Missing Data Imputation and Acquisition with Deep Hierarchical Models and Hamiltonian Monte Carlo

Ignacio Peis[1*], Chao Ma[2,3], José Miguel Hernández-Lobato[2]

[1]Universidad Carlos III de Madrid, Spain [2]University of Cambridge, UK [3]Microsoft Research Cambridge, UK

Universidad Carlos III de Madrid
UNIVERSITY OF CAMBRIDGE
NEURAL INFORMATION PROCESSING SYSTEMS
New Orleans, USA, 2022

Summary

- ☐ **Goal**: **impute, predict, acquire information** from heterogeneous incomplete data using $p(x_i, \boldsymbol{y} \mid \boldsymbol{x}_O) \approx \mathbb{E}_{q^{(T)}(\boldsymbol{h}|\boldsymbol{x}_O)}[p(x_i, \boldsymbol{y} \mid \boldsymbol{h})]$, where samples from $q^{(T)}(\boldsymbol{h} \mid \boldsymbol{x}_O)$ approximately follow the true posterior.
- ☐ We present **HH-VAEM**, a **deep hierarchical model** for handling mixed-type incomplete data with **automatically tuned HMC**.
- ☐ We propose a **sampling-based strategy for feature acquisition** that benefits from the improved inference of HH-VAEM.

