機械学習に基づく推薦システム

岩田具治
NTT コミュニケーション科学基礎研究所

推薦システム

- ユーザの興味に合わせておすすめのアイテムを提示
- 推薦対象
  - 本，音楽，ニュース，映画，論文，検索語，レストラン，友達，恋人
- 推薦システムの効用
  - ユーザ：膨大な商品のなかから好みに合ったものを見つけられる
  - オンラインストア：売上の向上が期待できる

目次

1. 推薦システムの分類
2. 類似度に基づく手法
3. 行列分解に基づく手法
4. ベイズ推定を用いた行列分解
5. 行列分解の非線形拡張
6. 顧客生涯価値を高めるための推薦

推薦手法の分類

- 協調フィルタリング
  - ユーザの行動履歴から推薦
  - ○人手で情報の付与が不要
  - ×新しいユーザ，新しい商品には対応できない
- 内容ベースフィルタリング
  - アイテムの内容から推薦
  - ○新しい商品でも推薦可能
  - ×内容を用意が必要，ありきたりの推薦
- 統合アプローチ

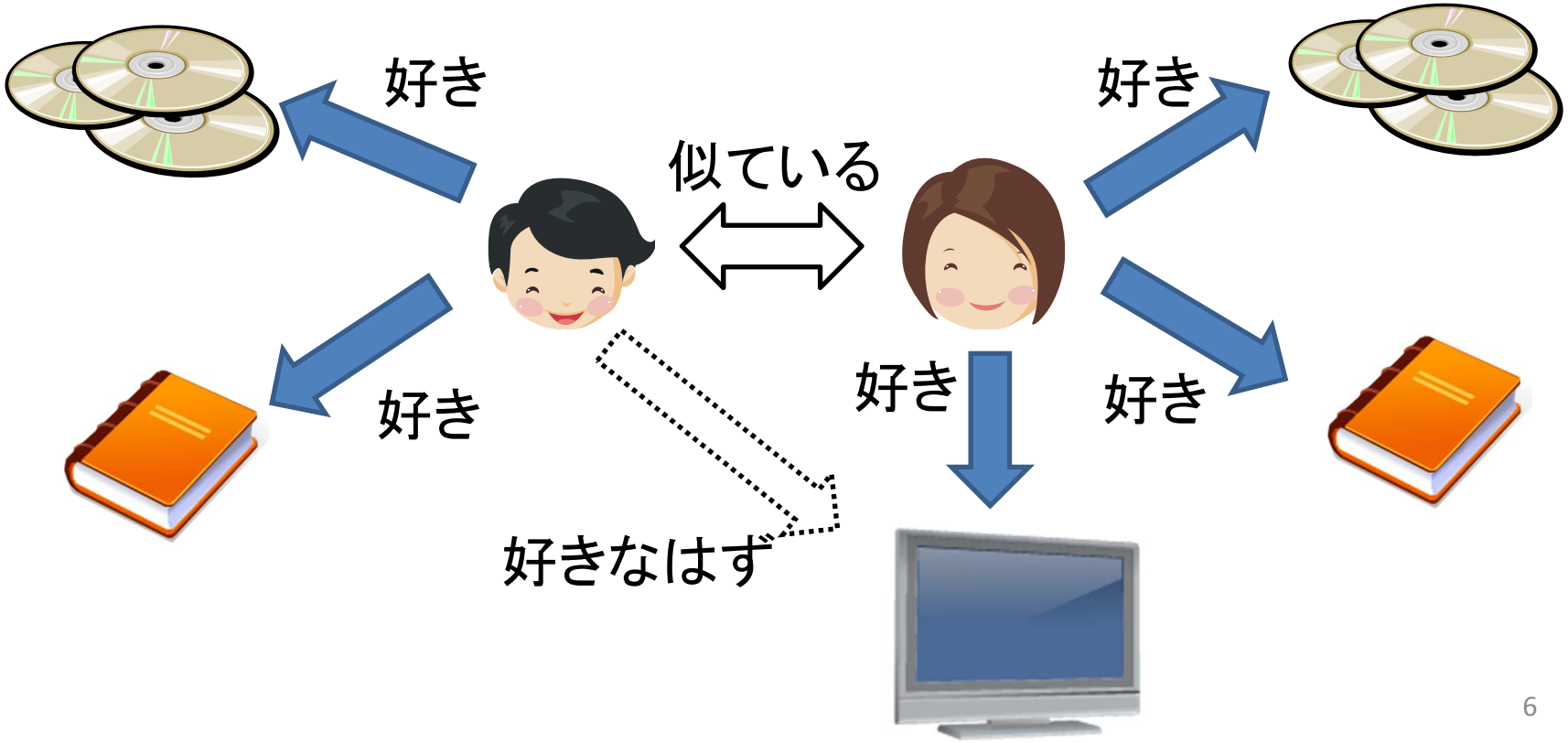
協調フィルタリング

- ユーザ×商品の行列の穴埋め問題
  - 評点データ：１～５、もしくは、未評価「?」
  - 未評価のうち最も評価が高いと推定される商品を推薦

	Aさん	Bさん	Cさん	Dさん
商品１	２	?	２	?
商品２	?	１	?	５
商品３	５	５	?	?
商品４	?	?	３	?

類似度に基づく手法

- 評点履歴が似ているユーザと同じような評点をつける



ユーザ類似度

- ユーザnとユーザn'の類似度（コサイン）

$$w_{nn'} = \frac{\boldsymbol{r}_n^\top \boldsymbol{r}_{n'}}{\| \boldsymbol{r}_n \| \| \boldsymbol{r}_{n'} \|}$$

ユーザn'の
評点ベクトル

- ユーザインデックス：$n \in \{1, \cdots, N\}$
- 商品インデックス：$m \in \{1, \cdots, M\}$

ユーザ数

商品数

- ユーザnの商品mに対する評点：r_{nm}
- ユーザnの評点ベクトル：$\boldsymbol{r}_n = (I_{n1} r_{n1}, \cdots, I_{nM} r_{nM})$
- ユーザnの商品mに対する評点の有無
  - $I_{nm} = 1$：評点あり、$I_{nm} = 0$：評点なし

評点推定

- 類似しているユーザの重み付き平均

ユーザnとユーザn'の類似度

$$r_{nm} = \sum_{n' \in \boldsymbol{U}_n} \frac{w_{nn'}}{\sum_{n'' \in \boldsymbol{U}_n} w_{n''n'}} r_{n'm}$$

ユーザnの類似ユーザ集合

ユーザn'の商品mに対する評点

類似度に基づく手法の問題点

- 類似度をどのように定義すればよいか自明でない
  - コサイン類似度？相関係数？
  - ユーザ間類似度？商品間類似度？
- 様々な類似度を試して精度がよいものを選択
  - データによって最良の類似度は異なる

