株式会社ニコン
NIKON CORPORATION

変分ベイズ学習の理論

中島伸一

ニコン光技術研究所

（主な）共同研究者：杉山将（東工大）、S.D. Babacan （Google）、
富岡亮太（TTIC）、武田朗子（東大）、竹内一郎（名工大）

よく使われる確率モデルの多くは特異

正則モデル	特異モデル
	行列分解
	ニューラルネット
線形回帰	混合分布
・	隠れマルコフモデル
・	・
・	・
	・

代表的学習法

	正則モデル	特異モデル
最尤（MAP）法		
ベイズ法		

学習理論

正則モデル

特異モデル

最尤（MAP）法

ベイズ法

過学習

	正則モデル	特異モデル
最尤（MAP）法		大
	〜 余分なパラメータ数	
ベイズ法		小

モデル選択規準

	正則モデル	特異モデル
最尤（MAP）法	AIC BIC ・ ・ ・	△
ベイズ法	**周辺尤度（ベイズ自由エネルギー）**	WAIC WBIC

パラメータ推定方法

	正則モデル	特異モデル
最尤（MAP）法	最小化問題 （解析解）	最小化問題 （解析解）
ベイズ法	（共役事前分布を使えば） 解析解	**サンプリング**

変分ベイズ法はベイズ法の近似

	正則モデル	特異モデル
最尤（MAP）法	最小化問題 （解析解）	最小化問題 （解析解）
変分ベイズ法 [Attias99]	ー	**最小化問題** ↑
ベイズ法	（共役事前分布を使えば） 解析解	**サンプリング**

モデル選択規準もOK!

	正則モデル	特異モデル
最尤（MAP）法		△
変分ベイズ法 [Attias99]	AIC BIC ・ ・ ・	**<u>変分ベイズ自由エネルギー</u>**
ベイズ法	**周辺尤度 （ベイズ自由エネルギー）**	WAIC WBIC

変分ベイズ学習

ベイズ事後分布を（広がりのある分布で）近似する一手法

点推定ではない。（↔最尤法、MAP法）

- ✤効率的に計算可能。
- ✤パラメータの推定精度に関する情報が得られる。
- ✤モデル自由度の自動選択が可能（Automatic relevance determination）。

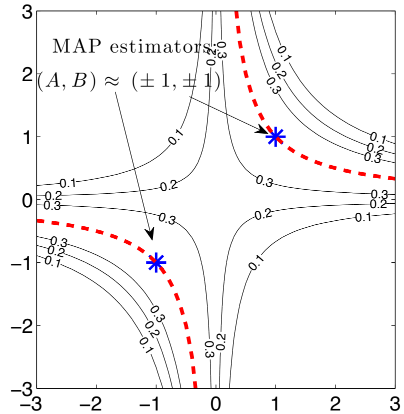
変分ベイズ学習

ベイズ事後分布を（広がりのある分布で）近似する一手法

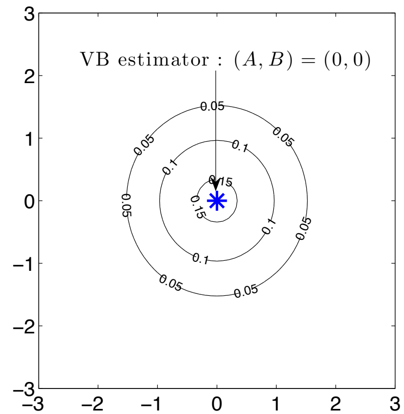
点推定ではない。（↔最尤法、MAP法）

- 効率的に計算可能。
- パラメータの推定精度に関する情報が得られる。
- モデル自由度の自動選択が可能（Automatic relevance determination）。

とはいえ、分布の形は結構違います。。。



ベイズ事後分布

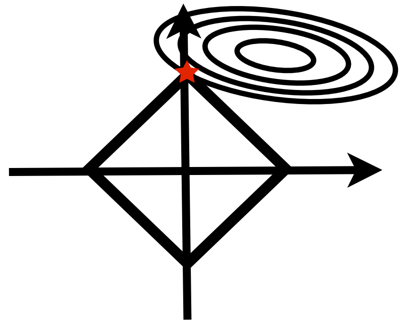


変分ベイズ事後分布

変分ベイズ学習の理論を！

	正則モデル	特異モデル
最尤（MAP）法		
変分ベイズ法 [Attias99]	–	？？？
ベイズ法		

凸形式 vs ベイズ！



凸形式： $\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1$

確率モデル： $p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left(-\frac{\|\boldsymbol{y} - A\boldsymbol{x}\|^2}{2\sigma^2}\right)$

$$p(\boldsymbol{x}) \propto \exp\left(-\frac{\|\boldsymbol{x}\|_1}{2c}\right)$$

$\lambda = \frac{\sigma^2}{c}$ でMAPをやれば凸形式と等価！

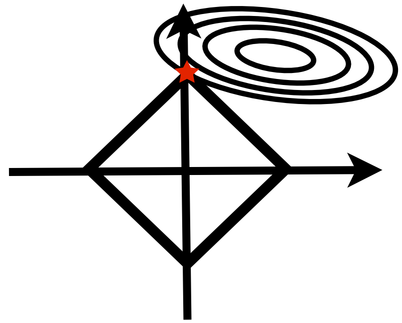
ある条件下で、λを適切に選べばうまくいく。

凸問題

理論保証有

[Candes&Tao, Donoho]

凸形式 vs ベイズ！



凸形式： $\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1$

確率モデル： $p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left(-\frac{\|\boldsymbol{y} - A\boldsymbol{x}\|^2}{2\sigma^2}\right)$

$$p(\boldsymbol{x}) \propto \exp\left(-\frac{\|\boldsymbol{x}\|_1}{2c}\right)$$

$\lambda = \frac{\sigma^2}{c}$ でMAPをやれば凸形式と等価！

凸問題
理論保証有
[Candes&Tao, Donoho]

ある条件下で、λを適切に選べばうまくいく。

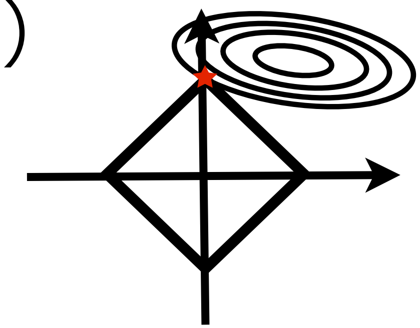
スパースベイズ法： $p(\boldsymbol{x}) \propto \exp\left(-\frac{x_m^2}{2c_m^2}\right)$

x と **c** と σ^2 を推定。

非凸問題
理論保証無

ある条件下でうまくいく（ことが実験で確認されている）。

凸形式 vs ベイズ！（低ランク行列推定）



凸形式： $\min_{U} \|V - U\|^2 + \lambda \|U\|_{\mathrm{tr}}$

確率モデル： $p(V|U) \propto \exp\left(-\frac{\|V - U\|^2}{2\sigma^2}\right)$

$$p(U) \propto \exp\left(-\frac{\|U\|_{\mathrm{tr}}}{2c}\right)$$

$\lambda = \frac{\sigma^2}{c}$ でMAPをやれば凸形式と等価！

凸問題（解析解）
理論保証有
[Candes&Recht2008]

ある条件下で、λを適切に選べばうまくいく。

スパースベイズ法：$p(V|A, B) \propto \exp\left(-\frac{\|V - BA^\top\|^2_{\mathrm{Fro}}}{2\sigma^2}\right)$

$$p(A) \propto \exp\left(-\frac{\mathrm{tr}(AC_A^{-1}A^\top) + \mathrm{tr}(BC_B^{-1}B^\top)}{2}\right)$$

~~非凸問題~~
~~理論保証無~~

解析解
理論保証有

ある条件下でうまくいく！！！

本日お話しすること

- ベイズモデル選択
- 変分ベイズ学習
- 行列分解モデル
  - 大域解析解
  - ランク推定性能の理論保証
- より一般のモデルへ

本日お話しすること

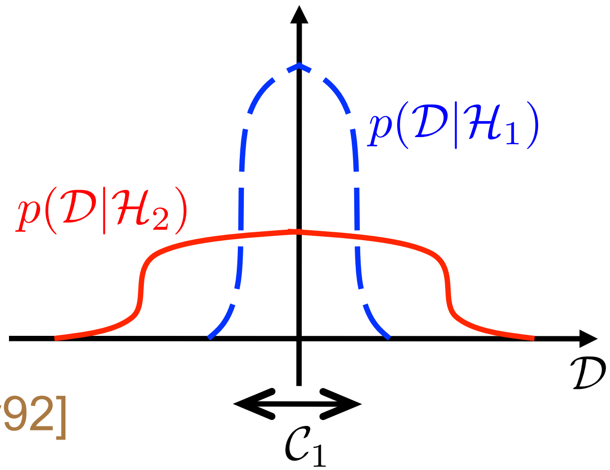
- ベイズモデル選択
- 変分ベイズ学習
- 行列分解モデル
  - 大域解析解
  - ランク推定性能の理論保証
- より一般のモデルへ

周辺尤度によるモデル選択

周辺尤度： $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$

周辺尤度によるモデル選択

周辺尤度： $p(\underbrace{\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}}_{\mathcal{D}\text{: データ}}|\underbrace{\{p(\boldsymbol{x}|\boldsymbol{w}), p(\boldsymbol{w})\}}_{\mathcal{H}\text{: 仮説}}) = \int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$



[Mackay92]

$\mathcal{H}_1$: 自由度の小さい仮説

$\mathcal{H}_2$: 自由度の大きい仮説

$\mathcal{C}_1$: 単純なモデルから発生するデータ

柔軟な仮説は薄く広くカバー

観測データが自由度の小さい仮説で表現できるなら、小さい仮説の尤度が高い。

➡ オッカムのカミソリ効果を持つ（過学習抑制）。

例：経験ベイズ法（type II 最尤推定法）[Efron&Moris73]

モデル分布： $p(y|u) \propto \exp\left(-\frac{(y-u)^2}{2\sigma^2}\right)$

事前分布： $p(u|c_u) \propto \frac{1}{c_u}\exp\left(-\frac{u^2}{2{c_u}^2}\right)$

超事前分布： $p(c_u) \propto 1$

$y \in \mathbb{R}$: observation

$u \in \mathbb{R}$: parameter

$\boxed{c_u \in \mathbb{R}_{++} : \text{hyper-param}}$

正の実数上の各点が仮説！

例：経験ベイズ法（type II 最尤推定法）[Efron&Moris73]