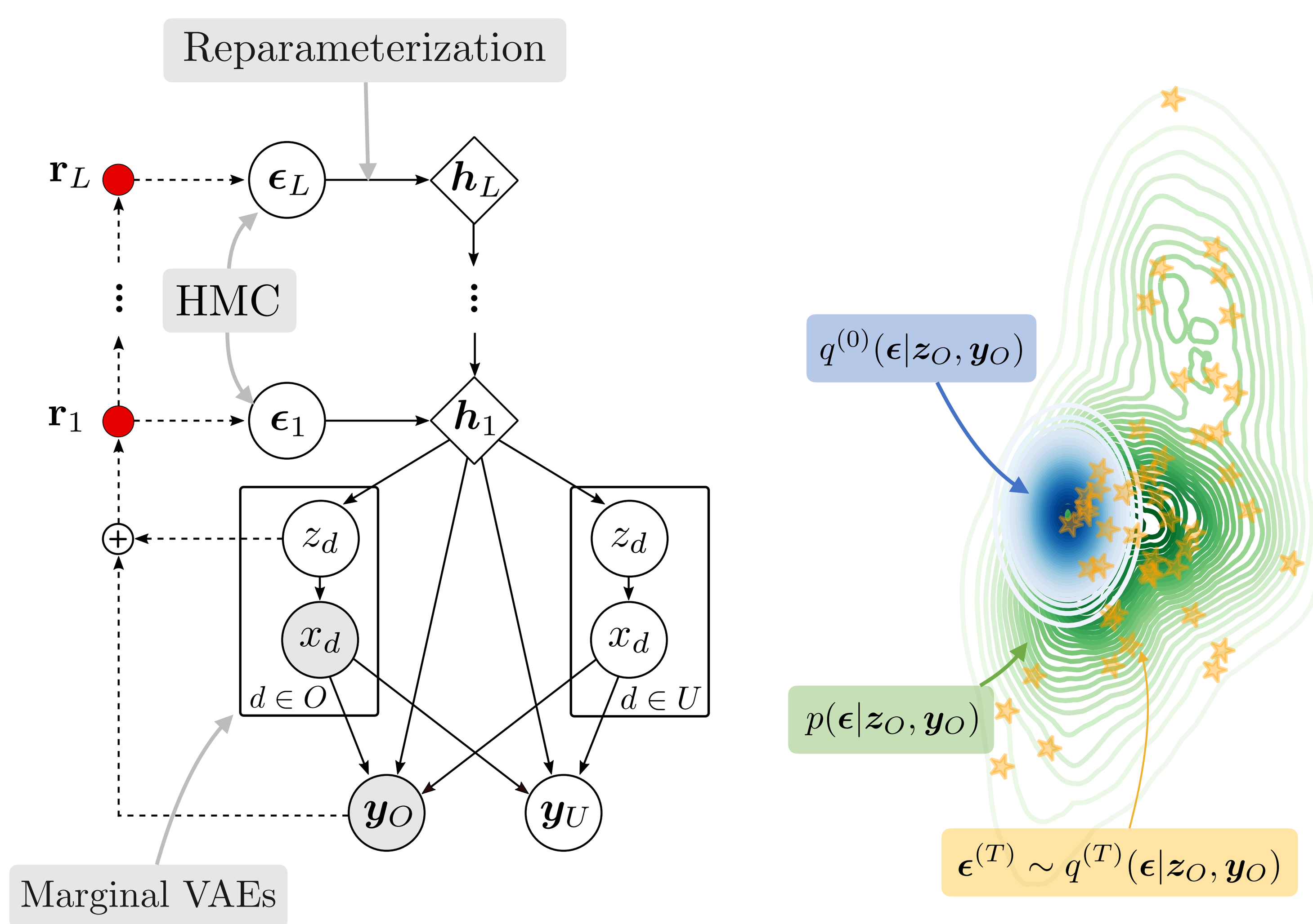


Figure 1: HH-VAEM model (left) and HMC example (right).

Hierarchical VAE and HMC

- Increase model flexibility with a **hierarchical latent space** $\boldsymbol{h} = \{\boldsymbol{h}_1, ..., \boldsymbol{h}_L\}$ where $\boldsymbol{h}_l \sim p(\boldsymbol{h}_l|\boldsymbol{h}_{l+1})$.
- **Reparameterization** for reducing strong curvature regions where $\nabla_{\boldsymbol{h}_{1:L}} \log p^*(\boldsymbol{h}_{1:L}) \uparrow\uparrow$:

$$\boldsymbol{h}_l = f_{\mu_l}(\boldsymbol{h}_{l+1}) + f_{\sigma_l}(\boldsymbol{h}_{l+1}) \cdot \boldsymbol{\epsilon}_l$$

- Improving posterior approximation by means of **well-posed HMC** to sample $\boldsymbol{\epsilon}^{(T)} \sim q^{(T)}(\boldsymbol{\epsilon}|\boldsymbol{z}_O, \boldsymbol{y}_O)$.
- **Automatically tune HMC hyperparams** for efficiently explore posterior distribution.

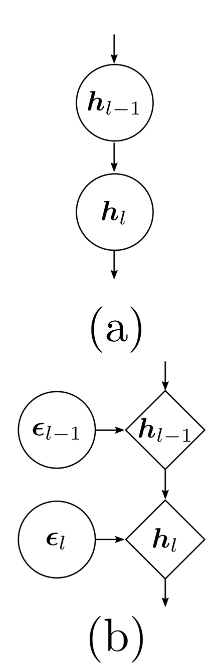


Figure 2: AR (a), Reparam. (b).

Motivation

Low bias samples from HMC and hierarchically enriched model improve the three considered tasks:

- ☐ **Imputation**: $p(\boldsymbol{x}_U|\boldsymbol{x}_O) \approx \mathbb{E}_{q^{(T)}(\boldsymbol{\epsilon}|\boldsymbol{x}_O)}[p(\boldsymbol{x}_U|\boldsymbol{\epsilon})]$.
- ☐ **Prediction**: $p(\boldsymbol{y}|\boldsymbol{x}_O) \approx \mathbb{E}_{q^{(T)}(\boldsymbol{\epsilon}|\boldsymbol{x}_O)}[p(\boldsymbol{y}|\boldsymbol{\epsilon}, \boldsymbol{x}_O, \hat{\boldsymbol{x}}_U)]$.
- ☐ **Sampling-based active learning**.