機械学習アプローチ

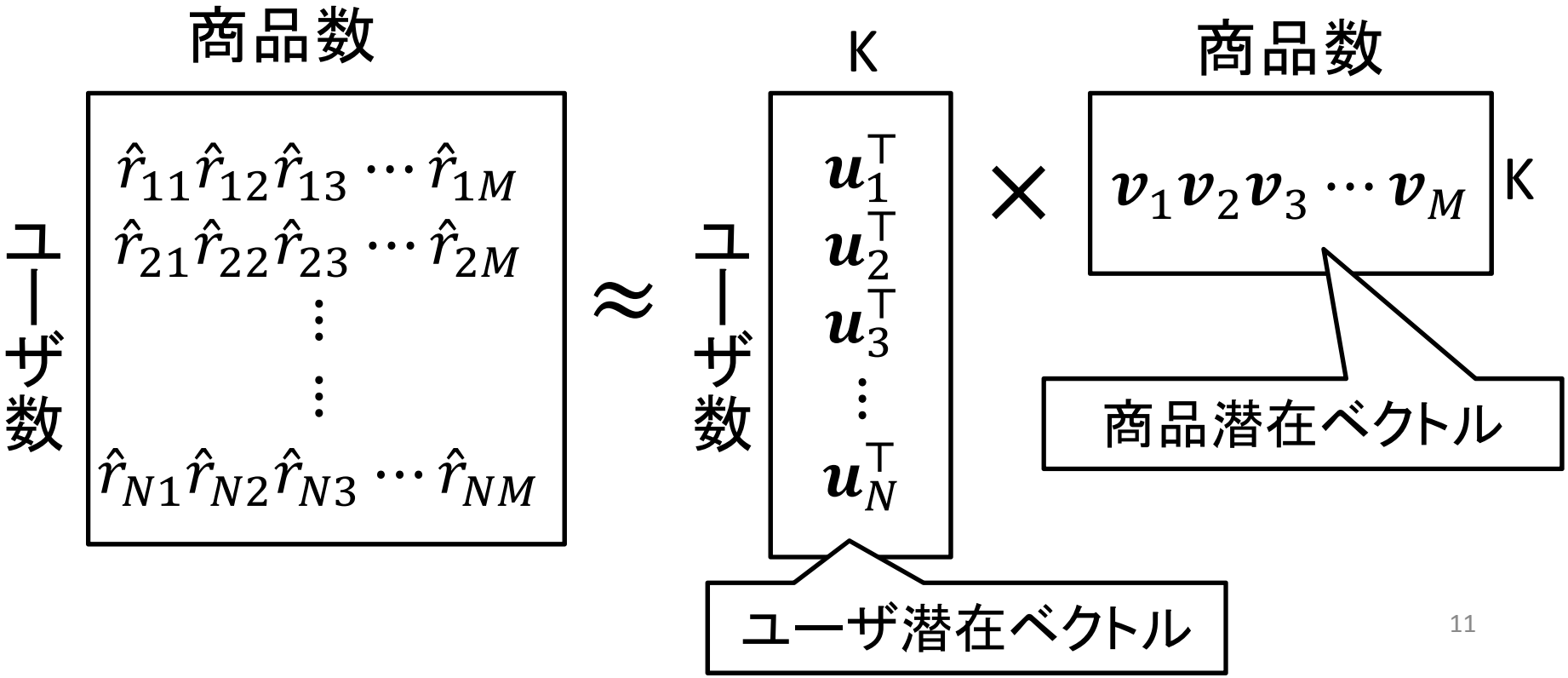
- 目的を達成するようにデータを用いてモデルを学習
- 協調フィルタリングの目的
  - 評点の誤差を小さくする
  - 二乗誤差：$\sum_{n=1}^{N}\sum_{m=1}^{M} I_{nm} \parallel r_{nm} - \hat{r}_{nm} \parallel^2$

評点の有無　正解の評点　推定の評点

行列分解に基づく手法

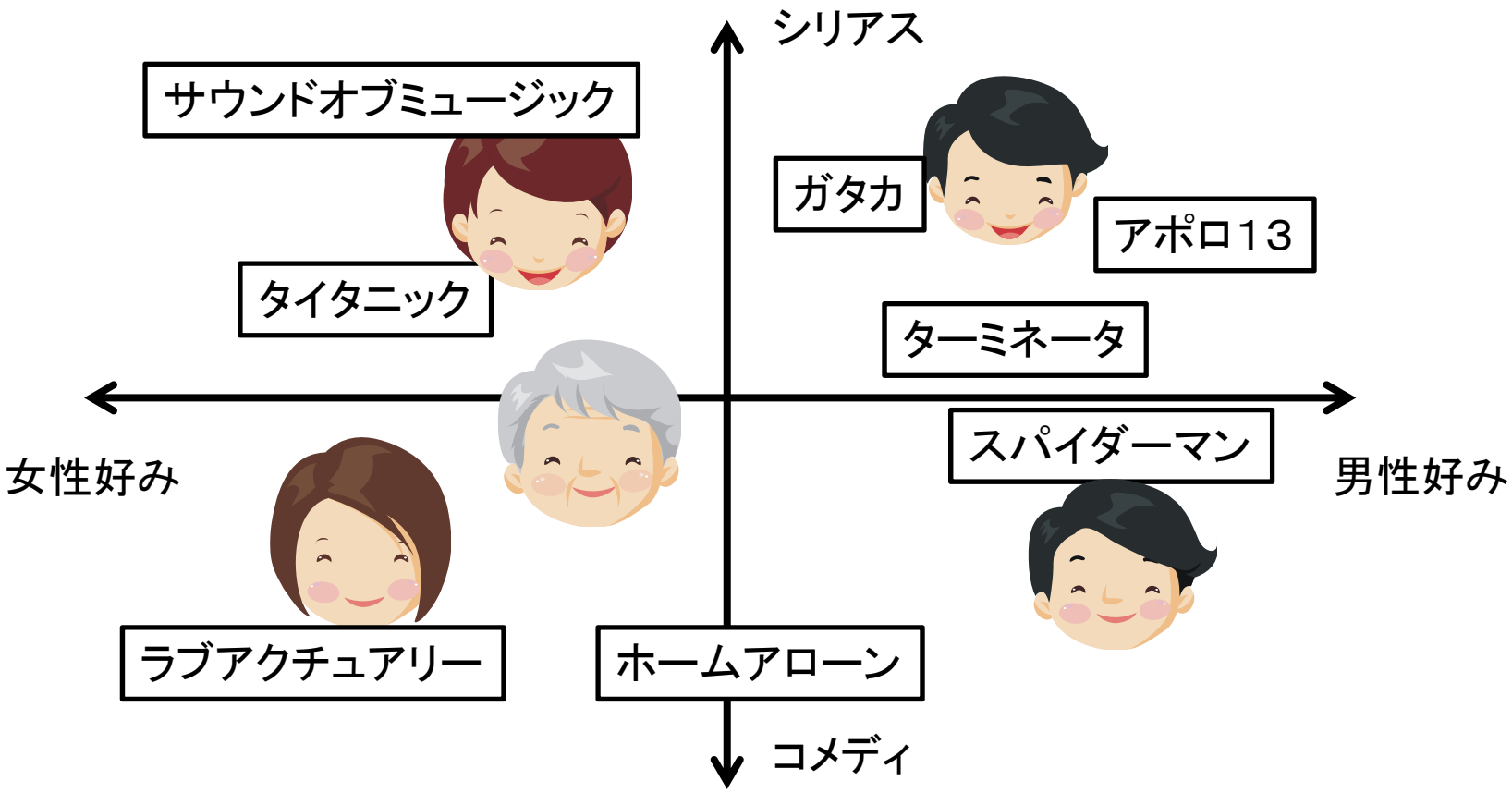
- ユーザnと商品mの評点r_{nm}を潜在ベクトルの内積で近似

$$\hat{r}_{nm} \approx \boldsymbol{u}_n^\top \boldsymbol{v}_m = \sum_{k=1}^{K} u_{nk} v_{mk}$$



