Extending Stochastic EP to Latent Variable Models

Vera Johne, Yingzhen Li, Richard Turner

University of Cambridge

Motivation

- Latent Variable Models is a powerful approach to probabilistic modeling.
- Exact inference and learning on LVMs can be challenging (intractability).
- We desire computational and memory efficient approximate inference techniques in non-linear LVMs.

Latent Variable Models

- Many interesting real life problems (e.g. Face recognition, topic modeling, finance)

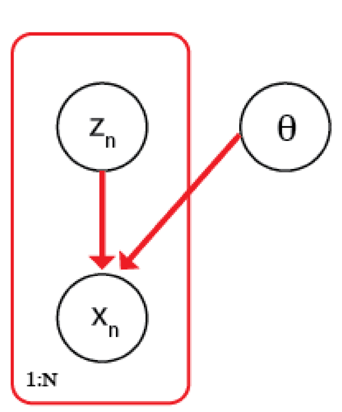
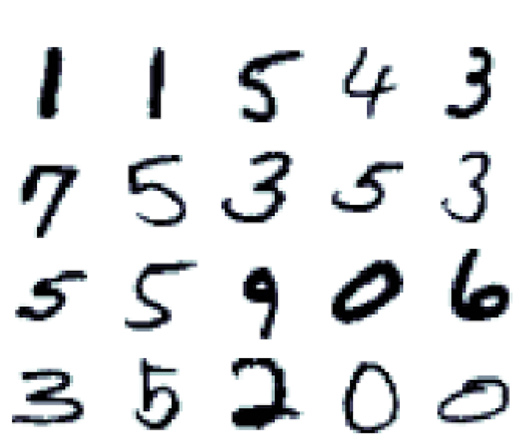


Figure 1: Example of LVM



Expectation Propagation

- Locally minimize reverse KL divergence
- Moment matching
- Factorisation Assumption:

$$p(D,\theta) = p(\theta)\prod_i f_i(\theta)$$

- Approximating factors: members of exponential family

$$q(\theta) = \frac{1}{Z}p(\theta)\prod_i \hat{f}_i(\theta)$$

Algorithm: until convergence

1. choose a factor f_n to refine:
2. compute cavity distribution
   $q_{-n}(\theta) \propto q(\theta)/f_n(\theta)$
3. compute tilted distribution
   $\tilde{p}_n(\theta) \propto p(x_n|\theta)q_{-n}(\theta)$
4. moment matching:
   $f_n(\theta) \leftarrow \texttt{proj}[\tilde{p}_n(\theta)]/q_{-n}(\theta)$
5. inclusion:
   $q(\theta) \leftarrow q_{-n}(\theta)f_n(\theta)$

Stochastic-EP

$$q(\theta) = \frac{1}{Z}p(\theta)\hat{f}(\theta)^N$$

Algorithm: until convergence

1. choose a datapoint $x_n \sim \mathcal{D}$:
2. compute cavity distribution
   $q_{-1}(\theta) \propto q(\theta)/f(\theta)$
3. compute tilted distribution
   $\tilde{p}_n(\theta) \propto p(x_n|\theta)q_{-1}(\theta)$
4. moment matching:
   $f_n(\theta) \leftarrow \texttt{proj}[\tilde{p}_n(\theta)]/q_{-1}(\theta)$
5. inclusion:
   $q(\theta) \leftarrow q_{-1}(\theta)f_n(\theta)$
6. *implicit update*:
   $f(\theta) \leftarrow f(\theta)^{1-\frac{1}{N}}f_n(\theta)^{\frac{1}{N}}$

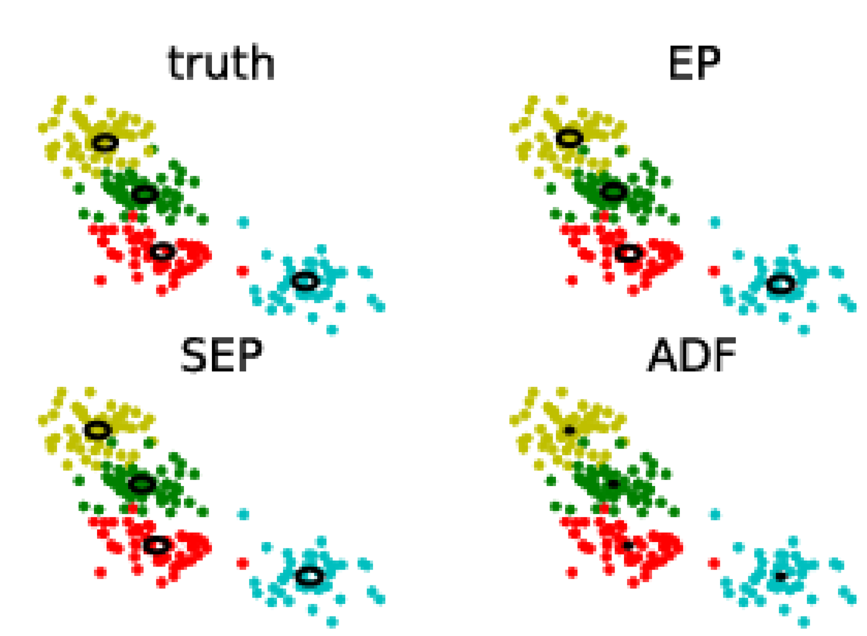


Figure 2: GMM Results

SEP for non-linear LVM

- Can we use SEP for approximate inference in LVMs?
- Gives objective function:
  $\int q(\theta, z_n)r(x_n,\theta,z_n)\log r(x_n,\theta,z_n)dz_n d\theta$
- Where:

$$r(x_n,\theta,z_n) = p(x_n,z_n|\theta)/(t(\theta)f_n(z_n))$$
$$q(\theta,z_n) = q(\theta)q(z_n)$$
$$q(z_{1:N},\theta) = p(\theta)\prod_n f_n(z_n)t(\theta)$$

- ψ: parameters of **recognition model**

$$\theta^{(n)}, z_n^{(n)} \sim q(\theta)f_\psi(z_n;x_n)$$

$$KL \approx \frac{1}{M}\sum_M r(x_n,\theta,z_n)log(r(x_n,\theta,z_n))$$

References

- Y. Li, J. Hernandez-Lobato and R. Turner, "Stochastic Expectation Propagation", NIPS, 2015.
- T. Minka, "Expectation Propagation for Approximate Bayesian Inference", 2001.
- D. P. Kingma, M. Welling, "Auto-Encoding Variational Bayes", ICLR, 2014.