Rebuilding Factorized Information Criterion: Asymptotically Accurate Marginal Likelihood

Kohei Hayashi[1,2] Shin-ichi Maeda[3]
Ryohei Fujimaki[4]

[1]National Institute of Informatics

[2]Kawarabayashi Large Graph Project, ERATO, JST

[3]Kyoto University

[4]NEC Knowledge Discovery Laboratories

September 18, 2015

Introduction

Factorized asymptotic Bayesian inference (FAB)

- Recently-developed approximate Bayesian method

✔ Accurate and tractable

✖ Limited to binary latent variable models (LVMs)

Introduction

Factorized asymptotic Bayesian inference (FAB)

- Recently-developed approximate Bayesian method

✔ Accurate and tractable

✖ Limited to binary latent variable models (LVMs)

Our contributions:

- Extend FAB to general LVMs (e.g. PCA)
- Analyze theoretical properties that are unclear in the previous studies

1. Revisiting FAB
2. Generalization of FAB

Bayesian Inference for Binary LVMs

Binary LVM:

$$p(\underbrace{\mathbf{X}}_{\text{data}}, \underbrace{\mathbf{Z}}_{\text{LVs}}, \underbrace{\mathbf{\Pi}}_{\text{params}} \mid \underbrace{K}_{\text{model}}) = \underbrace{p(\mathbf{\Pi})}_{\text{prior}} \underbrace{p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K)}_{\text{joint likelihood}}$$

Bayesian Inference for Binary LVMs

Binary LVM:

$$p(\underbrace{\mathbf{X}}_{\text{data}}, \underbrace{\mathbf{Z}}_{\text{LVs}}, \underbrace{\mathbf{\Pi}}_{\text{params}} \mid \underbrace{K}_{\text{model}}) = \underbrace{p(\mathbf{\Pi})}_{\text{prior}} \underbrace{p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K)}_{\text{joint likelihood}}$$

Assumptions:

- $\mathbf{X}$ and $\mathbf{Z}$ are jointly i.i.d.

$$p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) = \prod_{n=1}^{N} p(\mathbf{x}_n, \mathbf{z}_n | \mathbf{\Pi}, K)$$

- The prior doesn't depend on N
  - $\ln p(\mathbf{\Pi}) = O(1)$
  - "Flat" prior

Goal: To obtain

- the marginal likelihood:

$$p(\mathbf{X}|K) = \int p(\mathbf{X}, \mathbf{Z}, \mathbf{\Pi}|K)\mathrm{d}\mathbf{Z}\mathrm{d}\mathbf{\Pi}$$

Goal: To obtain

- the marginal likelihood:

$$p(\mathbf{X}|K) = \int p(\mathbf{X}, \mathbf{Z}, \mathbf{\Pi}|K)\mathrm{d}\mathbf{Z}\mathrm{d}\mathbf{\Pi}$$

- the marginal posteriors:

$$p(\mathbf{Z}|\mathbf{X}, K) = \int p(\mathbf{X}, \mathbf{Z}, \mathbf{\Pi}|K)\mathrm{d}\mathbf{\Pi}/p(\mathbf{X}|K)$$
$$p(\mathbf{\Pi}|\mathbf{X}, K) = \int p(\mathbf{X}, \mathbf{Z}, \mathbf{\Pi}|K)\mathrm{d}\mathbf{Z}/p(\mathbf{X}|K)$$

Goal: To obtain

- the marginal likelihood:

$$p(\mathbf{X}|K) = \int p(\mathbf{X}, \mathbf{Z}, \mathbf{\Pi}|K)\mathrm{d}\mathbf{Z}\mathrm{d}\mathbf{\Pi}$$

- the marginal posteriors:

$$p(\mathbf{Z}|\mathbf{X}, K) = \int p(\mathbf{X}, \mathbf{Z}, \mathbf{\Pi}|K)\mathrm{d}\mathbf{\Pi}/p(\mathbf{X}|K)$$

$$p(\mathbf{\Pi}|\mathbf{X}, K) = \int p(\mathbf{X}, \mathbf{Z}, \mathbf{\Pi}|K)\mathrm{d}\mathbf{Z}/p(\mathbf{X}|K)$$

