Correlated Variational Auto-Encoders

Da Tang[1] Dawen Liang[2] Tony Jebara[1,2] Nicholas Ruozzi[3]

[1]Columbia University

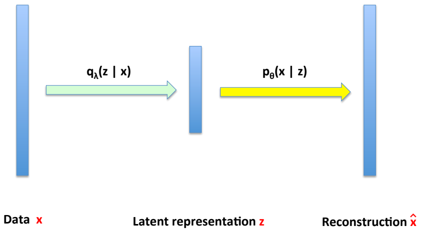
[2]Netflix Inc.

[3]The University of Texas at Dallas

June 11, 2019

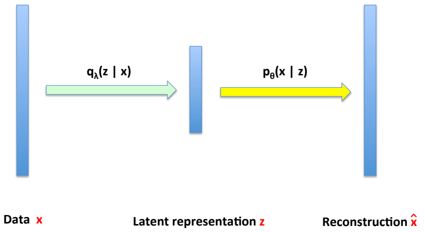
Variational Auto-Encoders (VAEs)

- Learn stochastic low dimensional latent representations for high dimensional data:



Variational Auto-Encoders (VAEs)

- Learn stochastic low dimensional latent representations for high dimensional data:



- Model the likelihood and the inference distribution independent among data points in the objective (the ELBO):

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \sum_{i=1}^{n} (\mathbb{E}_{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_i|\boldsymbol{x}_i)} [\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_i|\boldsymbol{z}_i)] - \mathrm{KL}(q_{\boldsymbol{\lambda}}(\boldsymbol{z}_i|\boldsymbol{x}_i)||p_0(\boldsymbol{z}_i))).$$

Motivation

- VAEs assume the prior is i.i.d. among data points.

Motivation

- VAEs assume the prior is i.i.d. among data points.
- If we know information about correlations between data points (e.g., networked data), we can incorporate it into the generative process of VAEs.

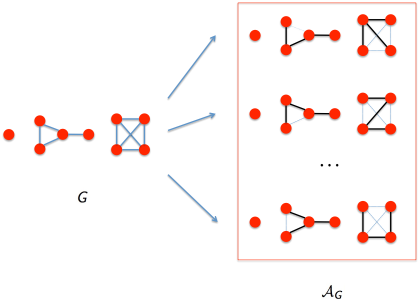
Learning with a Correlation Graph

- Given an undirected correlation graph $G = (V, E)$ for data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, where $V = \{v_1, \ldots, v_n\}$ and $E = \{(v_i, v_j) : \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are correlated}\}$.

Learning with a Correlation Graph

- Given an undirected correlation graph $G = (V, E)$ for data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, where $V = \{v_1, \ldots, v_n\}$ and $E = \{(v_i, v_j) : \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are correlated}\}$.
- Directly applying a correlated prior of $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$ on general undirected graphs is hard.

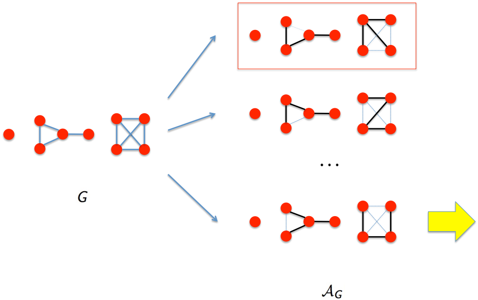
Correlated Priors



Define the prior of $\boldsymbol{z}$ as a uniform mixture over all *Maximal Acyclic Subgraphs* of G:

$$p_0^{\mathrm{corr}_g}(\boldsymbol{z}) = \frac{1}{|\mathcal{A}_G|} \sum_{G'=(V,E')\in\mathcal{A}_G} p_0^{G'}(\boldsymbol{z}).$$

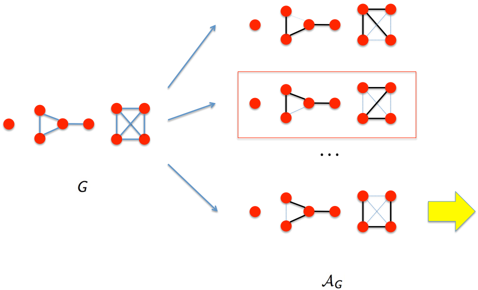
Correlated Priors



We apply a uniform mixture over acyclic subgraphs since we have closed-form correlated distributions for acyclic graphs:

$$p_0^{G'}(\boldsymbol{z}) = \prod_{i=1}^{n} p_0(\boldsymbol{z}_i) \prod_{(v_i,v_j)\in E'} \frac{p_0(\boldsymbol{z}_i,\boldsymbol{z}_j)}{p_0(\boldsymbol{z}_i)p_0(\boldsymbol{z}_j)}.$$

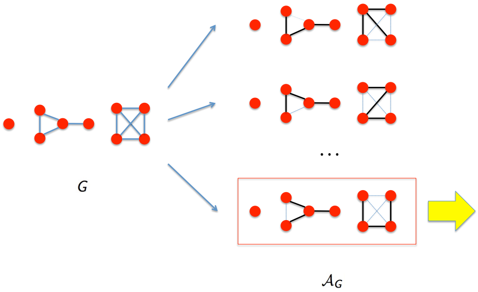
Correlated Priors



We apply a uniform mixture over acyclic subgraphs since we have closed-form correlated distributions for acyclic graphs:

$$p_0^{G'}(\mathbf{z}) = \prod_{i=1}^{n} p_0(\mathbf{z}_i) \prod_{(v_i,v_j)\in E'} \frac{p_0(\mathbf{z}_i, \mathbf{z}_j)}{p_0(\mathbf{z}_i)p_0(\mathbf{z}_j)}.$$