モデル分布： $p(y|u) \propto \exp\left(-\frac{(y-u)^2}{2\sigma^2}\right)$

事前分布： $p(u|c_u) \propto \frac{1}{c_u} \exp\left(-\frac{u^2}{2{c_u}^2}\right)$

超事前分布： $p(c_u) \propto 1$

周辺尤度： $p(y|c_u) = \int p(y|u, c_u)p(u|c_u)du$

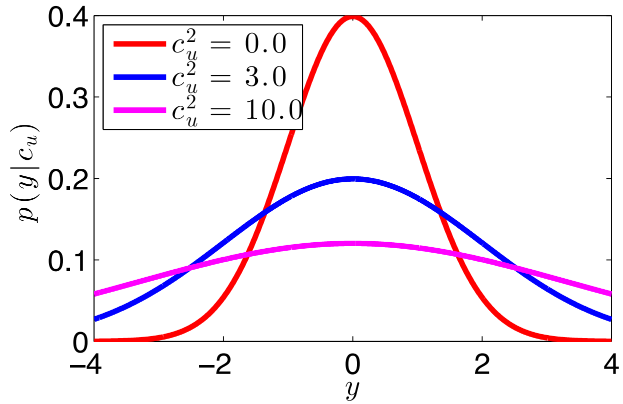
$$\propto \sqrt{\frac{1}{{c_u}^2 + \sigma^2}} \exp\left(-\frac{y^2}{2({c_u}^2 + \sigma^2)}\right)$$

$y \in \mathbb{R}$: observation

$u \in \mathbb{R}$: parameter

$\boxed{c_u \in \mathbb{R}_{++} : \text{hyper-param}}$

正の実数上の各点が仮説！



c_u 大 ➡ 表現力大

例：経験ベイズ法（type II 最尤推定法） [Efron&Moris73]

モデル分布： $p(y|u) \propto \exp\left(-\frac{(y-u)^2}{2\sigma^2}\right)$

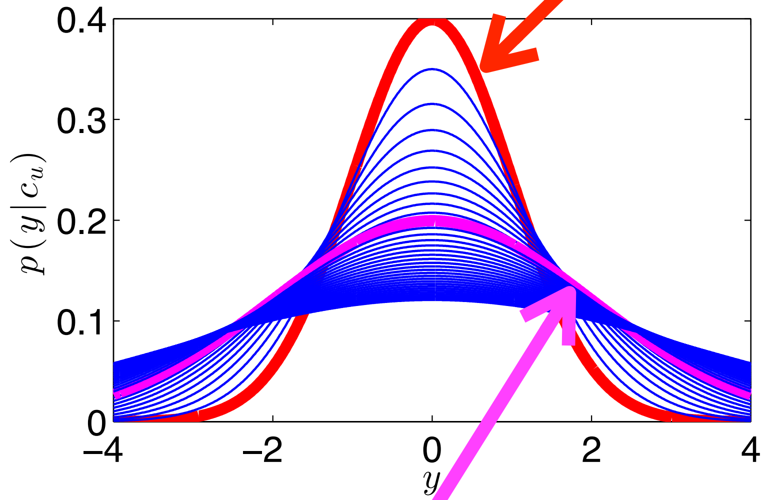
事前分布： $p(u|c_u) \propto \frac{1}{c_u}\exp\left(-\frac{u^2}{2{c_u}^2}\right)$

超事前分布： $p(c_u) \propto 1$

周辺尤度： $p(y|c_u) = \int p(y|u,c_u)p(u|c_u)du$

$$\propto \sqrt{\frac{1}{{c_u}^2+\sigma^2}}\exp\left(-\frac{y^2}{2({c_u}^2+\sigma^2)}\right)$$

$|y| \leq \sigma^2 (=1)$ のとき一番



$|y| = \sqrt{c_u^2+\sigma^2} (=2)$ のとき一番

$$\therefore \widehat{c}_u{}^2 = \underset{c_u}{\operatorname{argmax}}\, p(y|c_u)$$

$$= \begin{cases} 0 & (|y| \leq \sigma^2) \\ y^2 - \sigma^2 & (|y| > \sigma^2) \end{cases}$$

$$\widehat{u} = \langle u \rangle_{p(u|y,\widehat{c}_u{}^2)} = \begin{cases} 0 & (|y| \leq \sigma^2) \\ (1-\sigma^2/y^2)y & (|y| > \sigma^2) \end{cases}$$

positive-part James-Stein推定量 [James&Stein61]

多次元線形モデル

モデル分布： $p(\boldsymbol{y}|\boldsymbol{u}) \propto \exp\left(-\frac{\|\boldsymbol{y}-A\boldsymbol{u}\|^2}{2\sigma^2}\right)$

事前分布： $p(\boldsymbol{u}|\boldsymbol{c}) \propto \frac{1}{c_m} \exp\left(-\frac{{u_m}^2}{2{c_m}^2}\right)$

超事前分布： $p(c_m) \propto 1$

多くの m について、 $\hat{c}_m{}^2 \to 0, \hat{u}_m = 0$ となる。

多次元でやるとスパース解が得られる（スパースベイズ推定）。

MAP推定でもできるか？

モデル分布： $p(y|u) \propto \exp\left(-\frac{(y-u)^2}{2\sigma^2}\right)$

事前分布： $p(u|c_u) \propto \frac{1}{c_u} \exp\left(-\frac{u^2}{2{c_u}^2}\right)$

超事前分布： $p(c_u) \propto 1$

周辺尤度：
$$p(y|c_u) = \int p(y|u, c_u)p(u|c_u)du$$
$$\propto \sqrt{\frac{1}{{c_u}^2+\sigma^2}} \exp\left(-\frac{y^2}{2({c_u}^2+\sigma^2)}\right)$$

同時分布： $p(u, c_u|y) \propto p(y|u, c_u)p(u|c_u)p(c_u)$

同時分布を u, c_u について最大化

MAP推定でもできるか？

モデル分布： $p(y|u) \propto \exp\left(-\frac{(y-u)^2}{2\sigma^2}\right)$

事前分布： $p(u|c_u) \propto \frac{1}{c_u} \exp\left(-\frac{u^2}{2{c_u}^2}\right)$

超事前分布： $p(c_u) \propto 1$

周辺尤度： $p(y|c_u) = \int p(y|u, c_u)p(u|c_u)du$

$$\propto \sqrt{\underbrace{\frac{1}{{c_u}^2+\sigma^2}}} \exp\left(-\frac{y^2}{2({c_u}^2+\sigma^2)}\right)$$

同時分布： $p(u, c_u|y) \propto p(y|u, c_u)p(u|c_u)p(c_u)$

$$\propto \underbrace{\frac{1}{c_u}} \exp\left(-\frac{(y-u)^2}{2\sigma^2} - \frac{u^2}{2{c_u}^2}\right)$$

$c_u \to 0, u = 0$

で発散！

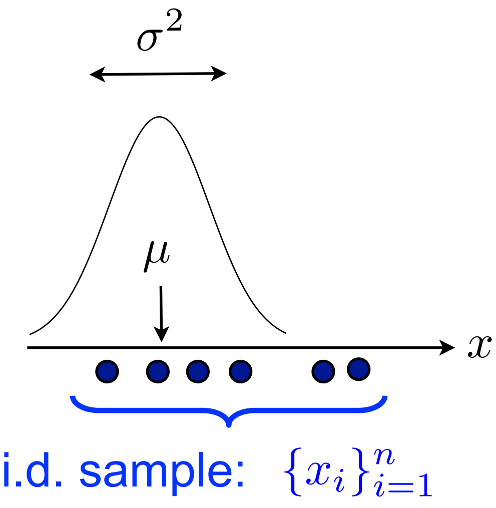
（広がりのある）事後分布を推定したい →

特異モデルでは
変分ベイズ学習！

本日お話しすること

- ベイズモデル選択
- 変分ベイズ学習
- 行列分解モデル
    - 大域解析解
    - ランク推定性能の理論保証
- より一般のモデルへ

ベイズ学習



観測データ： $\{\boldsymbol{x}^{(i)}\}_{i=1}^n$

パラメータ： $\boldsymbol{w} = (\mu, \sigma^2)$

モデル分布： $p(\boldsymbol{x}|\boldsymbol{w}) \propto \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

事前分布： $p(\boldsymbol{w}) \propto 1$

事後分布： $p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \dfrac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$

周辺尤度： $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$

ベイズ推定量： $\widehat{\boldsymbol{w}} = \int \boldsymbol{w} \cdot p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)d\boldsymbol{w} = \langle \boldsymbol{w} \rangle_{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$

予測分布： $p(\boldsymbol{x}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \int p(\boldsymbol{x}|\boldsymbol{w})p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)d\boldsymbol{w}$

（尤度）×（事前分布）の積分が必要

ベイズ学習が解析的にできるモデル

ガウス： $p(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-1/2} \exp\left(-\frac{\mathrm{tr}\left((\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}\right)}{2}\right)$

多項分布（２項分布）： $p(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{k=1}^{K} \frac{{\theta_k}^{x_k}}{x_k!}$

ベイズ学習が解析的にできるモデル

ガウス： $p(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-1/2} \exp\left(-\frac{\mathrm{tr}\left((\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^{\top}\Sigma^{-1}\right)}{2}\right)$

逆ウィシャート（逆ガンマ）： $p(\Sigma|\Psi, \nu) \propto |\Sigma|^{-\frac{\nu+p+1}{2}} \exp\left(-\frac{\mathrm{tr}(\Psi\Sigma^{-1})}{2}\right)$

多項分布（２項分布）： $p(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{k=1}^{K} \frac{{\theta_k}^{x_k}}{x_k!}$

ディリクレ分布（ベータ分布）： $p(\boldsymbol{\theta}|\boldsymbol{\phi}) \propto \prod_{k=1}^{K} {\theta_k}^{\phi_k - 1}$