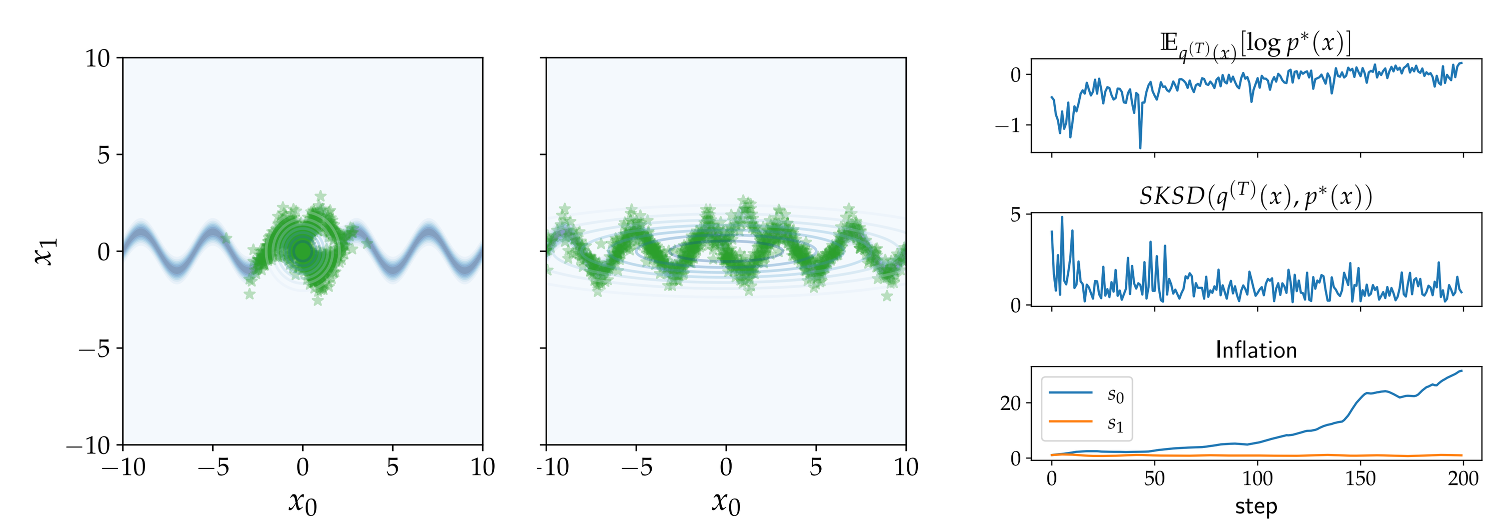
Training

1. Train **marginal VAEs** $\{\theta_d, \gamma_d\}_{d=1}^D$ to standarize heterogeneous data.
2. **Pre-train dependency VAE** $\{\psi, \theta_z, \theta_y\}$ using $\mathcal{L}_{VI}$.
3. **Fine-tune** encoder using $\mathcal{L}_{VI}$, decoder, predictor and **HMC hyperparams** ϕ using $\mathcal{L}_{HMC}$ with initial proposal $q^{(0)}$ inflated by $\boldsymbol{s}$, and **inflation parameter** $\boldsymbol{s}$ using $\mathcal{L}_{SKSD}$.

$$\mathcal{L}_{VI}(\boldsymbol{z}_O, \boldsymbol{y}_O; \{\theta, \psi\}) = \mathbb{E}_{q_\psi}\left[\log \frac{p_\theta(\boldsymbol{z}_O, \boldsymbol{y}_O, \boldsymbol{\epsilon})}{q_\psi(\boldsymbol{\epsilon}|\boldsymbol{z}_O, \boldsymbol{y}_O)}\right]$$

$$\mathcal{L}_{HMC}(\boldsymbol{z}_O, \boldsymbol{y}_O; \{\theta, \phi\}) = \mathbb{E}_{q_\phi^{(T)}(\boldsymbol{\epsilon})}[\log p_\theta(\boldsymbol{z}_O, \boldsymbol{y}_O, \boldsymbol{\epsilon})]$$

$$\mathcal{L}_{SKSD}(\boldsymbol{z}_O, \boldsymbol{y}_O; \boldsymbol{s}) = \text{SKSD}\left(q_\phi^{(T)}(\boldsymbol{\epsilon}|\boldsymbol{z}_O, \boldsymbol{y}_O; \boldsymbol{s}),\ p(\boldsymbol{\epsilon}|\boldsymbol{z}_O, \boldsymbol{y}_O)\right)$$



(a) Before training (b) After training (c) Training objectives

Figure 3: HMC training example.

Sampling-Based Active Learning

- **Sampling-based** information reward estimator for acquiring variables x_i to accurately predict $\boldsymbol{y}$:

$$\hat{R}_I = \hat{I}(\boldsymbol{y}, x_i \mid \boldsymbol{x}_O) \approx \sum_{ij} \hat{p}_{x_i, \boldsymbol{y}|\boldsymbol{x}_O}(i,j) \log \frac{\hat{p}_{x_i, \boldsymbol{y}|\boldsymbol{x}_O}(i,j)}{\hat{p}_{x_i|\boldsymbol{x}_O}(i)\ \hat{p}_{\boldsymbol{y}|\boldsymbol{x}_O}(j)}$$

that uses HMC in $p(x_i, \boldsymbol{y} \mid \boldsymbol{x}_O) \approx \mathbb{E}_{q^{(T)}(\boldsymbol{\epsilon}|\boldsymbol{x}_O)}[p(x_i, \boldsymbol{y} \mid \boldsymbol{\epsilon})]$, instead of the Gaussian alternative (Figure 1).