Problem: The marginalizations are intractable

Key idea: Use

- the variational representation for $\int \mathrm{d}\mathbf{Z}$
- Laplace's method for $\int \mathrm{d}\boldsymbol{\Pi}$

Key idea: Use

- the variational representation for $\int \mathrm{d}\mathbf{Z}$
- Laplace's method for $\int \mathrm{d}\mathbf{\Pi}$

Factorized information criterion (FIC)

$$\mathrm{FIC}(K) \equiv \max_q \mathbb{E}_q \left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) \right] \underbrace{- \mathbb{E}_q \left[\frac{D_{\mathbf{\Pi}}}{2} \sum_k \ln \sum_n z_{nk} \right]}_{\text{FIC penalty term}} + H(q)_{+O(\ln N)}$$

- $q(\mathbf{Z})$: trial distribution
- $H(q)$: entropy

Accuracy of FIC

✔ Asymptotically equivalent to the marginal likelihood

Accuracy of FIC

✔ Asymptotically equivalent to the marginal likelihood

Theorem 3 of [Fujimaki+ 12a]

In mixture models, under mild conditions,

$$\begin{aligned} \mathrm{FIC}(K) &= \ln p(\mathbf{X}|K) + O(1) \\ &\approx \ln p(\mathbf{X}|K) \end{aligned}$$

Accuracy of FIC

✔ Asymptotically equivalent to the marginal likelihood

Theorem 3 of [Fujimaki+ 12a]

In mixture models, under mild conditions,

$$\begin{aligned} \mathrm{FIC}(K) &= \ln p(\mathbf{X}|K) + O(1) \\ &\approx \ln p(\mathbf{X}|K) \end{aligned}$$

Similar results are obtained for:

- HMMs [Fujimaki+ 12b]
- Latent feature models [KH+ 13]
- Mixture of experts [Eto+ 14]
- Factorial relational models [Liu+ yesterday]

Optimizing FIC

Computation of FIC is difficult

$$\max_q \mathbb{E}_q \left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) \right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \mathbb{E}_q \left[\ln \sum_n z_{nk} \right] + H(q)$$

Optimizing FIC

Computation of FIC is difficult

$$\max_{q} \mathbb{E}_q \left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) \right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \mathbb{E}_q \left[\ln \sum_n z_{nk} \right] + H(q)$$

$$\geq \max_{q \in \mathcal{Q}} \mathbb{E}_q \left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) \right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \mathbb{E}_q \left[\ln \sum_n z_{nk} \right] + H(q)$$

Mean-field approx. ($\mathcal{Q} \equiv \{q(\mathbf{Z}) | q(\mathbf{Z}) = \prod_n q(\mathbf{z}_n)\}$)

Optimizing FIC

Computation of FIC is difficult

$$\max_{q} \mathbb{E}_q \left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) \right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \mathbb{E}_q \left[\ln \sum_n z_{nk} \right] + H(q)$$

$$\geq \max_{q \in \mathcal{Q}} \mathbb{E}_q \left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) \right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \mathbb{E}_q \left[\ln \sum_n z_{nk} \right] + H(q)$$

Mean-field approx. ($\mathcal{Q} \equiv \{q(\mathbf{Z}) | q(\mathbf{Z}) = \prod_n q(\mathbf{z}_n)\}$)

$$\geq \max_{q \in \mathcal{Q}, \mathbf{\Pi}} \mathbb{E}_q \left[\ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) \right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \ln \sum_n \mathbb{E}_q[z_{nk}] + H(q)$$

Jensen's ineq.

$$\equiv \underline{\mathrm{FIC}}(K)$$

Algorithm

Optimization problem:

$$\max_{q \in \mathcal{Q}, \mathbf{\Pi}} \mathbb{E}_q \left[\ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K)\right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \ln \sum_n \mathbb{E}_q[z_{nk}] + H(q)$$

Algorithm

Optimization problem:

$$\max_{q \in \mathcal{Q}, \mathbf{\Pi}} \mathbb{E}_q \left[\ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K) \right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \ln \sum_n \mathbb{E}_q [z_{nk}] + H(q)$$

Can be solved by EM-like alternating updates:

1. Initialize q and $\mathbf{\Pi}$
2. Update q (Fix $\mathbf{\Pi}$)
3. Update $\mathbf{\Pi}$ (Fix q)
4. Repeat step 2 and 3 until convergence