尤度関数（分布をパラメータの関数として見たもの）が、
名前の付いている（モーメントが知られている）分布と同じ関数形

→ ベイズ学習の計算は、

1. 尤度と共役事前分布をかける。　（事後分布の形を計算）
2. 共役事前分布と同じ形に書きなおす。　（モーメント計算可能！）

ベイズ学習が解析的にできないモデル

行列分解モデル： $p(V|A,B) \propto \exp\left(-\frac{\|V - BA^\top\|_{\mathrm{Fro}}^2}{2\sigma^2}\right)$

指数の上に4次関数

混合ガウス分布: $p(\{\boldsymbol{x}\}_{i=1}^n | \{\alpha_h\}_{h=1}^H, \{\boldsymbol{\mu}_h\}_{h=1}^H) = \prod_{i=1}^n \sum_{h=1}^H \alpha_h \mathcal{N}(\boldsymbol{x}^{(i)}; \boldsymbol{\mu}_h)$

足し算の掛け算

ベイズ学習が解析的にできないモデル

相関を無視したら（独立性制約）

行列分解モデル： $p(V|A,B) \propto \exp\left(-\frac{\|V - BA^\top\|_{\mathrm{Fro}}^2}{2\sigma^2}\right)$

Gaussian on *A*, Gaussian on *B*.

$$\boxed{r(A,B) = r(A)r(B)}$$

混合ガウス分布: $p(\{\boldsymbol{x}\}_{i=1}^n | \{\alpha_h\}_{h=1}^H, \{\boldsymbol{\mu}_h\}_{h=1}^H) = \prod_{i=1}^n \sum_{h=1}^H \alpha_h \mathcal{N}(\boldsymbol{x}^{(i)}; \boldsymbol{\mu}_h)$

$$p(\{\boldsymbol{x}\}_{i=1}^n | \{\alpha_h\}_{h=1}^H, \{\boldsymbol{\mu}_h\}_{h=1}^H, \{\boldsymbol{y}^{(i)}\}_{i=1}^n) = \prod_{i=1}^n \prod_{h=1}^H \left(\alpha_h \mathcal{N}(\boldsymbol{x}^{(i)}; \boldsymbol{\mu}_h)\right)^{\boldsymbol{y}_h^{(i)}}$$

Dirichlet on α_h , Gaussian on μ_h , Polynomial on $\boldsymbol{y}^{(i)}$

$$\boxed{r\left(\{\alpha_h\}_{h=1}^H, \{\mu_h\}_{h=1}^H, \{\boldsymbol{y}^{(i)}\}_{i=1}^n\right) = r\left(\{\alpha_h\}_{h=1}^H, \{\mu_h\}_{h=1}^H\right) r\left(\{\boldsymbol{y}^{(i)}\}_{i=1}^n\right)}$$

自由エネルギー最小化

任意の分布（変数）：　$r(A,B)$

自由エネルギー：$F(r) = \left\langle \log \frac{r(A,B)}{p(V|A,B)p(A)p(B)} \right\rangle_{r(A,B)}$

$$= \underbrace{\mathrm{KL}\left(r(A,B)\|p(A,B|V)\right)}_{\text{事後分布へのカルバック擬距離}} - \underbrace{\log p(V)}_{\text{対数周辺尤度（r に依存しない）}}$$

変分ベイズ法：　$\min_r F(r)$

$\text{s.t.}\ \ r(A,B) = r(A)r(B)$

変分ベイズアルゴリズム

一般に、Iterated conditional modes （局所探索）によって解かれる。

行列分解モデル：

$$\widehat{A} = V^\top \widehat{B} \frac{\Sigma_A}{\sigma^2} \qquad \Sigma_A = \sigma^2 \left(\widehat{B}^\top \widehat{B} + L\Sigma_B + \sigma^2 C_A^{-1} \right)^{-1}$$

$$\widehat{B} = V \widehat{A} \frac{\Sigma_B}{\sigma^2} \qquad \Sigma_B = \sigma^2 \left(\widehat{A}^\top \widehat{A} + M\Sigma_A + \sigma^2 C_B^{-1} \right)^{-1}$$

混合ガウス分布:

$$\widehat{\phi}_h = \overline{N}_h + \phi_0,$$

$$\widehat{\mu}_h = \widehat{\sigma}^2_{\mu_h} \overline{N}_h \overline{x}_h,$$

$$\widehat{\sigma}^2_{\mu_h} = \frac{1}{\overline{N}_h + c^{-2}},$$

$$\overline{N}_h = \sum_{i=1}^{n} \overline{y}_h^{(i)},$$

$$\overline{x}_h = \frac{1}{\overline{N}_h} \sum_{i=1}^{n} \overline{y}_h^{(i)} x^{(i)},$$

$$\overline{y}_h^{(i)} = \frac{\exp\left(\Psi(\widehat{\phi}_h) - \frac{1}{2}\left((x_i - \widehat{\mu}_h)^2 + \widehat{\sigma}^2_{\mu_h} \right) \right)}{\sum_{h=1}^{H} \exp\left(\Psi(\widehat{\phi}_h) - \frac{1}{2}\left((x_i - \widehat{\mu}_h)^2 + \widehat{\sigma}^2_{\mu_h} \right) \right)}$$

経験（empirical）変分ベイズ法

任意の分布（変数）： $r(A,B)$

自由エネルギー： $F(r,C_A,C_B) = \left\langle \log \frac{r(A,B)}{p(V|A,B)p(A|C_A)p(B|C_B)} \right\rangle_{r(A,B)}$

$$= \underbrace{\mathrm{KL}\left(r(A,B)\|p(A,B|V,C_A,C_B)\right)}_{\text{事後分布へのカルバック擬距離}} \underbrace{- \log p(V|C_A,C_B)}_{\text{対数周辺尤度}}$$

経験変分ベイズ法： $\min_{r,C_A,C_B} F(r,C_A,C_B)$

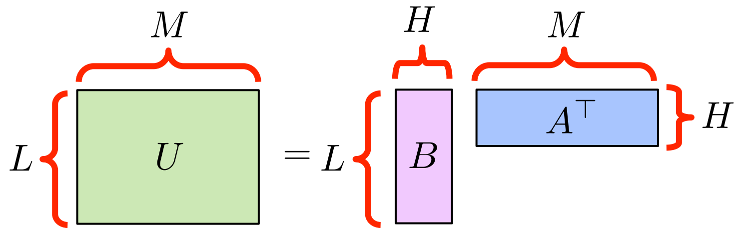
$$\text{s.t. } r(A,B) = r(A)r(B), \quad C_A, C_B \in \mathcal{D}_{++}$$

自由エネルギーをハイパーパラメータについても
最小化すれば、自動モデル選択が実現
（automatic relevance determination）[Neal96]

本日お話しすること

- ベイズモデル選択
- 変分ベイズ学習
- 行列分解モデル
  - 大域解析解
  - ランク推定性能の理論保証
- より一般のモデルへ

行列分解モデル



尤度： $p(V|A,B) \propto \exp\left(-\frac{\|V - BA^\top\|_{\mathrm{Fro}}^2}{2\sigma^2}\right)$

事前分布： $p(A) \propto \exp\left(-\frac{\mathrm{tr}(AC_A^{-1}A^\top)}{2}\right)$

$$p(B) \propto \exp\left(-\frac{\mathrm{tr}(BC_B^{-1}B^\top)}{2}\right)$$

観測値：

$$V \in \mathbb{R}^{L\times M}$$

パラメータ:

$$A \in \mathbb{R}^{M\times H}$$

$$B \in \mathbb{R}^{L\times H}$$

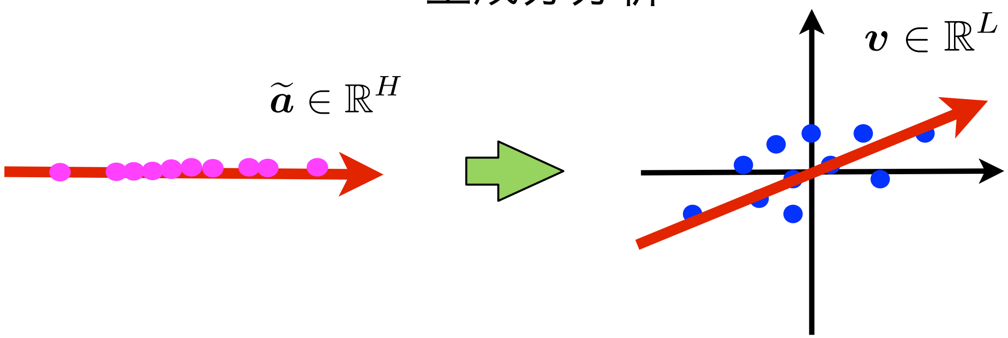
超パラメータ:

$$C_A = \mathrm{diag}(c_{a_1}^2, \ldots, c_{a_H}^2)$$

$$C_B = \mathrm{diag}(c_{b_1}^2, \ldots, c_{b_H}^2)$$

$$\sigma^2 > 0$$

主成分分析



※ V は欠損値を持たないとする。

※ $M \geq L$ とする。$M < L$ のときは $V \to V^\top$ に対して実行する。

変分ベイズ事後分布

※ノイズ分散 σ^2 は当面、given とする。

標準的な手続きにより、以下がわかる。

VB事後分布： $$r(A,B) \propto \exp\left(-\frac{\mathrm{tr}\left((A-\widehat{A})\Sigma_A(A-\widehat{A})^\top\right)}{2}\right)\exp\left(-\frac{\mathrm{tr}\left((B-\widehat{B})\Sigma_B(B-\widehat{B})^\top\right)}{2}\right)$$

where $$(\widehat{A},\widehat{B},\Sigma_A,\Sigma_B,C_A,C_B) = \underset{(\widehat{A},\widehat{B},\Sigma_A,\Sigma_B,C_A,C_B)}{\mathrm{argmin}} F$$

自由エネルギー： $$2F = LM\log(2\pi\sigma^2) - (L+M)H + M\log\frac{|C_A|}{|\Sigma_A|} + L\log\frac{|C_B|}{|\Sigma_B|} + \mathrm{tr}\left(C_A^{-1}(\widehat{A}^\top\widehat{A}+M\Sigma_A)\right) + \mathrm{tr}\left(C_B^{-1}(\widehat{B}^\top\widehat{B}+L\Sigma_B)\right) + \frac{1}{\sigma^2}\left(\|V-\widehat{B}\widehat{A}^\top\|^2 - \|\widehat{B}\widehat{A}^\top\|^2 + \mathrm{tr}\left((\widehat{A}^\top\widehat{A}+M\Sigma_A)(\widehat{B}^\top\widehat{B}+L\Sigma_B)\right)\right)$$