行列分解に基づく手法

- 各ユーザ、各商品を潜在ベクトルで表現



最小化すべき目的関数

- 評点がある部分の正解と予測の二乗誤差＋正則化項

評点の有無　正解の評点　推定の評点

$$E = \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M} I_{nm} \parallel r_{nm} - \boldsymbol{u}_n^{\top}\boldsymbol{v}_m \parallel^2$$

$$+\frac{\lambda}{2}\sum_{n=1}^{N} \parallel \boldsymbol{u}_n \parallel^2 + \frac{\lambda}{2}\sum_{m=1}^{M} \parallel \boldsymbol{v}_m \parallel^2$$

ユーザ潜在ベクトル　商品潜在ベクトル

潜在ベクトルの推定

- 勾配法で目的関数Eを最小化するU,Vを推定

$$\boldsymbol{u}_n \leftarrow \boldsymbol{u}_n - \eta \frac{\partial E}{\partial \boldsymbol{u}_n}$$

$$\boldsymbol{v}_m \leftarrow \boldsymbol{v}_m - \eta \frac{\partial E}{\partial \boldsymbol{v}_m}$$

比較実験

- データ：MovieLens 1M
- 評価尺度：RMSE（平均二乗誤差の平方根; Root Mean Square Error）

手法	誤差
ユーザ平均	1.036
商品平均	0.983
類似度に基づく手法	0.871
行列分解	0.857

http://mymedialite.net/examples/datasets.html

ベイズ推定を用いた行列分解

Bayesian Probabilistic Matrix Factorization using MCMC.
Ruslan Salakhutdinov and Andriy Mnih, ICML2008

行列分解の確率的解釈

- 二乗誤差の最小化＝正規分布尤度の最大化

- 正規分布

$$\mathrm{Normal}(r_{nm}|\boldsymbol{u}_n^\top \boldsymbol{v}_m, \alpha^{-1})$$
$$= \left(\frac{\alpha}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\alpha}{2} \parallel r_{nm} - \boldsymbol{u}_n^\top \boldsymbol{v}_m \parallel^2\right)$$

- 対数尤度

$$\log \mathrm{Normal}(r_{nm}|\boldsymbol{u}_n^\top \boldsymbol{v}_m, \alpha^{-1}) \propto -\frac{1}{2} \parallel r_{nm} - \boldsymbol{u}_n^\top \boldsymbol{v}_m \parallel^2$$

行列分解の確率的解釈

- L2正則化＝事前分布が正規分布の事後確率最大化（MAP）推定

- 正規分布

$$\mathrm{Normal}(\boldsymbol{u}_n|0,\lambda^{-1}\boldsymbol{I}) = \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \exp(-\frac{\lambda}{2} \parallel \boldsymbol{u}_n \parallel^2)$$

- 対数

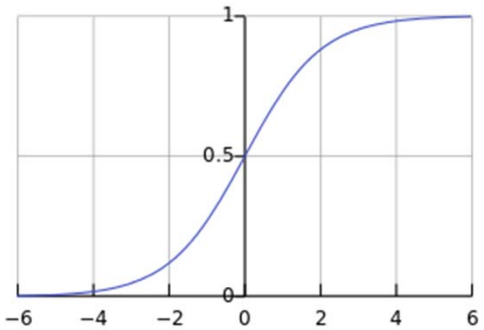
$$\log \mathrm{Normal}(\boldsymbol{u}_n|0,\lambda^{-1}\boldsymbol{I}) \propto -\frac{\lambda}{2} \parallel \boldsymbol{u}_n \parallel^2$$

行列分解の確率的解釈

- 確率モデルの事後確率最大化（MAP）推定
- for ユーザ n=1,…,N
  - ユーザ潜在ベクトルを生成
    - $\boldsymbol{u}_n \sim \mathrm{Normal}(0, \lambda^{-1}\boldsymbol{I})$
- for 商品 m=1,…,M
  - 商品潜在ベクトルを生成
    - $\boldsymbol{v}_n \sim \mathrm{Normal}(0, \lambda^{-1}\boldsymbol{I})$
- for ユーザ n=1,…,N
  - for 商品 m: $I_{nm} = 1$
    - 評点を生成
      - $r_{nm} \sim \mathrm{Normal}(\boldsymbol{u}_n^\top \boldsymbol{v}_m, \alpha^{-1})$

0-1データの場合の行列分解

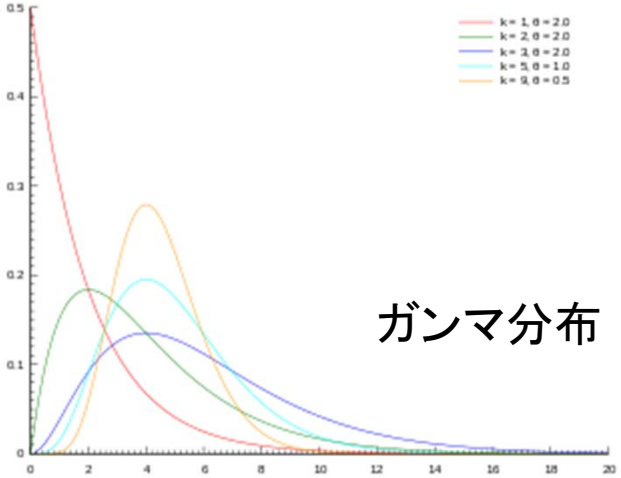
- シグモイド関数で0-1に変換
  - $p(r_{nm} = 1|\boldsymbol{u}_n, \boldsymbol{v}_m) = \frac{1}{1+\exp(-\boldsymbol{u}_n^\top \boldsymbol{v}_m)}$
- ベルヌーイ分布
  - $\mathrm{Bernoulli}(r_{nm}|\boldsymbol{u}_n, \boldsymbol{v}_m) = \left(\frac{1}{1+\exp(-\boldsymbol{u}_n^\top \boldsymbol{v}_m)}\right)^{r_{nm}} \left(1 - \frac{1}{1+\exp(-\boldsymbol{u}_n^\top \boldsymbol{v}_m)}\right)^{1-r_{nm}}$