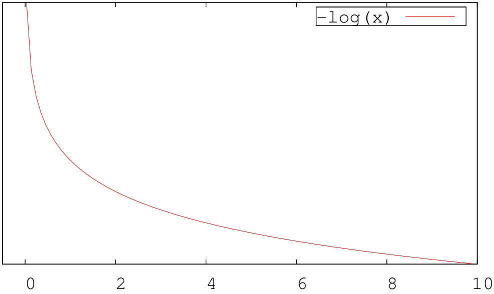
Model Pruning

The FAB algorithm eliminates irrelevant components automatically

Model Pruning

The FAB algorithm eliminates irrelevant components automatically

$$\mathbb{E}_q\left[\ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K)\right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \underbrace{\ln \sum_n \mathbb{E}_q[z_{nk}]}_{\text{penalty term}} + H(q)$$

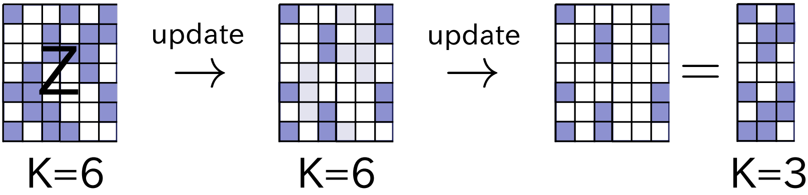


Model Pruning

The FAB algorithm eliminates irrelevant components automatically

$$\mathbb{E}_q\left[\ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)\right] - \frac{D_{\mathbf{\Pi}}}{2} \sum_k \underbrace{\ln \sum_n \mathbb{E}_q[z_{nk}]}_{\text{penalty term}} + H(q)$$

- The penalty term introduces group sparsity to $\mathbf{Z}$



Summary of FIC/FAB

- ✔ Asymptotically equivalent to the marginal likelihood
  - Fits to "Big Data" situations

Summary of FIC/FAB

- ✔ Asymptotically equivalent to the marginal likelihood
  - Fits to "Big Data" situations
- ✔ Performs parameter inference and model selection simultaneously
  - EM-like updates of q and $\mathbf{\Pi}$
  - ARD-like model pruning

Summary of FIC/FAB

- ✔ Asymptotically equivalent to the marginal likelihood
  - Fits to "Big Data" situations
- ✔ Performs parameter inference and model selection simultaneously
  - EM-like updates of q and $\mathbf{\Pi}$
  - ARD-like model pruning
- ✔ Doesn't depend on the choice of $p(\mathbf{\Pi})$
  - More frequentist than Bayesian

Summary of FIC/FAB

- ✔ Asymptotically equivalent to the marginal likelihood
  - Fits to "Big Data" situations
- ✔ Performs parameter inference and model selection simultaneously
  - EM-like updates of q and $\mathbf{\Pi}$
  - ARD-like model pruning
- ✔ Doesn't depend on the choice of $p(\mathbf{\Pi})$
  - More frequentist than Bayesian
- ✔ Works in many binary LVMs

Limitations of FIC/FAB

✖ Limited to binary LVMs

- In real $\mathbf{Z}$, $\sum_n z_{nk}$ can be negative
- $-\ln \sum_n z_{nk}$ may diverge

Limitations of FIC/FAB

- ✖ Limited to binary LVMs
  - In real $\mathbf{Z}$, $\sum_n z_{nk}$ can be negative
  - $-\ln \sum_n z_{nk}$ may diverge
- ✖ Missing relations to EM and VB
  - Similar approaches, but which are better?