停留条件： $$\widehat{A} = V^\top\widehat{B}\frac{\Sigma_A}{\sigma^2}, \qquad \Sigma_A = \sigma^2\left(\widehat{B}^\top\widehat{B}+L\Sigma_B+\sigma^2C_A^{-1}\right)^{-1},$$
$$\widehat{B} = V\widehat{A}\frac{\Sigma_B}{\sigma^2}, \qquad \Sigma_B = \sigma^2\left(\widehat{A}^\top\widehat{A}+M\Sigma_A+\sigma^2C_B^{-1}\right)^{-1}.$$

$$c_{a_h}^2 = \|\widehat{\boldsymbol{a}}_h\|^2/M + (\Sigma_A)_{hh}, \qquad c_{b_h}^2 = \|\widehat{\boldsymbol{b}}_h\|^2/L + (\Sigma_B)_{hh}.$$

最尤解はSVDで求まる

対数周辺尤度： $2F^{\mathrm{ML}} = LM\log(2\pi\sigma^2) + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2\right)$
（の符号反転）

Vの特異値分解を $V = \sum_{h=1}^{L} \gamma_h \boldsymbol{\omega}_{b_h} \boldsymbol{\omega}_{a_h}^\top$ とすると、ML解は

$$\widehat{U}^{\mathrm{ML}} = (\widehat{B}\widehat{A}^\top)^{\mathrm{ML}} = \sum_{h=1}^{H} \gamma_h \boldsymbol{\omega}_{b_h} \boldsymbol{\omega}_{a_h}^\top,$$

MAP解 （given C_A, C_B ）も既知

対数事後確率：（の符号反転）

$$2F^{\mathrm{MAP}} = LM\log(2\pi\sigma^2) + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2\right) + \mathrm{tr}\left(C_A^{-1}\widehat{A}^\top\widehat{A}\right) + \mathrm{tr}\left(C_B^{-1}\widehat{B}^\top\widehat{B}\right)$$

V の特異値分解を $V = \sum_{h=1}^{L} \gamma_h \boldsymbol{\omega}_{b_h} \boldsymbol{\omega}_{a_h}^\top$ とすると、MAP解は

$$\widehat{U}^{\mathrm{MAP}} = (\widehat{B}\widehat{A}^\top)^{\mathrm{MAP}} = \sum_{h=1}^{H} \widehat{\gamma}_h^{\mathrm{MAP}} \boldsymbol{\omega}_{b_h} \boldsymbol{\omega}_{a_h}^\top,$$

where $$\widehat{\gamma}_h^{\mathrm{MAP}} = \max\left(0, \gamma_h - \frac{\sigma^2}{c_{a_h} c_{b_h}}\right)$$

MAP解はreweighted SVD。

$*\ \min_U \|V - U\|^2 + \lambda\|U\|_{\mathrm{tr}}$ と等価 [Srebro&Jaakkola03] $\because 2\|U\|_{\mathrm{tr}} = \min_{U=BA^\top}(\|A\|_{\mathrm{Fro}}^2 + \|B\|_{\mathrm{Fro}}^2)$

VB解も reweighted SVD?

対数事後確率：
（の符号反転）

$$2F^{\mathrm{MAP}} = LM\log(2\pi\sigma^2) + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2\right) + \mathrm{tr}\left(C_A^{-1}\widehat{A}^\top\widehat{A}\right) + \mathrm{tr}\left(C_B^{-1}\widehat{B}^\top\widehat{B}\right)$$

自由エネルギー：

$$2F = LM\log(2\pi\sigma^2) - (L+M)H + M\log\frac{|C_A|}{|\Sigma_A|} + L\log\frac{|C_B|}{|\Sigma_B|} + \mathrm{tr}\left(C_A^{-1}(\widehat{A}^\top\widehat{A} + M\Sigma_A)\right) + \mathrm{tr}\left(C_B^{-1}(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right) + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2 - \|\widehat{B}\widehat{A}^\top\|^2 + \mathrm{tr}\left((\widehat{A}^\top\widehat{A} + M\Sigma_A)(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)\right)$$

VB解も reweighted SVD?

対数事後確率：
（の符号反転）

$$2F^{\mathrm{MAP}} = LM\log(2\pi\sigma^2) + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2\right) + \mathrm{tr}\left(C_A^{-1}\widehat{A}^\top\widehat{A}\right) + \mathrm{tr}\left(C_B^{-1}\widehat{B}^\top\widehat{B}\right)$$

自由エネルギー：

$$2F = LM\log(2\pi\sigma^2) - (L+M)H + M\log\frac{|C_A|}{|\Sigma_A|} + L\log\frac{|C_B|}{|\Sigma_B|} + \mathrm{tr}\left(C_A^{-1}(\widehat{A}^\top\widehat{A} + M\Sigma_A)\right) + \mathrm{tr}\left(C_B^{-1}(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right) + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2 - \|\widehat{B}\widehat{A}^\top\|^2 + \mathrm{tr}\left((\widehat{A}^\top\widehat{A} + M\Sigma_A)(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)\right)$$

Σ_A, Σ_B を対角に制限すれば reweighted SVD [Nakajima&Sugiyama:JMLR2011]。

対角に制限する解としない解の挙動が殆ど同じ[Nakajima+:ICML2011]。

分布の広がりが解の方向を変える気が（あまり）しない。

Σ_A, Σ_Bが対角であることを証明する

自由エネルギー：
$$2F = LM\log(2\pi\sigma^2) - (L+M)H + M\log\frac{|C_A|}{|\Sigma_A|} + L\log\frac{|C_B|}{|\Sigma_B|} + \mathrm{tr}\left(C_A^{-1}(\widehat{A}^\top\widehat{A} + M\Sigma_A)\right) + \mathrm{tr}\left(C_B^{-1}(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right) + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2 - \|\widehat{B}\widehat{A}^\top\|^2 + \mathrm{tr}\left((\widehat{A}^\top\widehat{A} + M\Sigma_A)(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)\right)$$

任意の直交変換

$$\widehat{A} = A^*\Omega^\top, \qquad \Sigma_A = \Omega\Sigma_A^*\Omega^\top,$$
$$\widehat{B} = B^*\Omega^\top, \qquad \Sigma_B = \Omega\Sigma_B^*\Omega^\top.$$

に対して殆どの項が不変。

Σ_A, Σ_Bが対角であることを証明する

自由エネルギー：
$$2F = LM\log(2\pi\sigma^2) - (L+M)H + M\log\frac{|C_A|}{|\Sigma_A|} + L\log\frac{|C_B|}{|\Sigma_B|}$$
$$\boxed{+\operatorname{tr}\left(C_A^{-1}(\widehat{A}^\top\widehat{A} + M\Sigma_A)\right) + \operatorname{tr}\left(C_B^{-1}(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)}$$
$$+\frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2 - \|\widehat{B}\widehat{A}^\top\|^2 + \operatorname{tr}\left((\widehat{A}^\top\widehat{A} + M\Sigma_A)(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)\right)$$

任意の直交変換

$$\widehat{A} = A^*\Omega^\top, \qquad \Sigma_A = \Omega\Sigma_A^*\Omega^\top,$$
$$\widehat{B} = B^*\Omega^\top, \qquad \Sigma_B = \Omega\Sigma_B^*\Omega^\top.$$

に対して殆どの項が不変。

Σ_A, Σ_Bが対角であることを証明する

自由エネルギー： $$2F = LM\log(2\pi\sigma^2) - (L+M)H + M\log\frac{|C_A|}{|\Sigma_A|} + L\log\frac{|C_B|}{|\Sigma_B|} + \mathrm{tr}\left(C_A^{-1}(\widehat{A}^\top\widehat{A} + M\Sigma_A)\right) + \boxed{\mathrm{tr}\left(C_B^{-1}(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)} + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2 - \|\widehat{B}\widehat{A}^\top\|^2 + \mathrm{tr}\left((\widehat{A}^\top\widehat{A} + M\Sigma_A)(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)\right)$$

任意の直交行列 $\Omega^\top$対して

$$\widehat{A} = A^* C_A^{-1/2}\Omega^\top C_A^{1/2}, \qquad \Sigma_A = C_A^{1/2}\Omega C_A^{-1/2}\Sigma_A^* C_A^{-1/2}\Omega^\top C_A^{1/2},$$
$$\widehat{B} = B^* C_A^{1/2}\Omega^\top C_A^{-1/2}, \qquad \Sigma_B = C_A^{-1/2}\Omega C_A^{1/2}\Sigma_B^* C_A^{1/2}\Omega^\top C_A^{-1/2}.$$

1項を除いて不変。

Σ_A, Σ_Bが対角であることを証明する

$(A^*, \Sigma_A^*, B^*, \Sigma_B^*)$ を解と仮定する。

自由エネルギー：
$$2F = LM\log(2\pi\sigma^2) - (L+M)H + M\log\frac{|C_A|}{|\Sigma_A|} + L\log\frac{|C_B|}{|\Sigma_B|}$$
$$+\,\mathrm{tr}\left(C_A^{-1}(\widehat{A}^\top\widehat{A} + M\Sigma_A)\right) + \boxed{\mathrm{tr}\left(C_B^{-1}(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)}$$
$$+\,\frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2 - \|\widehat{B}\widehat{A}^\top\|^2 + \mathrm{tr}\left((\widehat{A}^\top\widehat{A} + M\Sigma_A)(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)\right)$$

任意の直交行列 $\Omega^\top$対して

$$\widehat{A} = A^* C_A^{-1/2}\Omega^\top C_A^{1/2}, \qquad \Sigma_A = C_A^{1/2}\Omega C_A^{-1/2}\Sigma_A^* C_A^{-1/2}\Omega^\top C_A^{1/2},$$
$$\widehat{B} = B^* C_A^{1/2}\Omega^\top C_A^{-1/2}, \qquad \Sigma_B = C_A^{-1/2}\Omega C_A^{1/2}\Sigma_B^* C_A^{1/2}\Omega^\top C_A^{-1/2}.$$