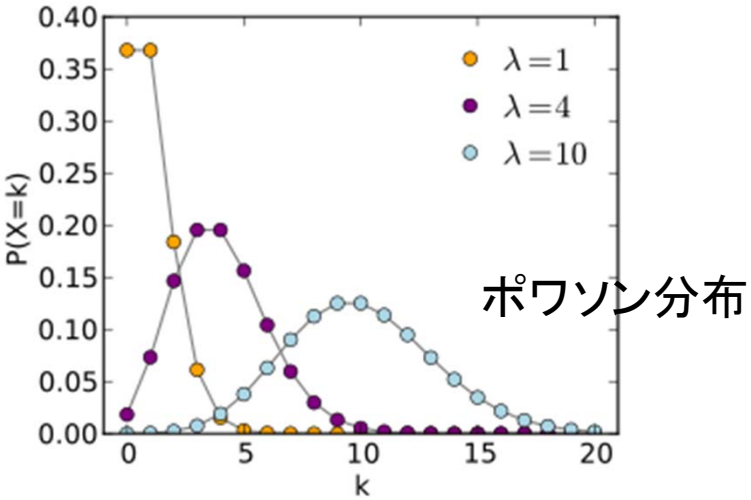


http://en.wikipedia.org/wiki/Sigmoid_function

非負値行列分解

- 潜在ベクトルをガンマ分布で生成
  - $p(\boldsymbol{u}_n|a,b) = \mathrm{Gamma}(a,b)$
  - $p(\boldsymbol{v}_m|a,b) = \mathrm{Gamma}(a,b)$
- ポワソン分布で評点を生成
  - $\mathrm{Poisson}(r_{nm}|\boldsymbol{u}_n,\boldsymbol{v}_m) = \dfrac{\exp(-\boldsymbol{u}_n^\top \boldsymbol{v}_m)(\boldsymbol{u}_n^\top \boldsymbol{v}_m)^{r_{nm}}}{\Gamma(r_{nm}+1)}$





http://en.wikipedia.org/wiki/[Gamma,Poisson]_distribution

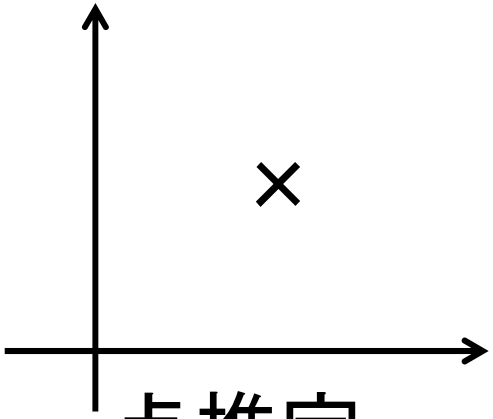
ベイズ推定

- 未知パラメータの事後分布をベイズの定理を用いて推定

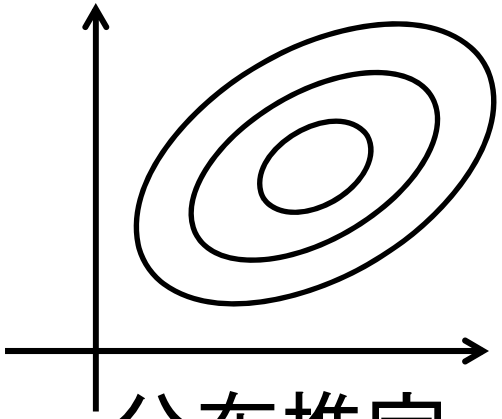
事後分布　事前分布　尤度

$$p(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{R}) = \frac{p(\boldsymbol{U},\boldsymbol{V})p(\boldsymbol{R}|\boldsymbol{U},\boldsymbol{V})}{p(\boldsymbol{R})}$$

- 推定値の確信度も考慮できる



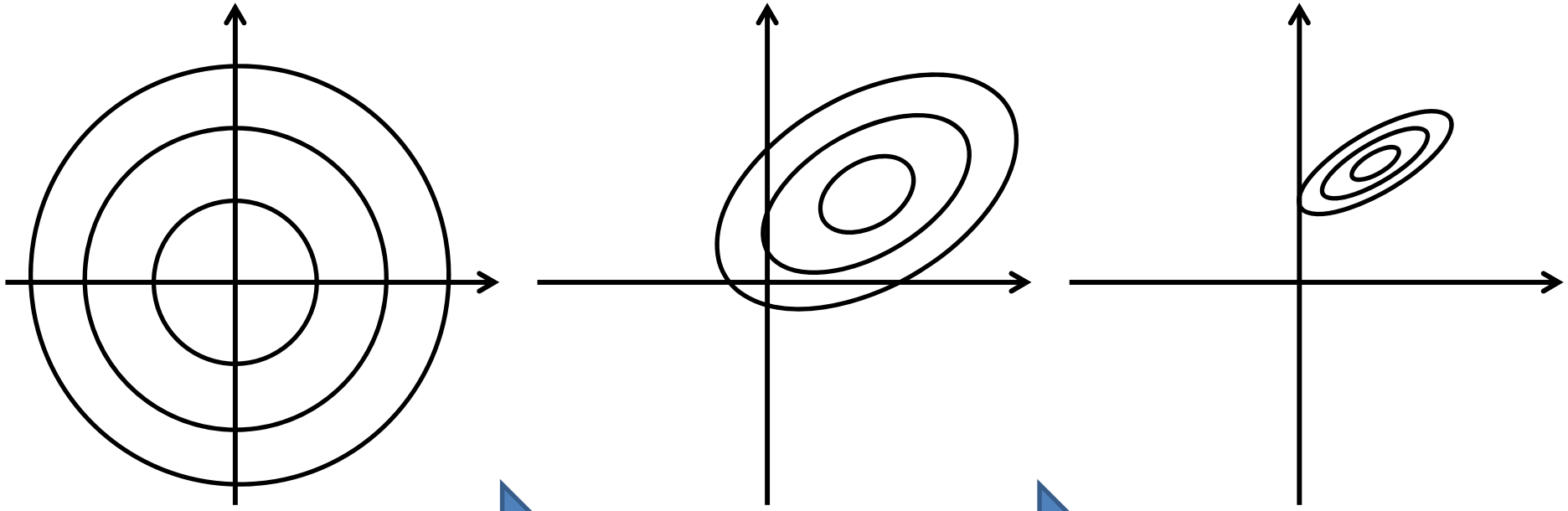
点推定



分布推定

ベイズ推定

- 一般にデータを観測するごとに事後分布は狭まっていく（確信が高まっていく）



$p(\boldsymbol{u}_n)$
観測なし
事前分布

$p(\boldsymbol{u}_n|\boldsymbol{R}_1)$
少し観測
事後分布１

$p(\boldsymbol{u}_n|\boldsymbol{R}_1, \boldsymbol{R}_2)$
たくさん観測
事後分布２

ベイズ予測

- 事後分布を用いて推定値の期待値で予測
  - 観測データ数が少ない場合に特によい

$$\hat{r}_{nm} = \int \boldsymbol{u}_n^\top \boldsymbol{v}_m p(\boldsymbol{u}_n|\boldsymbol{R}) p(\boldsymbol{v}_m|\boldsymbol{R}) d\boldsymbol{u}_n d\boldsymbol{v}_m$$