Correlated Priors



We apply a uniform mixture over acyclic subgraphs since we have closed-form correlated distributions for acyclic graphs:

$$p_0^{G'}(\boldsymbol{z}) = \prod_{i=1}^{n} p_0(\boldsymbol{z}_i) \prod_{(v_i,v_j)\in E'} \frac{p_0(\boldsymbol{z}_i, \boldsymbol{z}_j)}{p_0(\boldsymbol{z}_i)p_0(\boldsymbol{z}_j)}.$$

Inference with a Weighted Objective

Define a new ELBO for general graphs:

$$
\begin{aligned}
\log p_{\boldsymbol{\theta}}(\boldsymbol{x}) &= \log \mathbb{E}_{p_0^{\mathrm{corr}_g}(\boldsymbol{z})}[p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})] \\
&\geq \frac{1}{|\mathcal{A}_G|} \sum_{G' \in \mathcal{A}_G} \left(\mathbb{E}_{q_{\boldsymbol{\lambda}}^{G'}(\boldsymbol{z}|\boldsymbol{x})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})] - \mathrm{KL}(q_{\boldsymbol{\lambda}}^{G'}(\boldsymbol{z}|\boldsymbol{x}) || p_0^{G'}(\boldsymbol{z})) \right) \\
&:= \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\theta})
\end{aligned}
$$

where $q_{\boldsymbol{\lambda}}^{G'}$ is defined in the same way as for the priors:

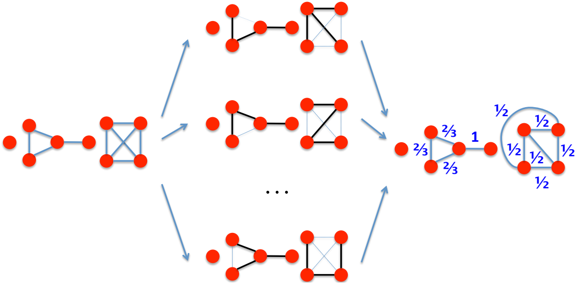
$$
q_{\boldsymbol{\lambda}}^{G'}(\boldsymbol{z}) = \prod_{i=1}^{n} q_{\boldsymbol{\lambda}}(\boldsymbol{z}_i|\boldsymbol{x}_i) \prod_{(v_i,v_j)\in E'} \frac{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_i, \boldsymbol{z}_j|\boldsymbol{x}_i, \boldsymbol{x}_j)}{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_i|\boldsymbol{x}_i) q_{\boldsymbol{\lambda}}(\boldsymbol{z}_j|\boldsymbol{x}_j)}.
$$

Inference with a Weighted Objective

- The loss function is intractable due to the potentially exponential many subgraphs.

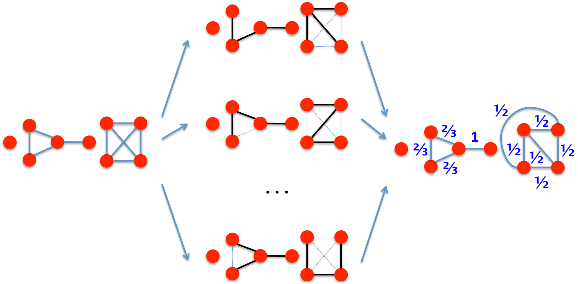
Inference with a Weighted Objective

- The loss function is intractable due to the potentially exponential many subgraphs.
- Represent the average loss on acyclic subgraphs as a weighted average loss on edges.



Inference with a Weighted Objective

- The loss function is intractable due to the potentially exponential many subgraphs.
- Represent the average loss on acyclic subgraphs as a weighted average loss on edges.



- The weighted loss is tractable. The weights can be computed from the pseudo-inverse of the Laplacian matrix of G.

Empirical Results

Table: Link prediction test NCRR

Method	Test NCRR
VAE	0.0052 ± 0.0007
GraphSAGE	0.0115 ± 0.0025
CVAE	$\mathbf{0.0171 \pm 0.0009}$

Table: Spectral clustering scores

Method	NMI scores
VAE	0.0031 ± 0.0059
GraphSAGE	0.0945 ± 0.0607
CVAE	$\mathbf{0.2748 \pm 0.0462}$

Table: User matching test RR

Method	Test RR
VAE	0.3498 ± 0.0167
CVAE	$\mathbf{0.7129 \pm 0.0096}$

Conclusion and Future Work

- CVAE accounts for correlations between data points that are known *a priori*. It can adopt a correlated variational density function to achieve a better variational approximation.

Conclusion and Future Work

- CVAE accounts for correlations between data points that are known *a priori*. It can adopt a correlated variational density function to achieve a better variational approximation.
- Future work includes extending to correlated VAEs with higher-order correlations.

Thanks!

Poster #219

Code available at https://github.com/datang1992/Correlated-VAEs.