Experiments

		VAE(M)	MIWAE(M)	H-VAE(M)	HMC-VAE(M)	**HH-VAE(M**
$-\log p(\boldsymbol{x}_U\|\boldsymbol{x}_O)$	MNIST	0.124 ± 0.001	0.121 ± 0.001	0.119 ± 0.001	0.101 ± 0.004	$\mathbf{0.094 \pm 0.00}$
	F-MNIST	0.162 ± 0.002	0.160 ± 0.002	0.156 ± 0.002	0.150 ± 0.002	$\mathbf{0.144 \pm 0.00}$
	Avocado	1.89 ± 0.01	1.92 ± 0.04	1.89 ± 0.01	1.89 ± 0.02	$\mathbf{1.88 \pm 0.0}$
	Bank	2.84 ± 0.07	2.74 ± 0.05	2.82 ± 0.06	2.69 ± 0.05	$\mathbf{2.63 \pm 0.0}$
$-\log p(\boldsymbol{y}\|\boldsymbol{x}_O)$	MNIST	0.153 ± 0.009	0.151 ± 0.007	0.146 ± 0.006	0.067 ± 0.007	$\mathbf{0.056 \pm 0.01}$
	F-MNIST	0.501 ± 0.012	0.496 ± 0.008	0.494 ± 0.007	0.357 ± 0.060	$\mathbf{0.337 \pm 0.06}$
	Avocado	1.18 ± 0.02	1.15 ± 0.03	1.18 ± 0.02	1.12 ± 0.03	$\mathbf{1.10 \pm 0.0}$
	Bank	0.56 ± 0.06	0.51 ± 0.03	0.50 ± 0.03	0.52 ± 0.02	$\mathbf{0.49 \pm 0.0}$
Accuracy	MNIST	0.953 ± 0.004	0.953 ± 0.003	0.953 ± 0.003	0.978 ± 0.003	$\mathbf{0.981 \pm 0.00}$
	F-MNIST	0.824 ± 0.005	0.824 ± 0.004	0.824 ± 0.004	0.869 ± 0.015	$\mathbf{0.876 \pm 0.01}$

Table 1: Imputation and prediction evaluation.

$$\log p(\boldsymbol{x}_U|\boldsymbol{x}_O) \approx \log \mathbb{E}_{q^{(T)}(\boldsymbol{\epsilon}|\boldsymbol{x}_O)}[p(\boldsymbol{x}_U|\boldsymbol{\epsilon})] \approx \log \frac{1}{k}\sum_i^k p(\boldsymbol{x}_U|\boldsymbol{\epsilon}_i^{(T)})$$

$$\log p(\boldsymbol{y}|\boldsymbol{x}_O) \approx \log \mathbb{E}_{q^{(T)}(\boldsymbol{\epsilon}|\boldsymbol{x}_O)}[p(\boldsymbol{y}|\boldsymbol{\epsilon}, \hat{\boldsymbol{x}})] \approx \log \frac{1}{k}\sum_i^k p(\boldsymbol{y}|\boldsymbol{\epsilon}_i^{(T)}, \hat{\boldsymbol{x}}_i)$$