Limitations of FIC/FAB

- ✖ Limited to binary LVMs
  - In real $\mathbf{Z}$, $\sum_n z_{nk}$ can be negative
  - $-\ln \sum_n z_{nk}$ may diverge
- ✖ Missing relations to EM and VB
  - Similar approaches, but which are better?
- ✖ Unclear legitimacy of optimizing $\underline{\mathrm{FIC}}$
  - e.g. tightness

1. Revisiting FAB
2. Generalization of FAB

Setting

- Now $\mathbf{Z}$ can take general values (e.g. $\mathbf{Z} \in \mathbb{R}^{N \times K}$)

Setting

- Now $\mathbf{Z}$ can take general values (e.g. $\mathbf{Z} \in \mathbb{R}^{N \times K}$)
- Consider separating the parameters: $\mathbf{\Pi} = \{\mathbf{\Theta}, \mathbf{\Xi}\}$
  - $\mathbf{\Theta}$: k-*independent* params
  - $\mathbf{\Xi} = \{\boldsymbol{\xi}_k\}_{k=1}^{K}$: k-*dependent* params (e.g. mixing coefficients)

Generalized FIC (gFIC)

Definition

$$\mathrm{gFIC}(K) \equiv \mathbb{E}_{q^*}\left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K) \underbrace{-\frac{1}{2}\ln|\mathbf{F}_{\mathbf{\Xi}}|}_{\text{penalty}}\right] + H(q)_{+O(\ln N)}$$

- $q^*(\mathbf{Z}) \equiv p(\mathbf{Z}|\mathbf{X}, K)$: marginal posterior
- $\mathbf{F}_{\mathbf{\Xi}}$: Hessian of $-\ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)/N$ (i.e. empirical Fisher information)
  - In PCA, $\mathbf{F}_{\mathbf{\Xi}} = \mathbf{Z}^\top \mathbf{Z}$

Generalized FIC (gFIC)

Definition

$$\mathrm{gFIC}(K) \equiv \mathbb{E}_{q^*}\left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K) \underbrace{-\frac{1}{2}\ln|\mathbf{F}_{\mathbf{\Xi}}|}_{\text{penalty}}\right] + H(q)_{+O(\ln N)}$$

- $q^*(\mathbf{Z}) \equiv p(\mathbf{Z}|\mathbf{X}, K)$: marginal posterior
- $\mathbf{F}_{\mathbf{\Xi}}$: Hessian of $-\ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)/N$ (i.e. empirical Fisher information)
  - In PCA, $\mathbf{F}_{\mathbf{\Xi}} = \mathbf{Z}^\top \mathbf{Z}$

	FIC	gFIC
Applicable class	Binary LVMs	General LVMs
Penalty term	$-\sum_k \ln \sum_n z_{nk}$	$-\ln|\mathbf{F}_{\mathbf{\Xi}}|$
Regularization	Group sparsity	"Low-rank"

Generalized FAB (gFAB)

✔ Use the same technique as FAB

$$\mathbb{E}_{q^*}\left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K)\right] \qquad - \frac{1}{2}\mathbb{E}_{q^*}\left[\ln |\mathbf{F}_{\mathbf{\Xi}}|\right] + H(q^*)$$

Generalized FAB (gFAB)

✔ Use the same technique as FAB

$$
\begin{aligned}
&\mathbb{E}_{q^*}\left[\max_{\boldsymbol{\Pi}} \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\Pi}, K)\right] && - \frac{1}{2}\mathbb{E}_{q^*}\left[\ln|\mathbf{F}_{\boldsymbol{\Xi}}|\right] + H(q^*) \\
\geq &\max_{q \in \mathcal{Q}} \mathbb{E}_q\left[\max_{\boldsymbol{\Pi}} \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\Pi}, K)\right] && - \frac{1}{2}\mathbb{E}_q\left[\ln|\mathbf{F}_{\boldsymbol{\Xi}}|\right] + H(q)
\end{aligned}
$$

Mean-field approx.

Generalized FAB (gFAB)

✔ Use the same technique as FAB

$$\mathbb{E}_{q^*}\left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K)\right] - \frac{1}{2}\mathbb{E}_{q^*}\left[\ln |\mathbf{F}_{\mathbf{\Xi}}|\right] + H(q^*)$$

$$\geq \max_{q \in \mathcal{Q}} \mathbb{E}_q\left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K)\right] - \frac{1}{2}\mathbb{E}_q\left[\ln |\mathbf{F}_{\mathbf{\Xi}}|\right] + H(q)$$

Mean-field approx.

$$\geq \max_{q \in \mathcal{Q}, \mathbf{\Pi}} \mathbb{E}_q\left[\ln p(\mathbf{X}, \mathbf{Z} | \mathbf{\Pi}, K)\right] - \frac{1}{2}\ln \mathbb{E}_q[|\mathbf{F}_{\mathbf{\Xi}}|] + H(q)$$