1項を除いて不変。

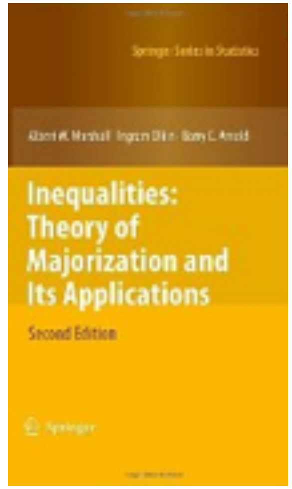
自由エネルギーは $\Omega = I_H$ のとき最小。

Σ_A, Σ_Bが対角であることを証明する

$(A^*, \Sigma_A^*, B^*, \Sigma_B^*)$ を解と仮定する。 以下は $\Omega = I_H$ のとき最小。

$$2F(\Omega) = \boxed{\mathrm{tr}\left(C_A^{-1} C_B^{-1} \Omega C_A^{1/2} \left(B^{*\top} B^* + L\Sigma_B^*\right) C_A^{1/2} \Omega^\top\right)} + \mathrm{const.}$$

Σ_A, Σ_Bが対角であることを証明する



$(A^*, \Sigma_A^*, B^*, \Sigma_B^*)$ を解と仮定する。 以下は $\Omega = I_H$ のとき最小。

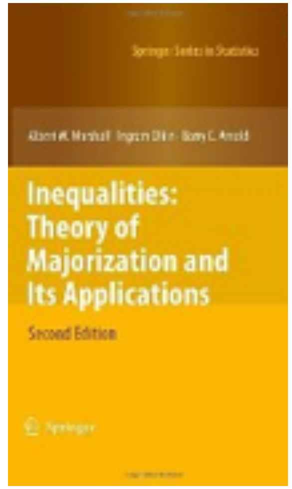
$$2F(\Omega) = \mathrm{tr}\left(\underbrace{C_A^{-1} C_B^{-1}}_{\text{対角}} \Omega \underbrace{C_A^{1/2}\left(B^{*\top} B^* + L\Sigma_B^*\right) C_A^{1/2}}_{？} \Omega^\top\right) + \mathrm{const.}$$

2つの行列をアライメントして、トレースを最小化する問題。

Proposition 1 *(Ruhe, 1970) Let $\lambda_h(\Phi), \lambda_h(\Psi)$ be the h-th largest eigenvalues of positive-definite symmetric matrices $\Phi, \Psi \in \mathbb{R}^{H \times H}$, respectively. Then, it holds that*

$$tr\{\Phi^{-1}\Psi\} \geq \sum_{h=1}^{H} \frac{\lambda_h(\Psi)}{\lambda_h(\Phi)}.$$

Σ_A, Σ_Bが対角であることを証明する



$(A^*, \Sigma_A^*, B^*, \Sigma_B^*)$ を解と仮定する。 以下は $\Omega = I_H$ のとき最小。

$$2F(\Omega) = \mathrm{tr}\left(\underbrace{C_A^{-1}C_B^{-1}}_{\text{対角}}\,\Omega \underbrace{C_A^{1/2}\left(B^{*\top}B^* + L\Sigma_B^*\right)C_A^{1/2}}_{\text{対角！}}\,\Omega^\top\right) + \mathrm{const.}$$

2つの行列をアライメントして、トレースを最小化する問題。

Lemma 1 *(Nakajima et al., 2013) Let $\Gamma, \Omega, \Phi \in \mathbb{R}^{H \times H}$ be a non-degenerate diagonal matrix, an orthogonal matrix, and a symmetric matrix, respectively. If*

$$G(\Omega) = tr\left\{\Gamma\Omega\Phi\Omega^\top\right\}$$

is minimized (as a function of Ω, given Γ, Φ) when $\Omega = I_H$, then Φ is diagonal.

Σ_A, Σ_Bが対角であることを証明する

$(A^*, \Sigma_A^*, B^*, \Sigma_B^*)$ を解と仮定する。 以下は $\Omega = I_H$ のとき最小。

$$2F(\Omega) = \mathrm{tr}\left(C_A^{-1} C_B^{-1} \Omega \underbrace{C_A^{1/2} \left(B^{*\top} B^* + L\Sigma_B^* \right) C_A^{1/2}}_{\text{対角！}} \Omega^\top \right) + \mathrm{const.}$$

停留条件：

$$\widehat{A} = V^\top \widehat{B} \frac{\Sigma_A}{\sigma^2}, \qquad \Sigma_A = \sigma^2 \left(\boxed{\widehat{B}^\top \widehat{B} + L\Sigma_B} + \sigma^2 C_A^{-1} \right)^{-1},$$

$$\widehat{B} = V \widehat{A} \frac{\Sigma_B}{\sigma^2}, \qquad \Sigma_B = \sigma^2 \left(\widehat{A}^\top \widehat{A} + M\Sigma_A + \sigma^2 C_B^{-1} \right)^{-1}.$$

$$c_{a_h}^2 = \|\widehat{\boldsymbol{a}}_h\|^2/M + (\Sigma_A)_{hh}, \qquad c_{b_h}^2 = \|\widehat{\boldsymbol{b}}_h\|^2/L + (\Sigma_B)_{hh}.$$

Σ_A は対角。

Σ_B の対角性も同様に示せる。

自由エネルギーを分解

$$2F(\widehat{A}, \widehat{B}, \Sigma_A, \Sigma_B, C_A, C_B) = LM\log(\pi\sigma^2) + \frac{\|V\|^2}{\sigma^2} + \sum_{h=1}^{H} 2F_h$$, where

$$2F_h(a_h, b_h, \sigma_{a_h}^2, \sigma_{b_h}^2, c_{a_h}^2, c_{b_h}^2) = -M\log\sigma_{a_h}^2 + \frac{a_h^2 + M\sigma_{a_h}^2}{c_{a_h}^2} - L\log\sigma_{b_h}^2 + \frac{b_h^2 + L\sigma_{b_h}^2}{c_{b_h}^2} - (M+L) - \frac{2}{\sigma^2}\gamma_h a_h b_h + \frac{1}{\sigma^2}\left(a_h^2 + M\sigma_{a_h}^2\right)\left(b_h^2 + L\sigma_{b_h}^2\right)$$

未知数の数： $O(MH) \to O(1)$

しかも、停留条件は多項式系（がんばれば解ける！）

$$\widehat{a}_h = \gamma_h \widehat{b}_h \frac{\sigma_{a_h}^2}{\sigma^2}, \qquad \sigma_{a_h}^2 = \sigma^2\left(\widehat{b}_h^2 + L\sigma_{b_h}^2 + \frac{\sigma^2}{c_{a_h}^2}\right)^{-1},$$

$$\widehat{b}_h = \gamma_h \widehat{a}_h \frac{\sigma_{b_h}^2}{\sigma^2}, \qquad \sigma_{b_h}^2 = \sigma^2\left(\widehat{a}_h^2 + M\sigma_{a_h}^2 + \frac{\sigma^2}{c_{b_h}^2}\right)^{-1},$$

$$c_{a_h}^2 = \widehat{a}_h^2/M + \sigma_{a_h}^2, \qquad c_{b_h}^2 = \widehat{b}_h^2/L + \sigma_{b_h}^2.$$

経験変分ベイズ解 (given σ^2)

定理:

EVB solution is given by $\widehat{U}^{\mathrm{EVB}} = \sum_{h=1}^{H} \widehat{\gamma}_h^{\mathrm{EVB}} \boldsymbol{\omega}_{b_h} \boldsymbol{\omega}_{a_h}^\top$, where $\widehat{\gamma}_h^{\mathrm{EVB}} = \begin{cases} 0 & (\gamma_h \le \underline{\gamma}_h^{\mathrm{EVB}}), \\ \breve{\gamma}_h^{\mathrm{EVB}} & (\gamma_h > \underline{\gamma}_h^{\mathrm{EVB}}). \end{cases}$

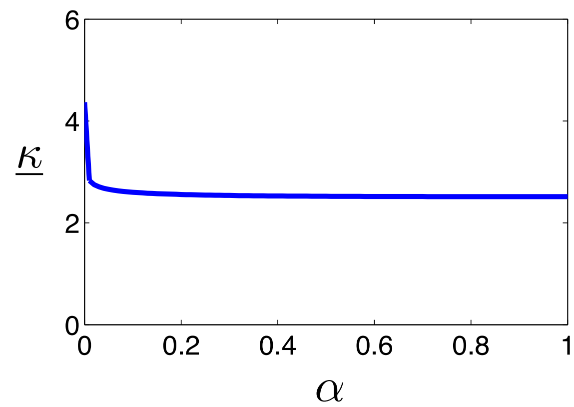
Threshold: $\underline{\gamma}^{\mathrm{EVB}} = \sigma \sqrt{M + L + \sqrt{LM}\left(\underline{\kappa} + \frac{1}{\underline{\kappa}}\right)},$

Amplitude: $\breve{\gamma}_h^{\mathrm{EVB}} = \frac{\gamma_h}{2}\left(1 - \frac{(M+L)\sigma^2}{\gamma_h^2} + \sqrt{\left(1 - \frac{(M+L)\sigma^2}{\gamma_h^2}\right)^2 - \frac{4LM\sigma^4}{\gamma_h^4}}\right).$

