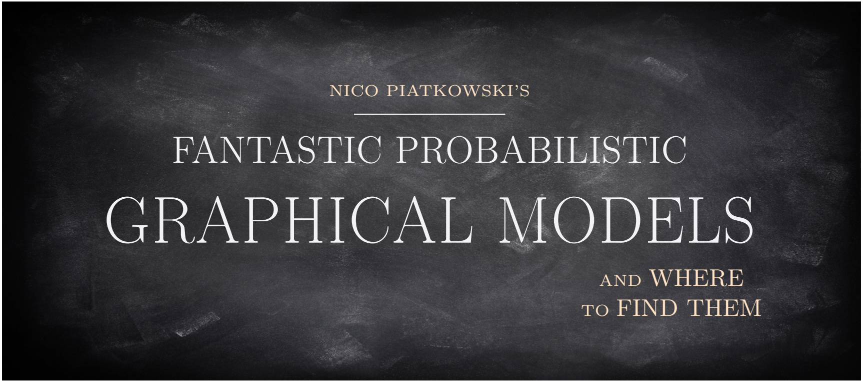
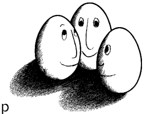


Probabilistic Graphical Models

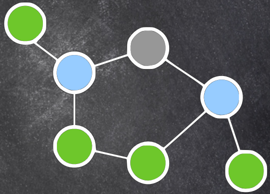
DoDSc colloquium, Jul 11, 2019



Dr. Nico Piatkowski

Artificial Intelligence Group

Undirected Graphical Models



Goals:

- Consistent estimation for **high-dimensional, discrete** random variables
- Learning from **unreasonable large** amounts of data

Some facts:

- Graphical models are in the exponential family

$$\mathbb{P}(\mathbf{X} \quad \mathbf{x}) \quad \underbrace{\frac{1}{Z(\boldsymbol{\theta})} \prod_{C \in \mathcal{C}} \exp(\langle \boldsymbol{\theta}_C, \phi_C(\mathbf{x}_C) \rangle)}_{\text{Factorization over cliques}} \quad \underbrace{\exp(\langle \boldsymbol{\theta}, \phi(\mathbf{x}) \rangle - A(\boldsymbol{\theta}))}_{\text{Exponential family member}}$$

- Covers

[Ising, Pitman, Fischer, Lauritzen, Besag, Hammersley, Clifford, . . . , 1925-today]

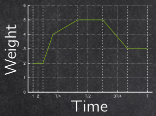
Compressed Spatio-Temporal Models

Motivation:

- Naive models overfit due to **many redundant parameters**
- l_1-regularization **destroys conditional independence structure**
- Existing techniques **do not work** without redundancies

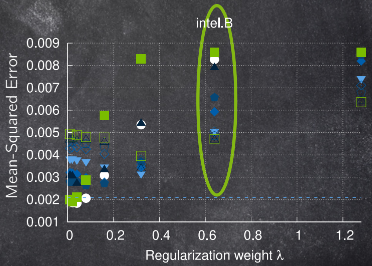
Contributions:

- **New:** Lossless representation via **ST reparametrization**



$$\boldsymbol{\theta}_C(t) \quad \sum_{i\ \ 1}^{t} \alpha_{i,t} \underbrace{\boldsymbol{\Delta}_C(i)}_{\text{New parameters}}$$

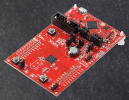
- **New:** Regularization of ST reparametrized model **removes** redundancies while cond. independence structure is kept intact



[Piatkowski, Lee, Morik, Machine Learning, 2013]

Integer Exponential Families

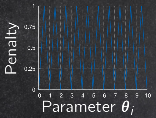
Motivation:



- Integer arithmetic **requires less resources**
- Learning and probabilistic inference is **inherently real-valued**
- Integer parameters **cannot be learned directly**

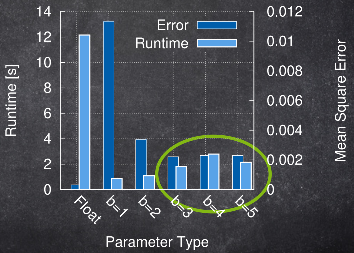
Contributions:

- **New: Integer regularization** enforces integrality of parameters



$$R_{\mathrm{int}}(\boldsymbol{\theta}) = \sum_{i=1}^{d} \underbrace{1 - |1 - 2(\lceil \boldsymbol{\theta}_i \rfloor - \boldsymbol{\theta}_i)|}_{\text{Penalty}}$$

- **New:** Base-2 exponential family with integer-valued probabilistic inference via **bit-length propagation** (bounded KL)



[Piatkowski, Lee, Morik, Neurocomputing, 2016]

Quadrature-based Probabilistic Inference

Motivation:

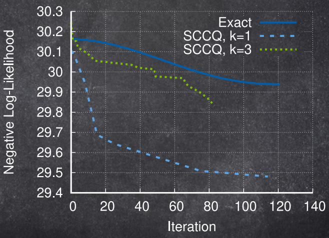
- Exact probabilistic inference has **exponential worst-case complexity**
- Variational inference **cannot guarantee approximation error**
- **Convergence** of Markov chains / dependence of MCMC samples

Contributions:

- **New:** Clenshaw-Curtis quadrature to approximate normalization/marginals **without structural approximation**

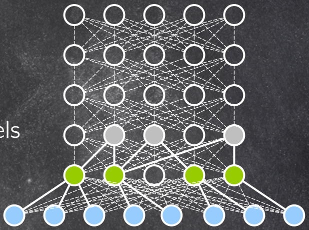
$$\hat{Z}^k(\boldsymbol{\theta}) \quad \int_{\mathcal{X}} \hat{f}_k(\mathbf{x})\,\mathrm{d}\,\nu(\mathbf{x}) \quad \mathbb{E}_{\mathrm{J},l}\left[\hat{Z}^k_{\mathrm{J},l}(\boldsymbol{\theta})\right]$$

- **New:** Approximation error depends on $\|\boldsymbol{\theta}\|_1$ which is controlled via l_1-regularization: $R_1(\boldsymbol{\theta}) \quad \|\boldsymbol{\theta}\|_1$



[Piatkowski, Morik, ICML, 2016], [Piatkowski, Morik, UAI, 2018]

Deep Boltzmann Trees



Motivation:

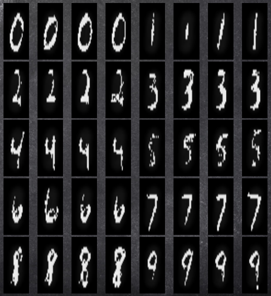
- **Heuristic hand-tuning** of deep architectures
- Known insights predict **unreasonable deep** models
- Contrastive divergence **is not consistent**

Contributions:

- **New:** Tree structured "deep" latent variable models with **guaranteed universal approximation** property

  $$\forall \epsilon > 0 : \exists \mathrm{DBT} : \mathrm{KL}[\mathbb{P}_G \| \sum_{\mathbf{h}} \mathbb{P}_{\mathrm{DBT}}(\,, \mathbf{h})] \leq \epsilon$$

- **New:** Proof that fixed-width DBMs of **depth two are universal approximators** if latent space is large enough



[Piatkowski, ECMLPKDD, 2019]