- 点推定：事後分布としてデルタ関数を想定

$$\hat{r}_{nm} = \widehat{\boldsymbol{u}}_n^\top \widehat{\boldsymbol{v}}_m$$

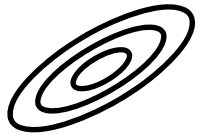
ベイズ行列分解

- 潜在ベクトルの共役事前分布を導入
  - 共役事前分布：事後分布の型が同じになる事前分布
  - 正規分布の平均の共役事前分布：正規分布
  - 正規分布の精度行列の共役事前分布：ウィッシャート分布
- ユーザ潜在ベクトルの精度行列を生成
  - $\boldsymbol{\Lambda}_{\mathrm{u}} \sim \mathrm{Wishart}(\boldsymbol{W}_0, \nu_0)$
- ユーザ潜在ベクトルの平均を生成
  - $\boldsymbol{\mu}_{\mathrm{u}} \sim \mathrm{Normal}(\boldsymbol{\mu}_0, (\beta_0 \boldsymbol{\Lambda}_{\mathrm{u}})^{-1})$
- for ユーザ n=1,…,N
  - ユーザ潜在ベクトルを生成
    - $\boldsymbol{u}_n \sim \mathrm{Normal}(\boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{\Lambda}_{\mathrm{u}}^{-1})$

ウィッシャート分布

$$\mathrm{Wishart}(\boldsymbol{\Lambda}|\boldsymbol{W}_0, \nu_0)$$
$$\propto |\boldsymbol{\Lambda}|^{\frac{1}{2}(\nu_0 - K - 1)} \exp(-\frac{1}{2}\mathrm{tr}(W_0^{-1}\boldsymbol{\Lambda}))$$

- 正定値行列の確率分布



- 正規分布$\mathrm{Normal}(\boldsymbol{x}|0, \boldsymbol{W}_0)$に従う
  ν_0個の$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{\boldsymbol{\nu_0}}$があったとき、
  $\boldsymbol{\Lambda} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{\nu_0}} \boldsymbol{x}_\mathrm{i}\boldsymbol{x}_\mathrm{i}^\top$は$\mathrm{Wishart}(\boldsymbol{\Lambda}|\boldsymbol{W}_0, \nu_0)$に従う
- 正規分布の精度行列の共役事前分布

ベイズ行列分解

- 商品潜在ベクトルの精度行列を生成
  - $\boldsymbol{\Lambda}_{\mathrm{v}} \sim \mathrm{Wishart}(\boldsymbol{W}_0, \nu_0)$
- 商品潜在ベクトルの平均を生成
  - $\boldsymbol{\mu}_{\mathrm{v}} \sim \mathrm{Normal}(\boldsymbol{\mu}_0, (\beta_0 \boldsymbol{\Lambda}_{\mathrm{v}})^{-1})$
- for 商品 m=1,…,M
  - 商品潜在ベクトルを生成
    - $\boldsymbol{v}_m \sim \mathrm{Normal}(\boldsymbol{\mu}_{\mathrm{v}}, \boldsymbol{\Lambda}_{\mathrm{v}}^{-1})$

未知変数

- ユーザ潜在ベクトルの精度行列:$\boldsymbol{\Lambda}_{\mathrm{u}}$
- ユーザ潜在ベクトルの平均:$\boldsymbol{\mu}_{\mathrm{u}}$
- 商品潜在ベクトルの精度行列:$\boldsymbol{\Lambda}_{\mathrm{v}}$
- 商品潜在ベクトルの平均:$\boldsymbol{\mu}_{\mathrm{v}}$
- ユーザ潜在ベクトル:$\boldsymbol{U} = \boldsymbol{u}_1, \cdots, \boldsymbol{u}_N$
- 商品潜在ベクトル:$\boldsymbol{V} = \boldsymbol{v}_1, \cdots, \boldsymbol{v}_M$

- 求めたい事後確率

データ

$$p(\boldsymbol{\Lambda}_{\mathrm{u}}, \boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{\Lambda}_{\mathrm{v}}, \boldsymbol{\mu}_{\mathrm{v}}, \boldsymbol{U}, \boldsymbol{V} | \boldsymbol{R})$$

ギブスサンプリング

- マルコフ連鎖モンテカルロ法（MCMC）の一種
- 無限個のサンプルを用いると真の事後分布が得られる（⇔変分ベイズ法）
- ある未知変数以外の変数が与えられたときのその未知変数の条件付確率を用いてサンプリングを繰り返す

1. $p(\boldsymbol{\Lambda}_{\mathrm{u}}|\boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{\Lambda}_{\mathrm{v}}, \boldsymbol{\mu}_{\mathrm{v}}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{R})$
2. $p(\boldsymbol{\mu}_{\mathrm{u}}|\boldsymbol{\Lambda}_{\mathrm{u}}, \boldsymbol{\Lambda}_{\mathrm{v}}, \boldsymbol{\mu}_{\mathrm{v}}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{R})$
3. $p(\boldsymbol{\Lambda}_{\mathrm{v}}|\boldsymbol{\Lambda}_{\mathrm{u}}, \boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{\mu}_{\mathrm{v}}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{R})$
4. $p(\boldsymbol{\mu}_{\mathrm{v}}|\boldsymbol{\Lambda}_{\mathrm{u}}, \boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{\Lambda}_{\mathrm{v}}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{R})$
5. $p(\boldsymbol{u}_n|\boldsymbol{\Lambda}_{\mathrm{u}}, \boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{\Lambda}_{\mathrm{v}}, \boldsymbol{\mu}_{\mathrm{v}}, \boldsymbol{U}_{\backslash n}, \boldsymbol{V}, \boldsymbol{R}), n = 1, \cdots, N$
6. $p(\boldsymbol{v}_m|\boldsymbol{\Lambda}_{\mathrm{u}}, \boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{\Lambda}_{\mathrm{v}}, \boldsymbol{\mu}_{\mathrm{v}}, \boldsymbol{U}, \boldsymbol{V}_{\backslash m}, \boldsymbol{R}), m = 1, \cdots, M$

mを除く

ユーザ潜在ベクトル事後確率

$$p(\boldsymbol{u}_n|\boldsymbol{\Lambda}_{\mathrm{u}}, \boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{V}, \boldsymbol{R}) = \mathrm{Normal}(\boldsymbol{\mu}_n^*, \boldsymbol{\Lambda}_n^{*\,-1})$$

- 事後分布の精度

$$\boldsymbol{\Lambda}_n^* = \boldsymbol{\Lambda}_{\mathrm{u}} + \alpha \sum_{m=1}^{M} I_{nm} \boldsymbol{v}_m \boldsymbol{v}_m^\top$$

- 事後分布の平均

$$\boldsymbol{\mu}_n^* = \boldsymbol{\Lambda}_n^{*\,-1}(\alpha \textstyle\sum_{m=1}^{M} I_{nm} \boldsymbol{v}_m r_{ij} + \boldsymbol{\Lambda}_{\mathrm{u}} \boldsymbol{\mu}_{\mathrm{u}})$$

商品潜在ベクトル事後確率

$$p(\boldsymbol{v}_m|\boldsymbol{\Lambda}_{\mathrm{v}}, \boldsymbol{\mu}_{\mathrm{v}}, \boldsymbol{U}, \boldsymbol{R}) = \mathrm{Normal}(\boldsymbol{\mu}_m^*, {\boldsymbol{\Lambda}_m^*}^{-1})$$