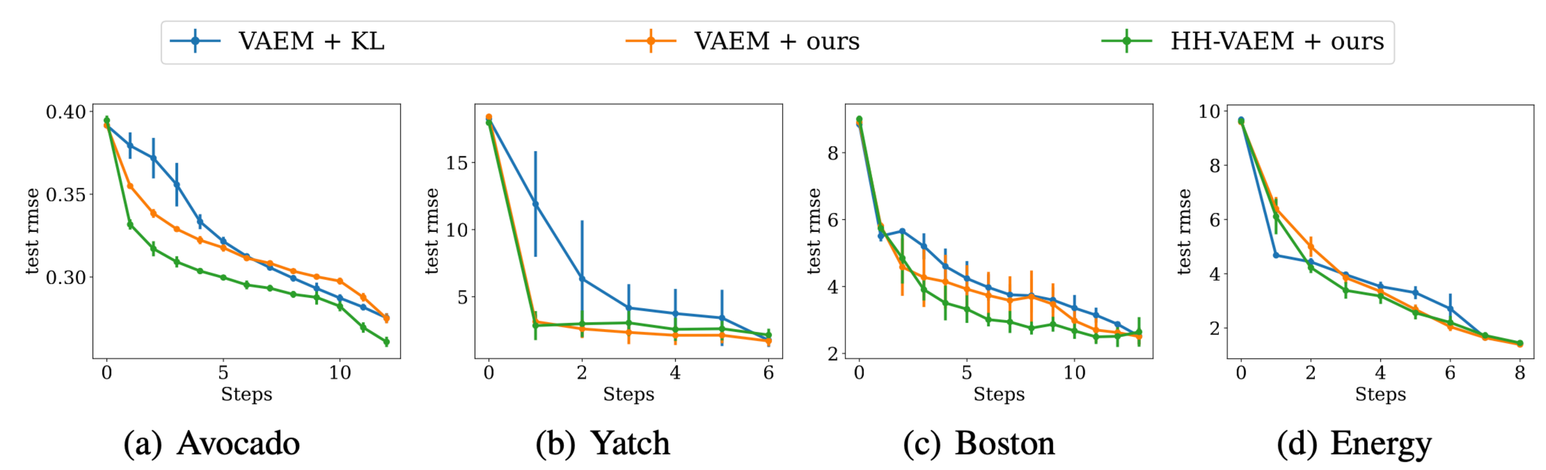


Figure 4: SAIA curves. RMSE vs number of discovered features.

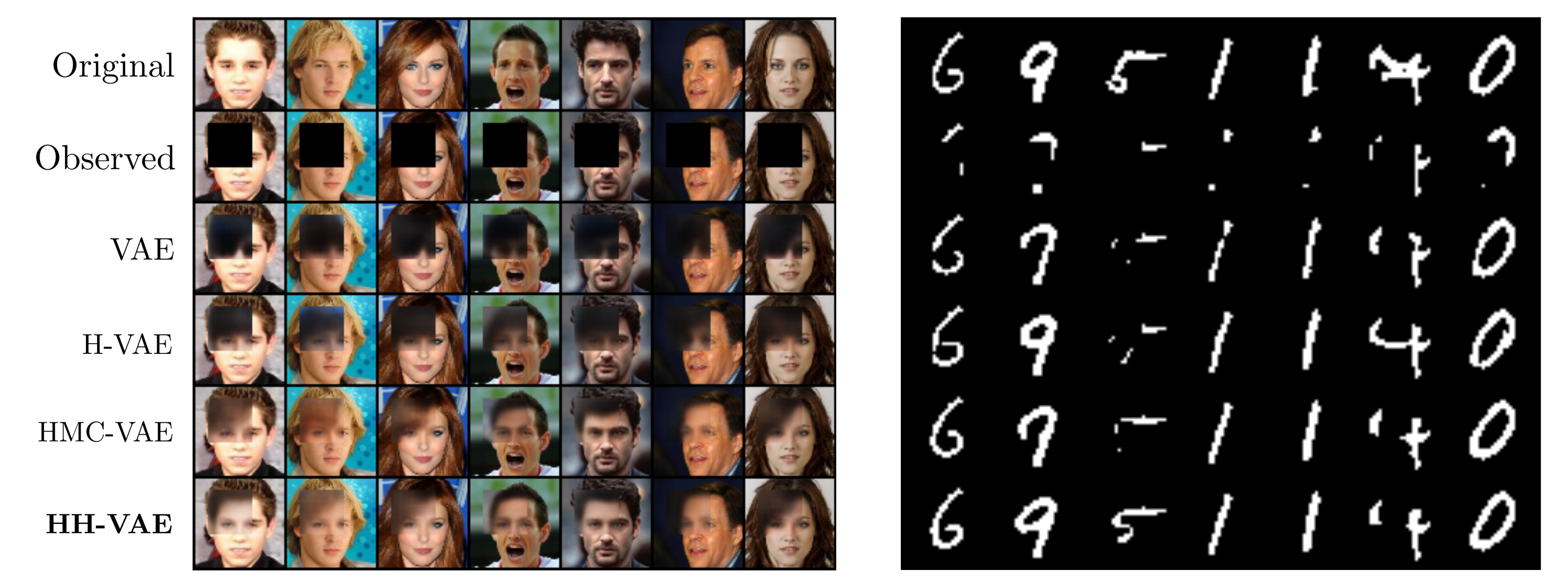


Figure 5: Image conditional inpainting on CelebA (left) and MNIST (right)

Acknowledgements

Ignacio Peis acknowledges support from Spanish government MICIU under grants FPU18/00516, RTI2018-099655-B-I00, EST21/00467 and PID2021-123182OB-I00 and from Comunidad de Madrid under grants Y2018/TCS-4705 PRACTICO-CM and the Multiannual Agreement with UC3M in the line of Excellence of University Professors (EPUC3M27), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation). José Miguel Hernández-Lobato acknowledges support from Boltzbit.

* Corresponding author: ipeis@tsc.uc3m.es

github.com/ipeis/HH-VAEM