Supplementary Information for ”Exact inference and learning in loopy cumulative distribution networks”

Jim C. Huang, Nebojsa Jojic and Christopher Meek
Microsoft Research
One Microsoft Way, Redmond, WA 98052

1 Nonparanormal distribution

The nonparanormal distribution $NPN(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{f})$ of [1] consists of a multivariate Gaussian distribution defined on nonlinearly transformed variables $f_\alpha(x_\alpha)$ such that the joint probability density function is given by

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{f}(\mathbf{x}); \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_\alpha |f'_\alpha(x_\alpha)| \tag{1}$$

where $\mathbf{f}$ is the vector whose elements are nonlinear transforms $f_\alpha(x_\alpha)$. The nonparanormal distribution corresponds to a Gaussian copula model with monotone and differentiable functions $f_\alpha(x_\alpha)$. We specified the functions $f_\alpha(x_\alpha)$ using the suggested parameterization of [1] whereby we set

$$f_\alpha(x_\alpha) = \tilde{\mu}_\alpha + \tilde{\sigma}_\alpha \Phi^{-1}(\tilde{F}_\alpha(x_\alpha)) \tag{2}$$

where Φ is the normal cumulative distribution function, $\tilde{\mu}_\alpha, \tilde{\sigma}_\alpha$ are the empirical mean and standard deviation for random variable X_α and $\tilde{F}_\alpha$ is the Winsorized estimator [1] of the CDF for X_α given by

$$\tilde{F}_\alpha(x_\alpha) = \begin{cases} \delta_n & \text{if } \hat{F}_\alpha(x_\alpha) < \delta_n \\ \hat{F}_\alpha(x_\alpha) & \text{if } \delta_n \le \hat{F}_\alpha(x_\alpha) \le 1 - \delta_n \\ 1 - \delta_n & \text{if } \hat{F}_\alpha(x_\alpha) > 1 - \delta_n \end{cases}$$

for empirical CDF $\hat{F}_\alpha(x_\alpha)$. The truncation constant δ_n was set according to [1] to the value $\delta_n = \frac{1}{4n^{1/4}\sqrt{\pi \log n}}$ where n is the size of the training set.

2 Multivariate logistic distribution

The multivariate logistic CDF [2] is given by

$$F(\mathbf{x}|\boldsymbol{\theta}) = \Big(1 + \sum_\alpha e^{-\frac{x_\alpha - \mu_\alpha}{\sigma_\alpha}}\Big)^{-(k+1)} \tag{3}$$

for $k > 0$ and $\sigma_\alpha > 0$. The corresponding PDF is given by

$$P(\mathbf{x}|\boldsymbol{\theta}) = \frac{k! e^{-\sum_\alpha \frac{x_\alpha - \mu_\alpha}{\sigma_\alpha}}}{\prod_\alpha \sigma_\alpha} F(\mathbf{x}|\boldsymbol{\theta}) \tag{4}$$

We estimated the above model by maximizing likelihood for $k = 1, \cdots, 10$ and selected the value of k that yielded the best average test likelihood.

3 Settings for learning CDN models

The disconnected CDN corresponds to a fully factorized Gumbel distribution

$$F(\mathbf{x}|\boldsymbol{\theta}) = \prod_{\alpha \in V} F(x_\alpha | \mu_\alpha, \sigma_\alpha) = \prod_{\alpha \in V} \exp(-e^{-\frac{x_\alpha - \mu_\alpha}{\sigma_\alpha}}) \tag{5}$$

with $\sigma_\alpha > 0$.

For learning the CDN models, we used first-order stochastic gradient descent methods with learning rates ν set to 0.05. Means were initialized to small random values and standard deviations were initialized to 1. The coefficient θ_s was initialized to 0.99. Parameters were then updated by a stochastic gradients algorithm with 100 epochs, with standard deviations constrained to be at least 0.1 and coefficients θ_s were constrained to be between 0.01 and 0.99. The learning rate was set to decay according to the rule $\nu \leftarrow \nu \frac{m}{m+1}$ for the m^{th} epoch.

References

[1] Liu, H., Lafferty, J. and Wasserman, L. (2009) The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research (JMLR)* **10**, 2295-2328.

[2] Malik, H.J. and Abraham, B. (1978) Multivariate logistic distributions. *Annals of Statistics* **1(3)**, 588-590.