- 事後分布の精度

$$\boldsymbol{\Lambda}_m^* = \boldsymbol{\Lambda}_{\mathrm{v}} + \alpha \sum_{n=1}^{N} I_{nm} \boldsymbol{u}_n \boldsymbol{u}_n^\top$$

- 事後分布の平均

$$\boldsymbol{\mu}_m^* = {\boldsymbol{\Lambda}_m^*}^{-1} (\alpha \textstyle\sum_{n=1}^{N} I_{nm} \boldsymbol{u}_n r_{ij} + \boldsymbol{\Lambda}_{\mathrm{v}} \boldsymbol{\mu}_{\mathrm{v}})$$

潜在ベクトル平均・精度事後分布

$$p(\boldsymbol{\mu}_{\mathrm{u}}, \boldsymbol{\Lambda}_{\mathrm{u}} | \boldsymbol{U})$$
$$= \mathrm{Normal}(\boldsymbol{\mu}_{\mathrm{u}} | \boldsymbol{\mu}_0^*, (\beta_0^* \boldsymbol{\Lambda}_{\mathrm{u}})^{-1}) \mathrm{Wishart}(\boldsymbol{\Lambda}_{\mathrm{u}} | \mathbf{W}_0^*, \nu_0^*)$$

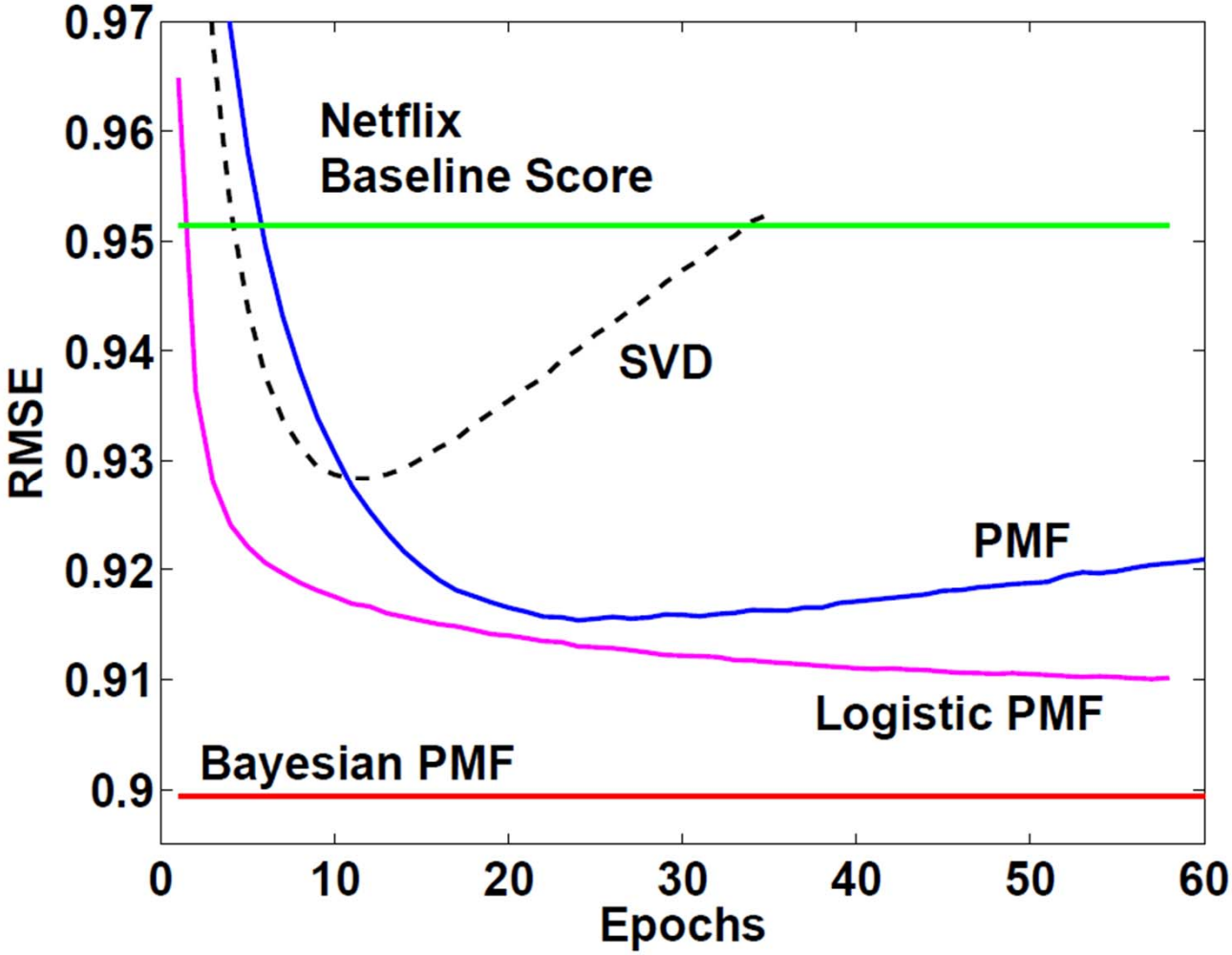
$$\beta_0^* = \beta_0 + N$$
$$\nu_0^* = \nu_0 + N$$
$$\boldsymbol{\mu}_0^* = \frac{\beta_0 \boldsymbol{\mu}_0 + N \overline{\boldsymbol{u}}}{\beta_0 + N}$$
$$\mathbf{W}_0^{*-1} = \boldsymbol{W}_0^{-1} + N \boldsymbol{S} + \frac{\beta_0 N}{\beta_0 + N} (\boldsymbol{\mu}_0 - \overline{\boldsymbol{u}})(\boldsymbol{\mu}_0 - \overline{\boldsymbol{u}})^\top$$