Jensen's ineq.

$$\equiv \underline{\mathrm{gFIC}}(K)$$

Generalized FAB (gFAB)

✔ Use the same technique as FAB

$$\mathbb{E}_{q^*}\left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)\right] - \frac{1}{2}\mathbb{E}_{q^*}\left[\ln|\mathbf{F}_{\mathbf{\Xi}}|\right] + H(q^*)$$

$$\geq \max_{q \in \mathcal{Q}} \mathbb{E}_q\left[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)\right] - \frac{1}{2}\mathbb{E}_q\left[\ln|\mathbf{F}_{\mathbf{\Xi}}|\right] + H(q)$$

Mean-field approx.

$$\geq \max_{q \in \mathcal{Q}, \mathbf{\Pi}} \mathbb{E}_q\left[\ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)\right] - \frac{1}{2}\ln \mathbb{E}_q[|\mathbf{F}_{\mathbf{\Xi}}|] + H(q)$$

Jensen's ineq.

$$\equiv \underline{\mathrm{gFIC}}(K)$$

- Able to solve by alternating updates of q and $\mathbf{\Pi}$

Comparison with EM and VB

- ✔ gFAB asymp. approx. $\ln p(\mathbf{X}|K)$ for all K, whereas EM and VB don't

Comparison with EM and VB

✔ gFAB asymp. approx. $\ln p(\mathbf{X}|K)$ for all K, whereas EM and VB don't

Theorem 2 & Corollary 5

Let K' be the "true" model of $\mathbf{X}$, then

$$\begin{aligned} \mathrm{gFIC}(K') &\approx \ln p(\mathbf{X}|K) \quad \text{for } K > K' \\ \mathrm{gFIC}(K) &\approx \ln p(\mathbf{X}|K) \quad \text{for } K \leq K' \end{aligned}$$

- K' can be obtained by model pruning

Comparison with EM and VB

✔ gFAB asymp. approx. $\ln p(\mathbf{X}|K)$ for all K, whereas EM and VB don't

Theorem 2 & Corollary 5

Let K' be the "true" model of $\mathbf{X}$, then

$$\begin{aligned} \mathrm{gFIC}(K') &\approx \ln p(\mathbf{X}|K) \quad \text{for } K > K' \\ \mathrm{gFIC}(K) &\approx \ln p(\mathbf{X}|K) \quad \text{for } K \leq K' \end{aligned}$$

- K' can be obtained by model pruning

Proposition 10+

✖ EM $+O(\ln N) \approx \ln p(\mathbf{X}|K)$ only for $K \leq K'$

✖ VB $\approx \ln p(\mathbf{X}|K)$ only for $K \leq K'$

Asymptotic Behavior of $\underline{\mathrm{gFIC}}$

- ✔ $\underline{\mathrm{gFIC}}(K) \approx \mathrm{gFIC}(K)$ in some cases

Asymptotic Behavior of $\underline{\mathrm{gFIC}}$

✔ $\underline{\mathrm{gFIC}}(K) \approx \mathrm{gFIC}(K)$ in some cases

Proposition 6

q^* is asymptotically mutually independent.

✔ Justify mean-field approximation

Asymptotic Behavior of $\underline{\mathrm{gFIC}}$

- ✔ $\underline{\mathrm{gFIC}}(K) \approx \mathrm{gFIC}(K)$ in some cases

Proposition 6

q^* is asymptotically mutually independent.

- ✔ Justify mean-field approximation

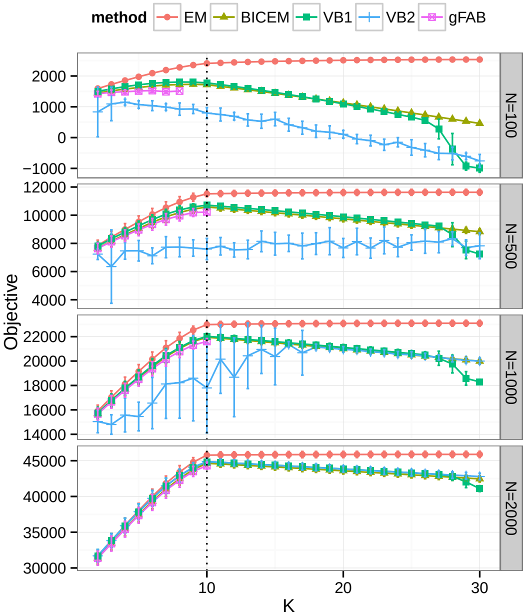
Proposition 7

If q is not degenerated and $\ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)$ is smooth and concave w.r.t. $\mathbf{\Pi}$,