Here, $\underline{\kappa}$ is the zero-cross point of

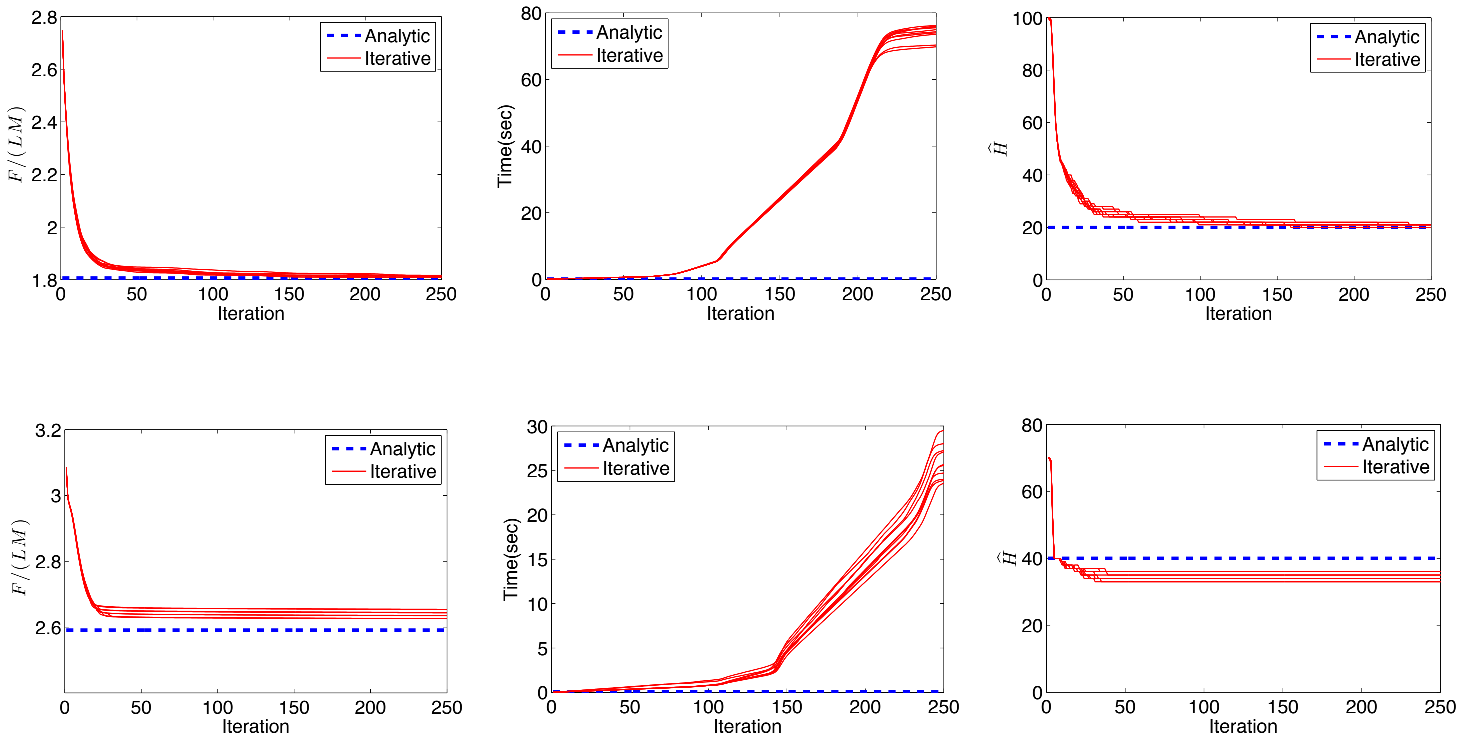
$$\Xi(\kappa; \alpha) = \Phi\left(\sqrt{\alpha}\kappa\right) + \Phi\left(\frac{\kappa}{\sqrt{\alpha}}\right),$$

where $\alpha = \frac{L}{M}, \quad \Phi(x) = \frac{\log(x+1)}{x} - \frac{1}{2}.$



σ^2 は1次元サーチを行う。

局所探索との比較実験



$$L = 100,\quad M = 300,\quad H^* = 20.$$

$$L = 70,\quad M = 300,\quad H^* = 40.$$

Free energy　　Computation time　　Estimated rank

高速に大域解が求まる。

本日お話しすること

- ベイズモデル選択
- 変分ベイズ学習
- 行列分解モデル
  - 大域解析解
  - ランク推定性能の理論保証
- より一般のモデルへ

性能保証を与えたい

尤度： $$p(V|A,B) \propto \exp\left(-\frac{\|V - BA^\top\|^2_{\mathrm{Fro}}}{2\sigma^2}\right)$$

事前分布： $$p(A) \propto \exp\left(-\frac{\mathrm{tr}(AC_A^{-1}A^\top)}{2}\right)$$

$$p(B) \propto \exp\left(-\frac{\mathrm{tr}(BC_B^{-1}B^\top)}{2}\right)$$

σ^2 推定を含めた完全自動学習のときの性能を保証すれば、

V を与えるだけで、

- （解析解によって）手軽に計算できる、
- 理論によって性能が保証された

モデル（ランク、主成分次元）選択付き主成分分析が実現できる。

凸形式： $\min_{\boldsymbol{x}} \|V - U\|^2 + \lambda\|U\|_{\mathrm{tr}}$ より便利！

自由エネルギー： $$2F = LM\log(2\pi\sigma^2) - (L+M)H + M\log\frac{|C_A|}{|\Sigma_A|} + L\log\frac{|C_B|}{|\Sigma_B|} + \mathrm{tr}\left(C_A^{-1}(\widehat{A}^\top\widehat{A} + M\Sigma_A)\right) + \mathrm{tr}\left(C_B^{-1}(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right) + \frac{1}{\sigma^2}\left(\|V - \widehat{B}\widehat{A}^\top\|^2 - \|\widehat{B}\widehat{A}^\top\|^2 + \mathrm{tr}\left((\widehat{A}^\top\widehat{A} + M\Sigma_A)(\widehat{B}^\top\widehat{B} + L\Sigma_B)\right)\right)$$

に σ^2 givenの経験変分ベイズ解を代入すると、

自由エネルギーはきれいに分解される！

Theorem

The noise estimator $\widehat{\sigma}^{2\ \mathrm{EVB}}$ is global minimizer of

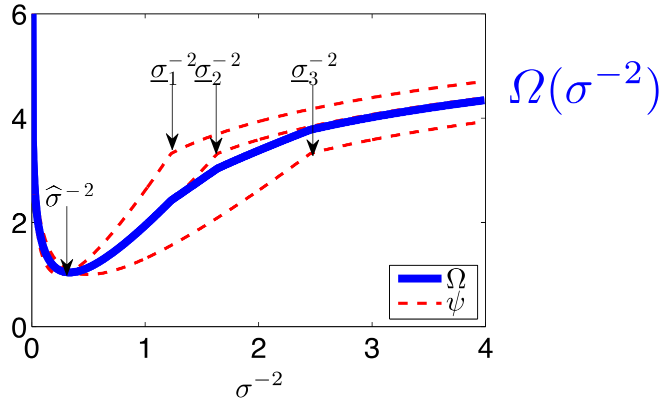
$$\Omega(\sigma^{-2}) = \sum_{h=1}^{H} \psi\left(\frac{\gamma_h^2}{M}\sigma^{-2}\right)$$

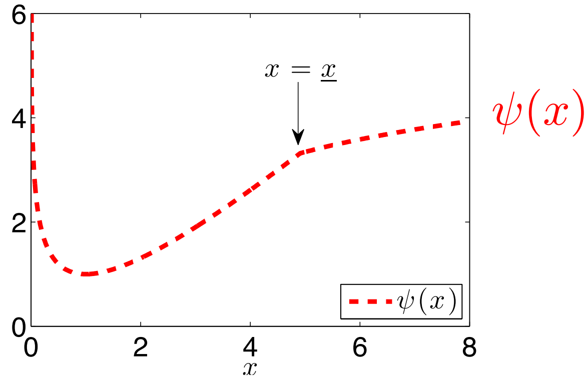
where $\psi(x) = \psi_0(x) + \theta(x > \underline{x})\,\psi_1(x)$, $\quad \underline{x} = 1 + \alpha + \sqrt{\alpha}\left(\underline{\kappa} + \underline{\kappa}^{-1}\right)$,

$$\psi_0(x) = x - \log x, \qquad \psi_1(x) = \log\left(\sqrt{\alpha}\kappa + 1\right) + \alpha \log\left(\frac{\kappa}{\sqrt{\alpha}} + 1\right) - \sqrt{\alpha}\kappa,$$

and κ is a function of x, defined by

$$\kappa(x) = \frac{1}{2\sqrt{\alpha}}\left((x - (1+\alpha)) + \sqrt{(x - (1+\alpha))^2 - 4\alpha}\right).$$





自由エネルギーはきれいに分解される！

Theorem

The noise estimator $\widehat{\sigma}^{2\ \mathrm{EVB}}$ is global minimizer of

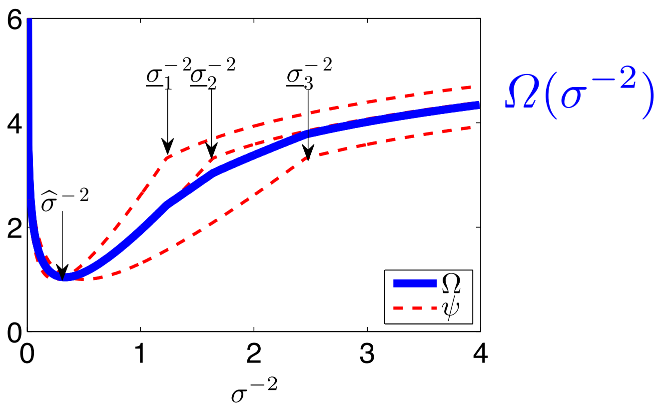
$$\Omega(\sigma^{-2}) = \sum_{h=1}^{H} \psi\left(\underbrace{\boxed{\frac{\gamma_h^2}{M}}}_{\text{scale}} \sigma^{-2}\right)$$

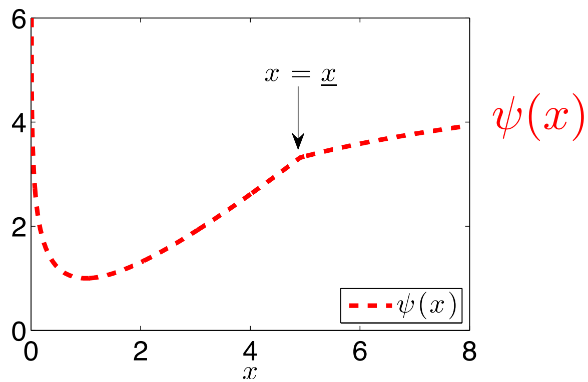
where $\psi(x) = \psi_0(x) + \theta(x > \underline{x})\,\psi_1(x)$, $\quad \underline{x} = 1 + \alpha + \sqrt{\alpha}\left(\underline{\kappa} + \underline{\kappa}^{-1}\right)$,

$$\psi_0(x) = x - \log x, \qquad \psi_1(x) = \log\left(\sqrt{\alpha}\kappa + 1\right) + \alpha \log\left(\frac{\kappa}{\sqrt{\alpha}} + 1\right) - \sqrt{\alpha}\kappa,$$

and κ is a function of x, defined by

$$\kappa(x) = \frac{1}{2\sqrt{\alpha}}\left((x - (1+\alpha)) + \sqrt{(x-(1+\alpha))^2 - 4\alpha}\right).$$