uの分散

uの平均

比較実験

誤差

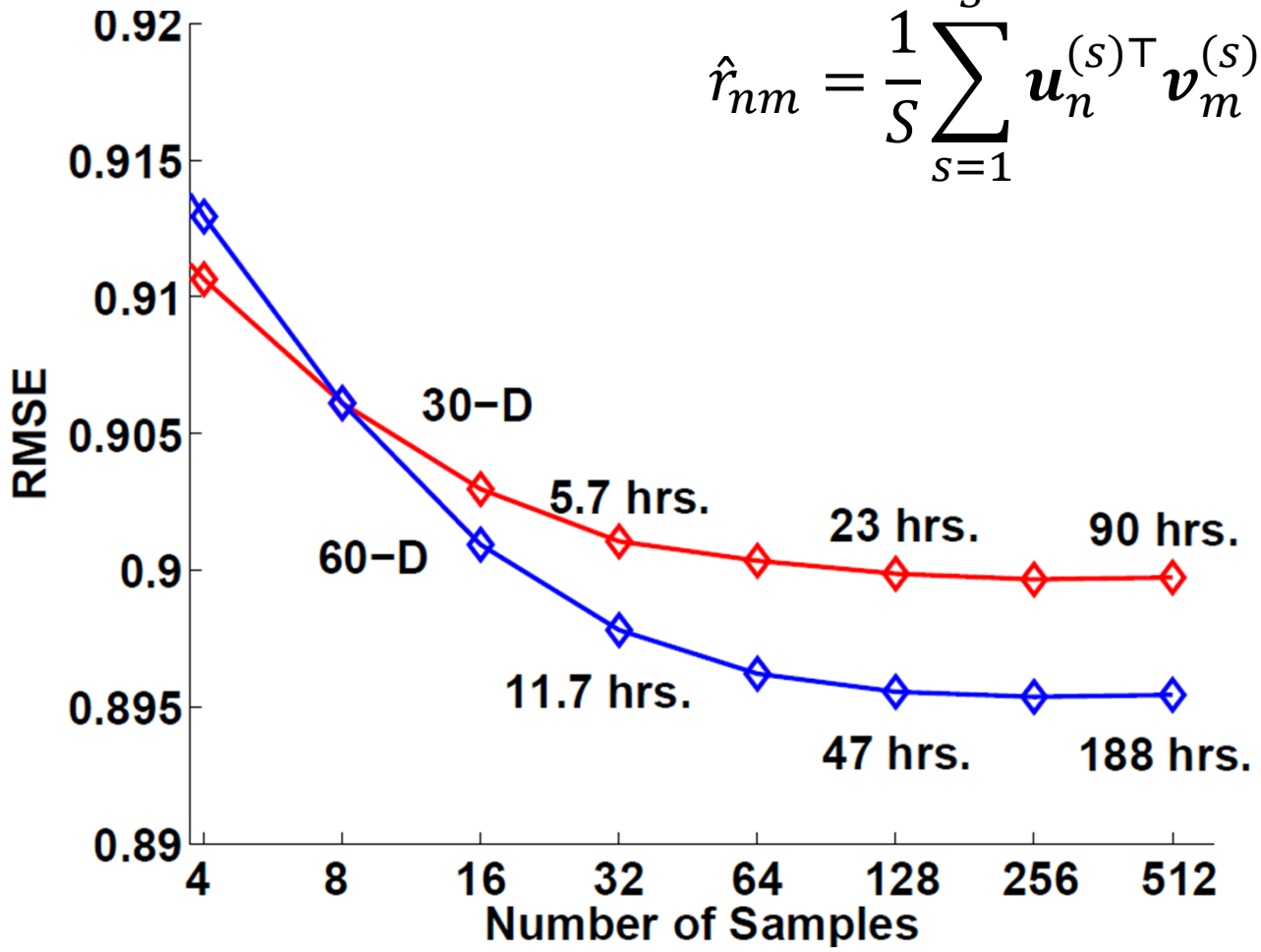


学習ステップ数

サンプル数による誤差の改善

$$\hat{r}_{nm} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{u}_n^{(s)\top} \boldsymbol{v}_m^{(s)}$$

