Generic Unsupervised Optimization for a Latent Variable Model With Exponential Family Observables

Hamid Mousavi, Jakob Drefs, Florian Hirschberger, Jörg Lücke

ML Lab, University of Oldenburg, Germany

NeurIPS, 2024

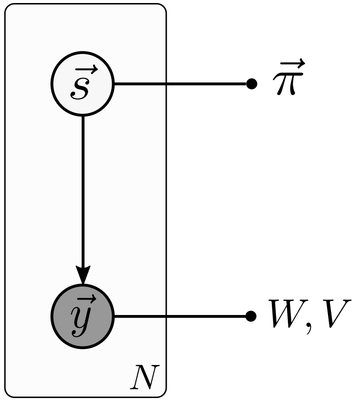
Proposed Latent Variable Models

- We propose Exponential Family Maximal Causes Analysis (EF-MCA).
- Main features of the proposed models are:
  - Binary latent variables
  - Exponential family observables
  - Maximum non-linear superposition

$$p(\vec{s}\,|\,\Theta) = \prod_{h=1}^{H} \pi_h^{s_h} (1 - \pi_h)^{1-s_h}$$

$$p(\vec{y}\,|\,\vec{s},\Theta) = \prod_{d=1}^{D} p\big(y_d; \vec{\eta}_d(\vec{s},\Theta)\big)$$

$$\text{where} \;\; p(y;\vec{\eta}) = h(y)\exp\big(\vec{\eta}^T \vec{T}(y) - A(\vec{\eta})\big)$$



$$s_h \sim \mathrm{Bern}(s_h; \pi_h)$$

$$y_d \sim p\big(y_d; \vec{\eta}_d(\vec{s},\Theta)\big)$$

$$\vec{\eta}_d(\vec{s},\Theta) = \vec{\Phi}\big(\bar{W}_d(\vec{s},\Theta), \bar{V}_d(\vec{s},\Theta)\big)$$

Lücke and Sahani, JMLR 2006; Lücke and Eggert, JMLR 2010; Mousavi*, Drefs* and Lücke, LOD 2020; Mousavi et al., Entropy 2021.

Generic Parameter Optimization

Theorem. Consider the proposed family of generative models with the given non-linear link function and assume a two-parameter distribution of the exponential family. Then the derivatives of the free energy (a.k.a. ELBO) w.r.t. parameters $\theta = (W, V, \vec{\pi})$ is zero, if for each d and h, we let:

$$W_{dh} = \frac{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s},\Theta) \rangle_{q^{(n)}} T_1(y_d^{(n)})}{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s},\Theta) \rangle_{q^{(n)}}}$$

$$V_{dh} = \frac{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s},\Theta) \rangle_{q^{(n)}} T_2(y_d^{(n)})}{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s},\Theta) \rangle_{q^{(n)}}}$$

$$\pi_h = \frac{1}{N} \sum_{n=1}^{N} \langle s_h \rangle_{q^{(n)}}$$

$$M_{dh}(\Theta) := F(W_{dh}, V_{dh}) \quad \text{where} \quad F(w,v) = \langle y \rangle_{p(y;\vec{\Phi}(w,v))}$$

$$\bar{W}_d(\vec{s},\Theta) := W_{dh(d,\vec{s},\Theta)}, \quad \bar{V}_d(\vec{s},\Theta) := V_{dh(d,\vec{s},\Theta)}$$

$$\text{where} \quad h(d,\vec{s},\Theta) := \text{argmax}_h \{ M_{dh}(\Theta)\, s_h \}.$$

$$\mathcal{A}_{dh}(\vec{s},\Theta) := \begin{cases} 1 & h = h(d,\vec{s},\Theta) \\ 0 & \text{otherwise} \end{cases}$$

Generic Parameter Optimization – EM Algorithm

- Initialize the model parameters
- Compute the posteriors and to allow for the large-scaled applications of the models, we use variational approximations in the form of truncated posteriors (we leverage the newly established method called Evolutionary Variational Optimization, EVO):

$$q^{(n)}(\vec{s} \mid \mathcal{K}, \Theta) := \frac{p(\vec{s}, \vec{y}^{(n)} \mid \Theta)}{\sum_{\vec{s}' \in \mathcal{K}^{(n)}} p(\vec{s}', \vec{y}^{(n)} \mid \Theta)} \delta(\vec{s} \in \mathcal{K}^{(n)})$$

- Update the model parameters using:

$$W_{dh}^{\mathrm{new}} = \frac{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s}, \Theta^{\mathrm{old}}) \rangle_{q^{(n)}} T_1(y_d^{(n)})}{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s}, \Theta^{\mathrm{old}}) \rangle_{q^{(n)}}}$$

$$V_{dh}^{\mathrm{new}} = \frac{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s}, \Theta^{\mathrm{old}}) \rangle_{q^{(n)}} T_2(y_d^{(n)})}{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s}, \Theta^{\mathrm{old}}) \rangle_{q^{(n)}}}$$

$$\pi_h^{\mathrm{new}} = \frac{1}{N} \sum_{n=1}^{N} \langle s_h \rangle_{q^{(n)}}$$

in general →

$$(W_{dh}^{(l)})^{\mathrm{new}} = \frac{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s}, \Theta^{\mathrm{old}}) \rangle_{q^{(n)}} T_l(y_d^{(n)})}{\sum_{n=1}^{N} \langle \mathcal{A}_{dh}(\vec{s}, \Theta^{\mathrm{old}}) \rangle_{q^{(n)}}}$$

$$\pi_h^{\mathrm{new}} = \frac{1}{N} \sum_{n=1}^{N} \langle s_h \rangle_{q^{(n)}}$$

Experiments – Poisson Denoising

- Comparison of the PSNR (in terms of dB) and ELBO values for Poisson denoising benchmarks at peak value of 1.

D = 400, *H* = 100, and 100 variational EM iterations

Model	Measure	House	Camera	Peppers
BSC	PSNR	22.40	19.91	19.47
	ELBO	-463.70	-418.86	-431.05
ES3C	PSNR	22.37	19.78	19.60
	ELBO	-458.23	-414.08	-426.98
TVAE	PSNR	22.84	20.13	19.57
	ELBO	-459.05	-416.48	-428.83
MCA	PSNR	22.66	19.78	19.33
	ELBO	-457.14	-413.17	-426.07
P-MCA	PSNR	**23.54**	**20.53**	**20.14**
	ELBO	**-392.13**	**-331.15**	**-355.55**

Noisy ES3C MCA

BSC TVAE P-MCA

Henniges et al., LVA 2010; Sheikh et al., JMLR 2014; Drefs et al., JMLR 2022; Sheikh et al., PLOS 2019.