自由エネルギーはきれいに分解される！

Theorem

The noise estimator $\widehat{\sigma}^{2\ \mathrm{EVB}}$ is global minimizer of

$$\Omega(\sigma^{-2}) = \sum_{h=1}^{H} \psi\left(\boxed{\frac{\gamma_h^2}{M}}\sigma^{-2}\right)$$

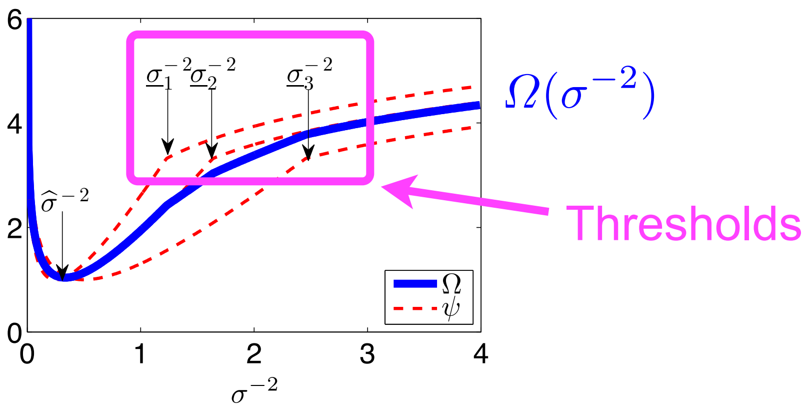
scale

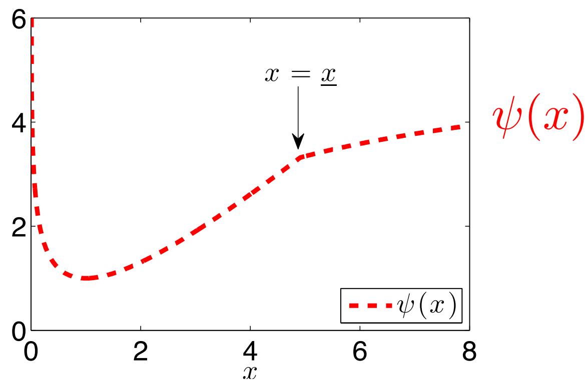
where $\psi(x) = \psi_0(x) + \theta(x > \underline{x})\,\psi_1(x)$, $\underline{x} = 1 + \alpha + \sqrt{\alpha}\left(\underline{\kappa} + \underline{\kappa}^{-1}\right)$,

$$\psi_0(x) = x - \log x, \qquad \psi_1(x) = \log\left(\sqrt{\alpha}\kappa + 1\right) + \alpha \log\left(\frac{\kappa}{\sqrt{\alpha}} + 1\right) - \sqrt{\alpha}\kappa,$$

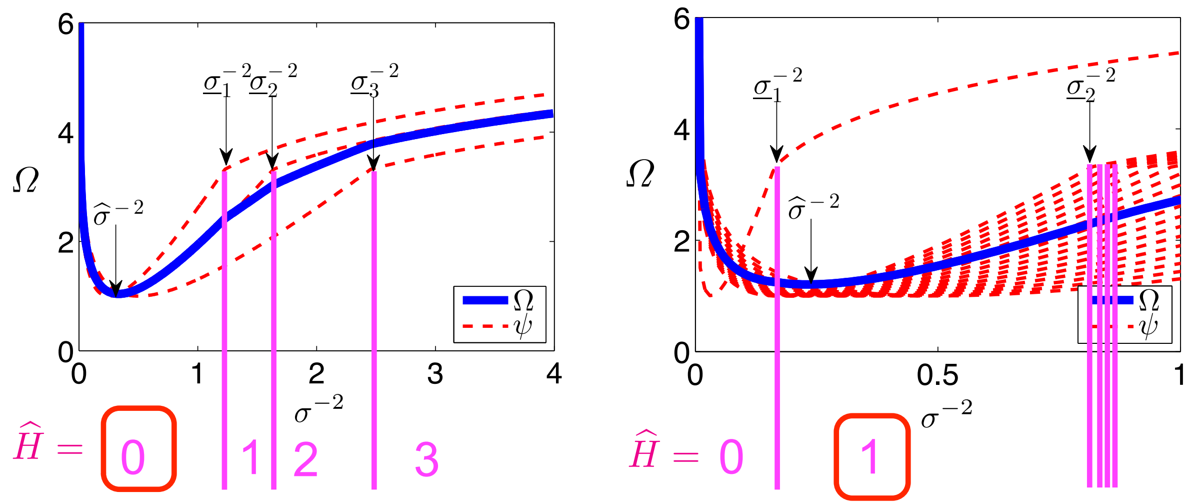
and κ is a function of x, defined by

$$\kappa(x) = \frac{1}{2\sqrt{\alpha}}\left((x - (1+\alpha)) + \sqrt{(x - (1+\alpha))^2 - 4\alpha}\right).$$





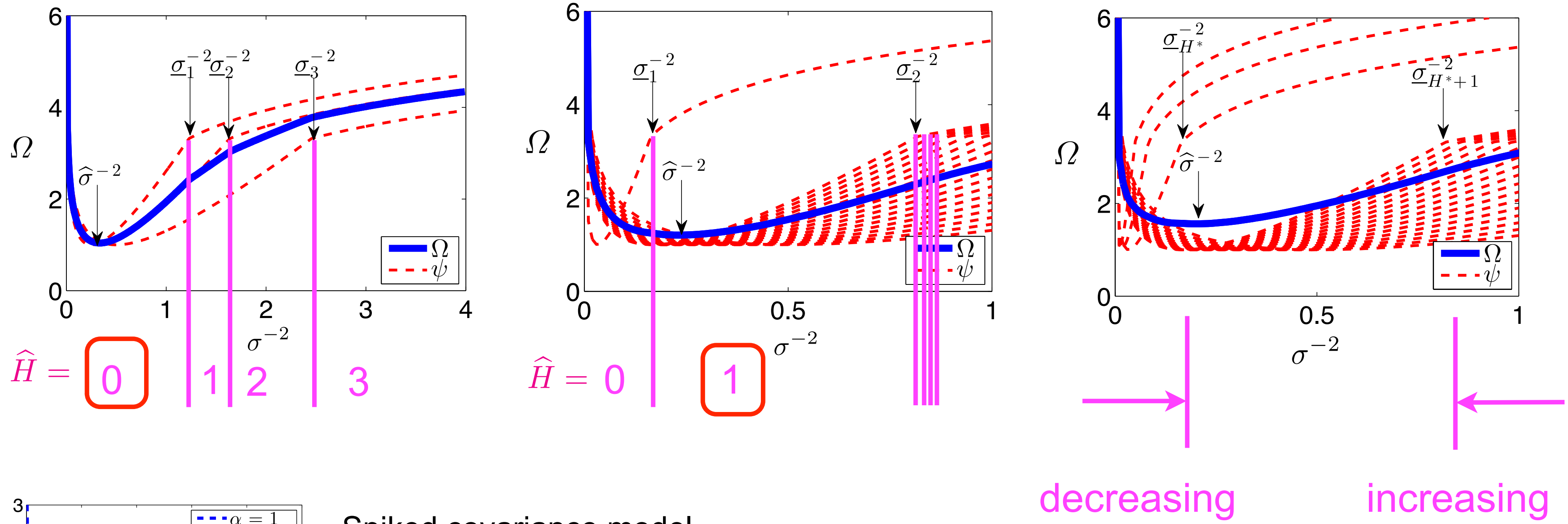
自由エネルギーはきれいに分解される！

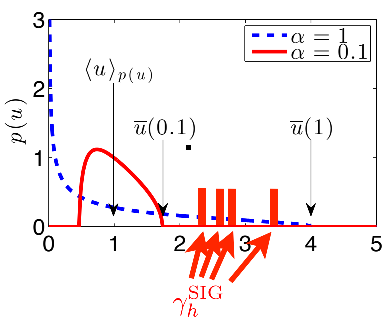


.

大域最小解とThresholdsの位置関係がランク推定値を決める！

自由エネルギーはきれいに分解される！





Spiked covariance model

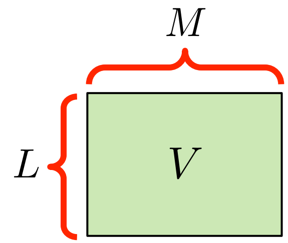
[Johnstone:AS2001, Baik&Silverstein:JMA2006]

$$V = U^* + \mathcal{E}$$

$$\Theta = \frac{1}{L}\frac{\partial \Omega}{\partial \sigma^{-2}} = \begin{cases} < 0 & \text{for } \sigma^{-2} < \underline{\sigma}_{H^*}^{-2} \\ > 0 & \text{for } \sigma^{-2} < \underline{\sigma}_{H^*+1}^{-2} \end{cases}$$

