mathbf{Z}^1|\mathbf{X})d\left\{\mathbf{Z}^l\right\}_{l=1}^{L-1}$$

$$q(\mathbf{F}^L_{\bar{S}}) = \int (q(\mathbf{F}^L_S|\mathbf{y}_S)p(\mathbf{F}^L_{\bar{S}}|\mathbf{F}^L_S, \mathbf{Z}^{L-1}))d\mathbf{F}^L_S\left(\prod_{l=2}^{L-1} q(\mathbf{Z}^l|\mathbf{Z}^{l-1}\right)q(\mathbf{Z}^1|\mathbf{X})d\left\{\mathbf{Z}^l\right\}_{l=1}^{L-1}$$
$$= \int q(\mathbf{F}^L_{\bar{S}}|\mathbf{Z}^{L-1}, \mathbf{y}_S)\left(\prod_{l=2}^{L-1} q(\mathbf{Z}^l|\mathbf{Z}^{l-1}\right)q(\mathbf{Z}^1|\mathbf{X})d\left\{\mathbf{Z}^l\right\}_{l=1}^{L-1}$$

where,

$$q(\mathbf{F}^L_{\bar{S}}|\mathbf{Z}^{L-1}, \mathbf{y}_S) = \int q(\mathbf{F}^L_S|\mathbf{y}_S)p(\mathbf{F}^L_{\bar{S}}|\mathbf{F}^L_S, \mathbf{Z}^{L-1})d\mathbf{F}^L_S$$
$$= \mathcal{N}(\mathbf{K}^L_{\bar{S},S}(\mathbf{K}^L_{S,S})^{-1}\hat{\mu}^L_S, \mathbf{V}^L_{\bar{S}} + \mathbf{K}^L_{\bar{S},S}(\mathbf{K}^L_{S,S})^{-1}\hat{\Sigma}^L_S(\mathbf{K}^L_{\bar{S},S}(\mathbf{K}^L_{S,S})^{-1})^T)$$

for $2 <= l <= L$

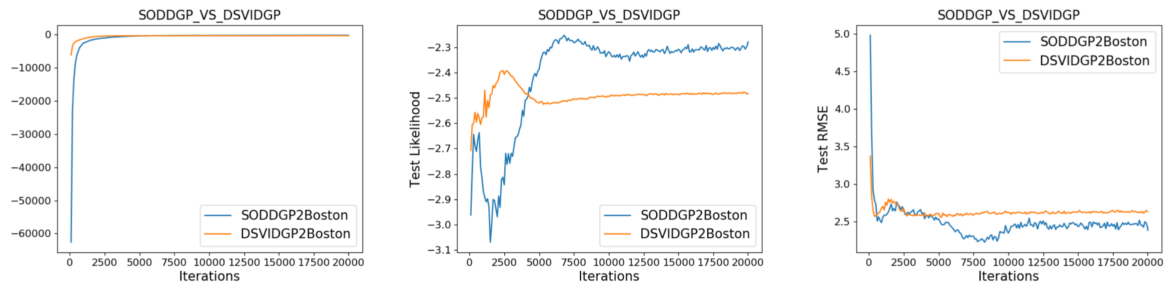
$$q(\mathbf{Z}^{l-1}|\mathbf{Z}^{l-2}) = \int p(\mathbf{Z}^{l-1}|\mathbf{F}^{l-1})p(\mathbf{F}^{l-1}_{\bar{S}}|\mathbf{F}^{l-1}_S, \mathbf{Z}^{l-2})q(\mathbf{F}^{l-1}_S)d\mathbf{F}^{l-1}$$
$$= \prod_{d=1}^{D^{l-1}} \int p(Z^{l-1}_{:,d}|\mathbf{f}^{l-1}_{:,d})p(\mathbf{f}^{l-1}_{\bar{S},d}|\mathbf{f}^{l-1}_{S,d}, \mathbf{Z}^{l-2})q(\mathbf{f}^{l-1}_{S,d})d\mathbf{f}^{l-1}_{:,d}$$
$$= \prod_{d=1}^{D^{l-1}} \int p(Z^{l-1}_{:,d}|\mathbf{f}^{l-1}_{:,d})q(\mathbf{f}^{l-1}_{:,d}|\mathbf{Z}^{l-2})d\mathbf{f}^{l-1}_{:,d}$$
$$= \prod_{d=1}^{D^{l-1}} \int \mathcal{N}(Z^{l-1}_{:,d}|\mathbf{f}^{l-1}_{:,d}, \sigma^2\mathbf{I})\mathcal{N}(\mathbf{f}^{l-1}_{:,d}|\mu^{l-1}_{:,d}, \Sigma^{l-1}_{:,d})d\mathbf{f}^{l-1}_{:,d}$$
$$= \prod_{d=1}^{D^{l-1}} \mathcal{N}(Z^{l-1}_{:,d}|\mu^{l-1}_{:,d}, \sigma^2\mathbf{I} + \Sigma^{l-1}_{:,d})$$

where,

$$\mu^{l-1}_{:,d} = \begin{bmatrix} \mu^{l-1}_{S,d} \\ \mathbf{K}^{l-1}_{\bar{S},S}(\mathbf{K}^{l-1}_{S,S})^{-1}\mu^{l-1}_{S,d} \end{bmatrix}. \quad ; \quad \Sigma^{l-1}_{:,d} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$
$$A = \Sigma^{l-1}_{S,d} \quad ; \quad B = \Sigma^{l-1}_{S,d}(\mathbf{K}^{l-1}_{S,S})^{-1}\mathbf{K}^{l-1}_{S,\bar{S}} \quad ; \quad C = (\Sigma^{l-1}_{S,d}(\mathbf{K}^{l-1}_{S,S})^{-1}\mathbf{K}^{l-1}_{S,\bar{S}})^T$$
$$D = \mathbf{V}^{l-1}_{\bar{S}} + \mathbf{K}^{l-1}_{\bar{S},S}(\mathbf{K}^{l-1}_{S,S})^{-1}\Sigma^{l-1}_{S,d}(\mathbf{K}^{l-1}_{S,S})^{-1}\mathbf{K}^{l-1}_{S,\bar{S}}$$

3 EXPERIMENTAL RESULTS

We show the change in objective function (ELBO) on the training split, and log-likelihood and root mean square error on the test split for the Boston data in Figure 7 over the training iterations. We can clearly see that the proposed SoD-DGP achieves a better performance than DSVI-DGP.



(a) ELBO (b) Test Log-likeihood (c) Test RMSE

Figure 7: ELBO, test log-likelihood and test RMSE on Boston housing

References

Andreas Damianou and Neil Lawrence. Deep Gaussian processes. In *International Conference on Artificial Intelligence and Statistics*, pages 207–215, 2013.

Hugh Salimbeni and Marc Deisenroth. Doubly stochastic variational inference for deep Gaussian processes. In *Advances in Neural Information Processing Systems 30*, pages 4588–4599. 2017.