$$\mathbb{E}_q[\max_{\mathbf{\Pi}} \ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)] \xrightarrow{\mathrm{p}} \max_{\mathbf{\Pi}} \mathbb{E}_q[\ln p(\mathbf{X}, \mathbf{Z}|\mathbf{\Pi}, K)].$$

- ✔ Justify Jensen's inequality

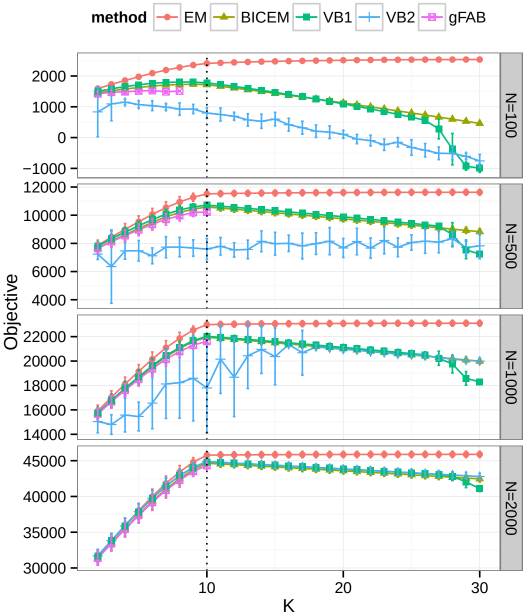
Experiments: Bayesian PCA



Task: model selection

- Choose K that maximizes the objective

Experiments: Bayesian PCA



Task: model selection

- Choose K that maximizes the objective

Results:

- ✔ gFAB: Successfully obtain true $K = 10$ w/ skipping $K = 10, \ldots, 29$
- ✖ EM: Always overestimates K (as suggested in Prop. 10+)
- ✖ VB1: Select true K but need to compute all $K = 1, \ldots, 30$

Conclusion

Summary of this talk:

- FAB: Tractable Bayesian method for binary LVMs
- Proposed gFAB for general LVMs (e.g. PCA)
- Theoretical Analysis
  - Showing the desirable properties of gFAB

Conclusion

Summary of this talk:

- FAB: Tractable Bayesian method for binary LVMs
- Proposed gFAB for general LVMs (e.g. PCA)
- Theoretical Analysis
  - Showing the desirable properties of gFAB

At the poster session (right after):

We will explain more details such as

- Full derivation of gFIC
- "High-level" mechanism of model pruning
- ...

Future work

- Potentially applicable to a wide class of LVMs
  - factor analysis, CCA, partial membership, linear dynamical systems, ...
- If you are interested in, let's collaborate!

Future work

- Potentially applicable to a wide class of LVMs
  - factor analysis, CCA, partial membership, linear dynamical systems, ...
- If you are interested in, let's collaborate!

Thank you!

References

[Fujimaki+ 12a] Fujimaki, Ryohei and Morinaga, Satoshi. Factorized asymptotic Bayesian inference for mixture modeling. In *AISTATS*, 2012.

[Fujimaki+ 12b] Fujimaki, Ryohei and Hayashi, Kohei. Factorized asymptotic Bayesian hidden Markov model. In *ICML*, 2012.

[KH+ 13] Hayashi, Kohei and Fujimaki, Ryohei. Factorized asymptotic Bayesian inference for latent feature models. In *NIPS*, 2013.

[Eto+ 14] Eto, Riki, Fujimaki, Ryohei, Morinaga, Satoshi, and Tamano, Hiroshi. Fully-automatic Bayesian piecewise sparse linear models. In *AISTATS*, 2014.

[Liu+ yesterday] Liu, Chunchen, Feng, Lu, Fujimaki, Ryohei, and Muraoka, Yusuke. Scalable model selection for large-scale factorial relational models. In *ICML*, 2015.