大域最小解とThresholdsの位置関係がランク推定値を決める！

正しいランク推定のための十分条件

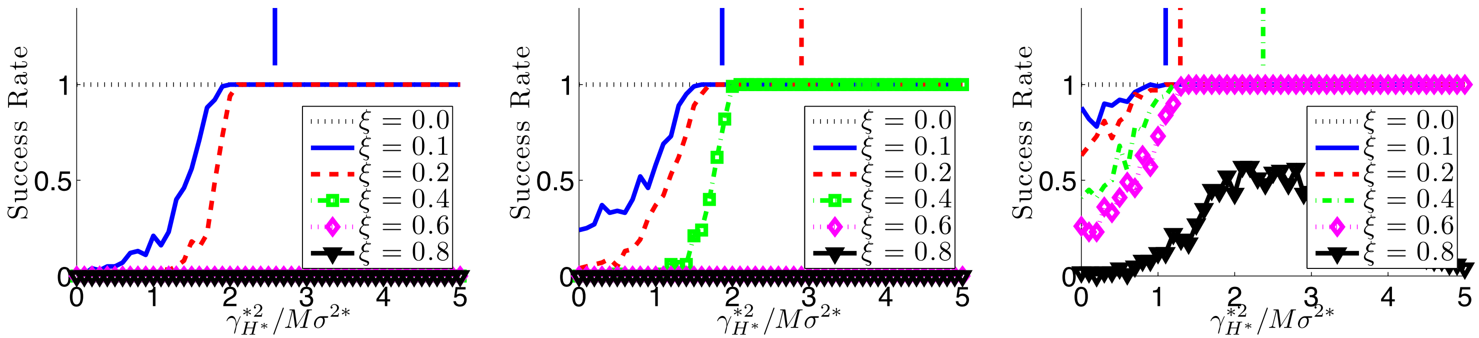


定理：

変分ベイズ主成分分析は、以下の条件を満たすとき
高確率で正しいランクを当てる。[Nakajima+:NIPS2012]

$$\alpha = \frac{L}{M}, \qquad \xi = \frac{H^*}{L}.$$

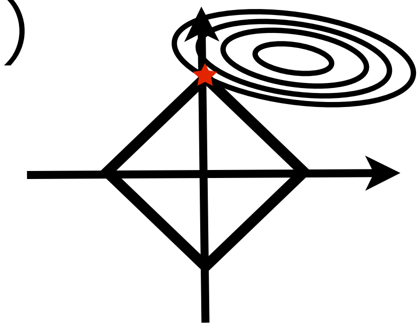
$$\xi < \frac{1}{\underline{x}} \quad \text{and} \quad \frac{\gamma_{H^*}^{*2}}{M\sigma^{2*}} > \left(\frac{\underline{x} - 1}{1 - \underline{x}\xi} - \alpha \right).$$



(a) $\alpha = 1$ (b) $\alpha = 0.5$ (c) $\alpha = 0.1$

変分ベイズ主成分分析のランク推定性能の理論保証！

凸形式 vs ベイズ！（低ランク行列推定）



凸形式： $\min_{U} \|V - U\|^2 + \lambda \|U\|_{\mathrm{tr}}$

確率モデル： $p(V|U) \propto \exp\left(-\frac{\|V-U\|^2}{2\sigma^2}\right)$

$$p(U) \propto \exp\left(-\frac{\|U\|_{\mathrm{tr}}}{2c}\right)$$

$\lambda = \frac{\sigma^2}{c}$ でMAPをやれば凸形式と等価！

ある条件下で、λを適切に選べばうまくいく。

凸問題
理論保証有
[Candes&Recht2008]

スパースベイズ法： $p(V|A,B) \propto \exp\left(-\frac{\|V - BA^\top\|^2_{\mathrm{Fro}}}{2\sigma^2}\right)$

$$p(A) \propto \exp\left(-\frac{\mathrm{tr}(AC_A^{-1}A^\top) + \mathrm{tr}(BC_B^{-1}B^\top)}{2}\right)$$

ある条件下でうまくいく！！！

~~非凸問題~~
~~理論保証無~~
解析解
理論保証有

変分ベイズの方が便利！

主成分分析では
変分ベイズの方が便利！

本日お話しすること

- ベイズモデル選択
- 変分ベイズ学習
- 行列分解モデル
  - 大域解析解
  - ランク推定性能の理論保証
- より一般のモデルへ

行列分解で大域解が求まった理由

- 多くの不要な自由度
  - 非自明な独立性の発見。
- 問題が分解可能
  - 解がreweighted SVDであることの発見。
- 停留条件が多項式系
  - がんばれば手で解ける。

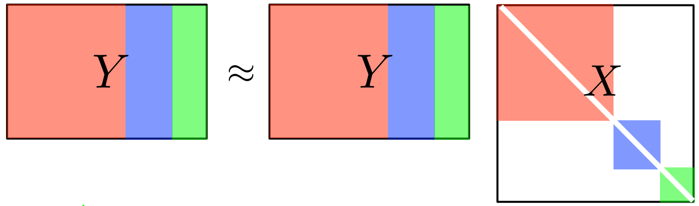
行列分解で大域解が求まった理由

- 多くの不要な自由度
  - 非自明な独立性の発見。
- 問題が分解可能
  - 解がreweighted SVDであることの発見。
- 停留条件が多項式系
  - がんばれば手で解ける。
  - Homotopy法で解ける。

Low-rank subspace clustering [Nakajima+:NIPS2013]

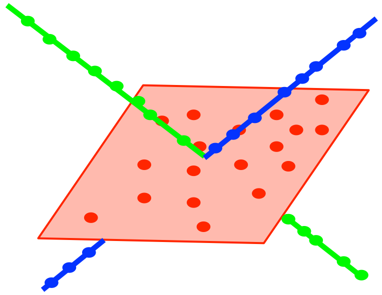
凸形式：[Liu+:ICML2010,Favaro+:CVPR2011,Babacan+:NIPS2012]

$$\min_{X} \|Y - YX\|_{\mathrm{Fro}}^2 + \lambda \|X\|_{\mathrm{tr}}.$$



確率モデル： [Babacan+:NIPS2012]

$$\begin{aligned} Y &= YX + \mathcal{E} \\ &= YBA^\top + \mathcal{E} \end{aligned}$$



1. O(J)個の多項式方程式をHomotopy法で解く

$$\widehat{a}_h = \frac{\gamma_h^2}{\sigma^2} \widehat{b}_h \sigma_{a_h}^2, \qquad \sigma_{a_h}^2 = \sigma^2 \left(\gamma_h^2 \widehat{b}_h^2 + \sum_{m=1}^{J} \gamma_m^2 \sigma_{B_{m,h}}^2 + \frac{\sigma^2}{c_{a_h}^2} \right)^{-1},$$

$$\widehat{b}_h = \frac{\gamma_h^2}{\sigma^2} \widehat{a}_h \sigma_{B_{h,h}}^2, \qquad \sigma_{B_{m,h}}^2 = \begin{cases} \sigma^2 \left(\gamma_m^2 \left(\widehat{a}_h^2 + M \sigma_{a_h}^2 \right) + \frac{\sigma^2}{c_{b_h}^2} \right)^{-1} & (m \le J), \\ c_{b_h}^2 & (m > J), \end{cases}$$

$$c_{a_h}^2 = \widehat{a}_h^2 / M + \sigma_{a_h}^2, \qquad c_{b_h}^2 = \left(\widehat{b}_h^2 + \sum_{m=1}^{J} \sigma_{B_{m,h}}^2 \right) / J.$$

2. 制約を追加してO(1)個の方程式を解く

$$\gamma_m^2 \sigma_{B_{m,h}}^2 = \overline{\sigma}_{b_h}^2 \quad \text{for all} \quad m \le J.$$

Homotopy法を用いた大域ソルバとその高速近似法を提案

変分ベイズ大域解法の拡張

- 多くの不要な自由度
  - 非自明な独立性の発見。

事前分布に相関がある場合、欠損値がある場合、テンソル分解などで条件式

- 問題が分解可能

Higher order SVDなどが使えないか？

  - 解がreweighted SVDであることの発見。

- 停留条件が多項式系
  - がんばれば手で解ける。
  - Homotopy法で解ける。

（マルチ）リニアモデルなら多項式系
Homotopy法の積極的な利用
多項式系でなくてもO(1)なら無理やり解ける？

もっと一般のモデル

- 混合分布モデル
- 隠れマルコフモデル
- ベイジアンネット

もっと一般のモデル

- 混合分布モデル [Watanabe&Watanabe:JMLR2006]
- 隠れマルコフモデル [Hosino+:IEICE2006]
- ベイジアンネット [Watanabe+:ML2009]

自由エネルギーの漸近挙動が解明されている。相転移現象の発見により、ハイパーパラメータ設定に指針を与えた！

ベイズ法とのKL距離が小さいことがわかった

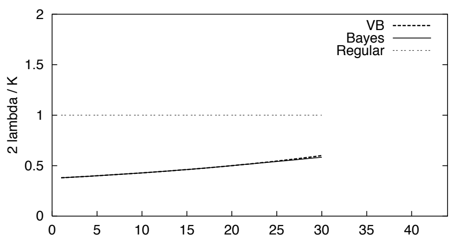
もっと一般のモデル

- 混合分布モデル [Watanabe&Watanabe:JMLR2006]
- 隠れマルコフモデル [Hosino+:IEICE2006]
- ベイジアンネット [Watanabe+:ML2009]

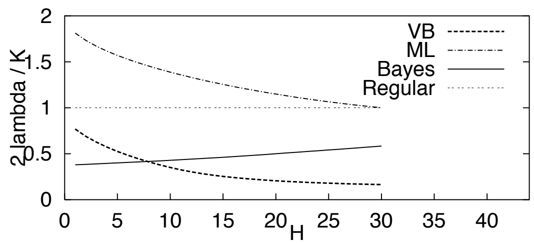
自由エネルギーの漸近挙動が解明されている。相転移現象の発見により、ハイパーパラメータ設定に指針を与えた！

ベイズ法とのKL距離が小さいことがわかったが、
汎化性能がベイズ法と同等という証明にはならない。

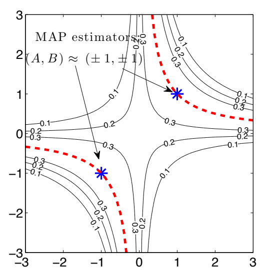
・縮小ランク回帰の自由エネルギーと汎化誤差：[Nakajima&Watanabe:NECO2007]



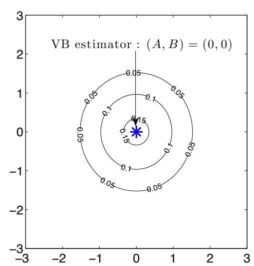
自由エネルギー



汎化誤差



Bayes



VB

まとめ

まとめ

- （全観測）行列分解モデルの変分ベイズ学習理論を紹介しました。
- 似たモデル（Subspace clustering, linear inverse problem, tensorなど）への拡張はいろいろとできそう。
- より一般のモデル（混合分布、隠れマルコフなど）への拡張はまだ見えず。
- その他、汎化誤差保証などもこれから。

中島 伸一, 杉山 将, "変分ベイズ学習理論の最新動向,"
日本応用数理学会論文誌, vol. 23, no. 3, pp.453-483, 2013,
http://sites.google.com/site/shinnkj23/home/manuscript_tjsiam2013.pdf.

ご清聴、ありがとうございました！

NIKON CORPORATION