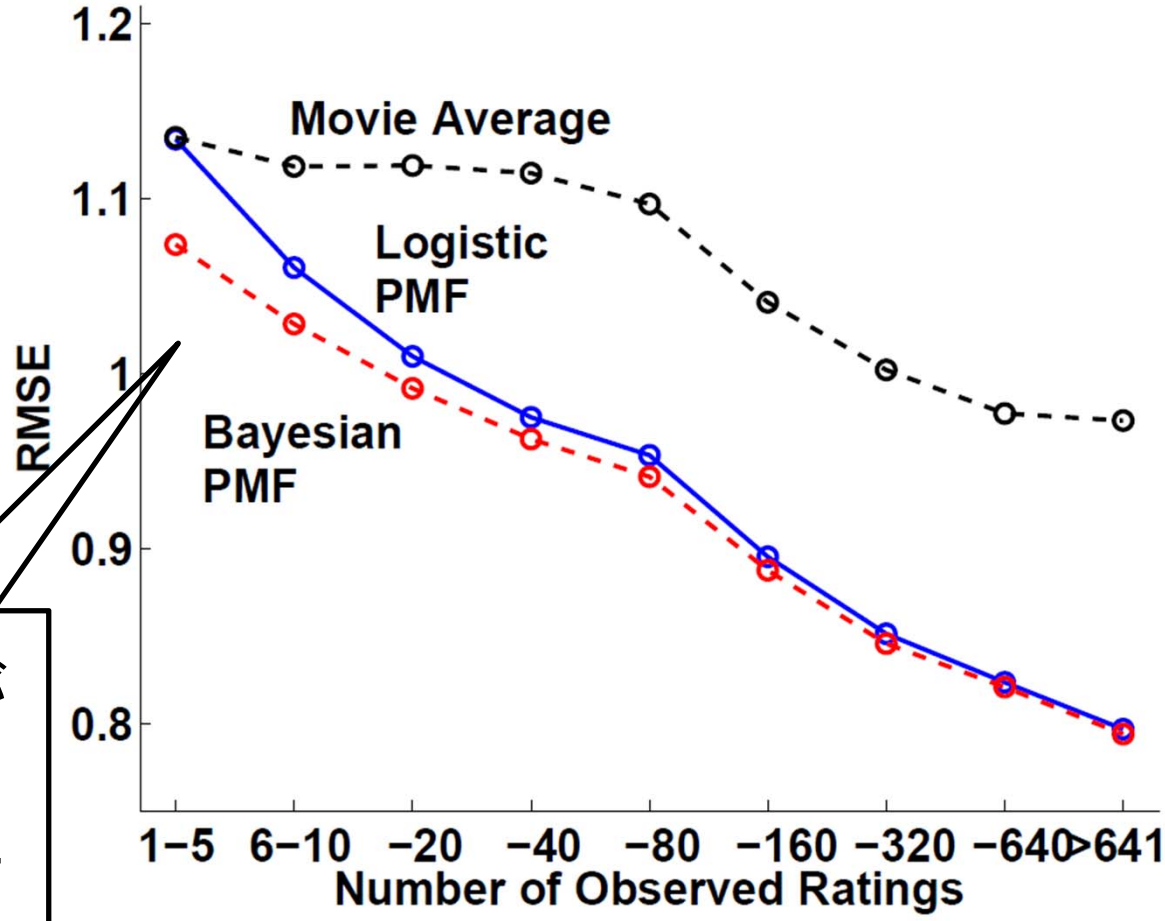
誤差



使用サンプル数

観測評点数毎の誤差

誤差

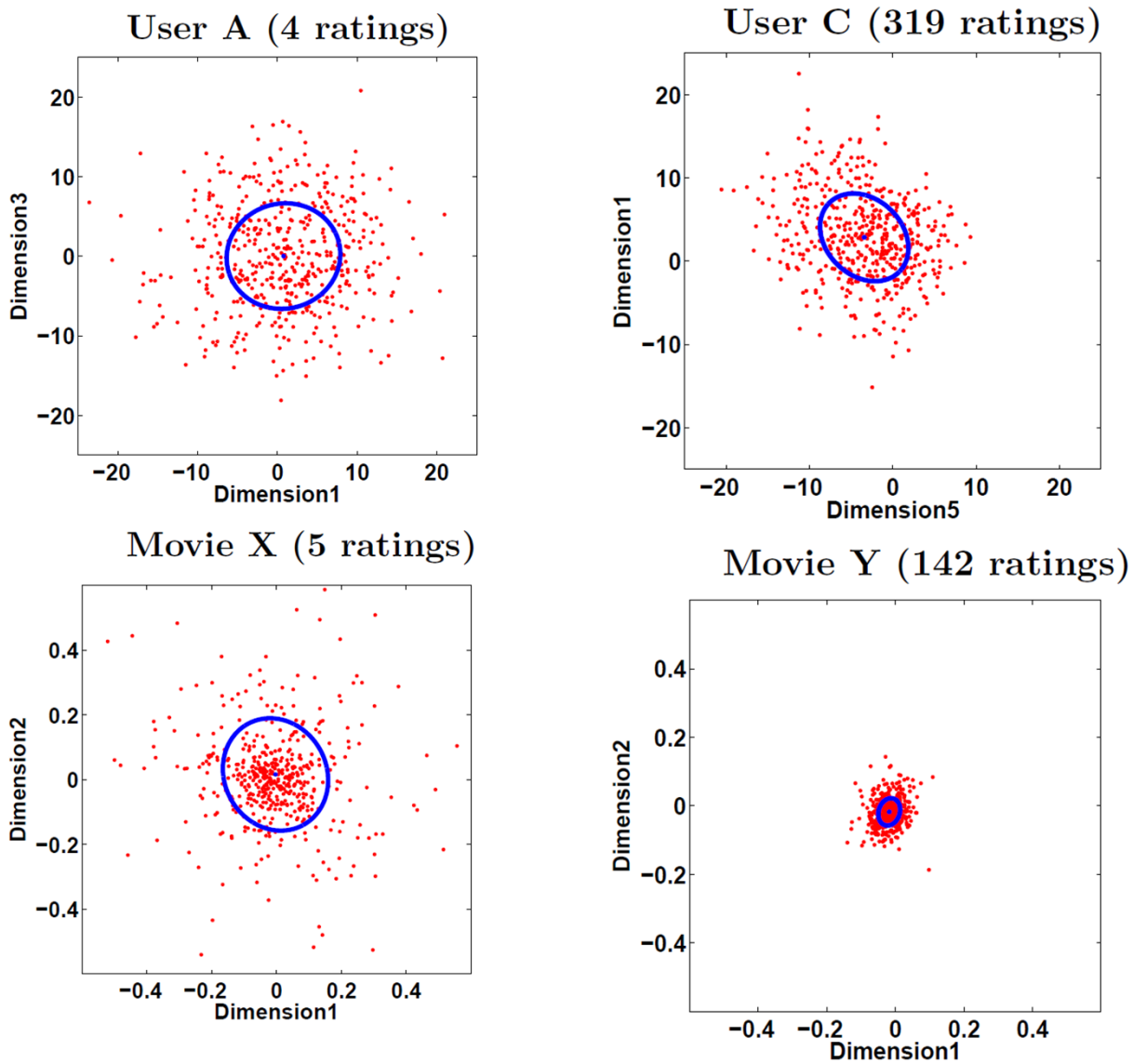


評点数が少ない場合に特によい

ユーザの観測評点数

推定された潜在ベクトル事後分布

データが
少ない



データが
多い

行列分解の
ガウス過程による非線形拡張

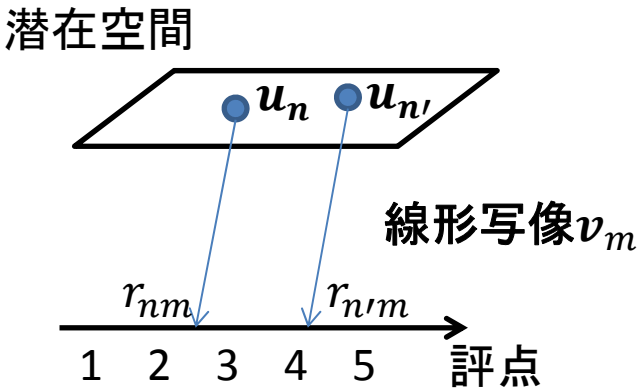
Non-linear matrix factorization with Gaussian processes
Neil Lawrence, and Rquel Urtasun, ICML 2009

線形モデルとしての行列分解

- 線形関数$f_m(\boldsymbol{u}) = \boldsymbol{u}^\top \boldsymbol{v}_m$で商品mの評点を予測

$$p(r_{nm}|\boldsymbol{u}_n, \boldsymbol{v}_m, \alpha) = \mathrm{Normal}(\boldsymbol{r}_n|\boldsymbol{u}_n^\top \boldsymbol{v}_m, \alpha^{-1}\boldsymbol{I})$$

$$\boxed{r_{nm}} = \underbrace{\boxed{\boldsymbol{u}_n^\top}}_{\text{潜在次元数}} \times \boxed{\boldsymbol{v}_m} \Big\} \text{潜在次元数}$$

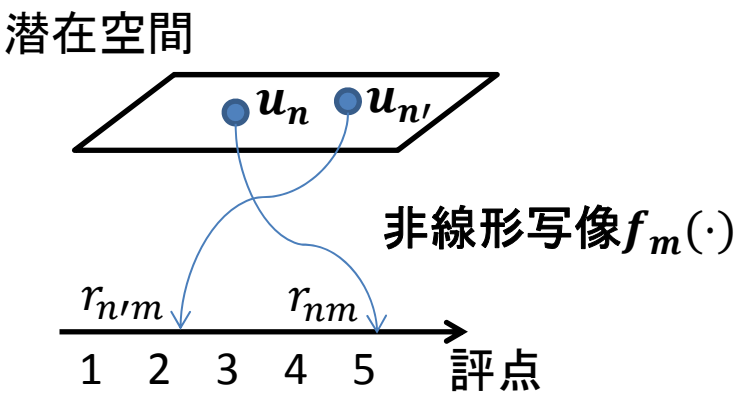


非線形関数

- 潜在ベクトル$\boldsymbol{u}_n$と非線形関数$f_m(\boldsymbol{u})$で評点r_{nm}を予測

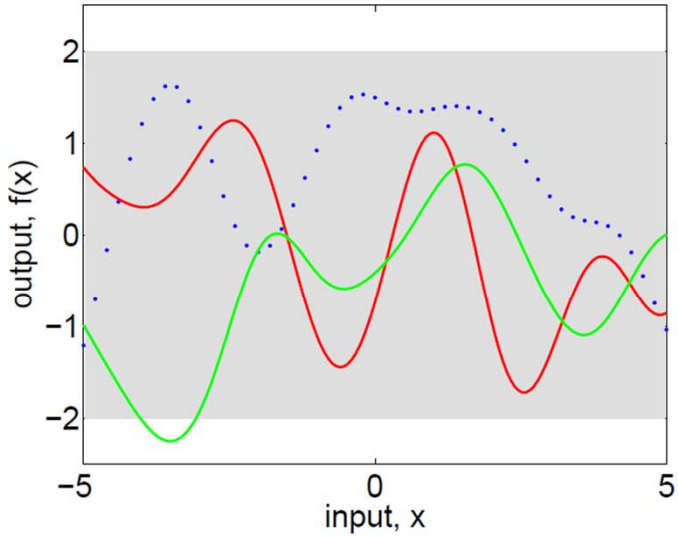
$$p(r_{nm}|\boldsymbol{u}_n, \boldsymbol{v}_m, \alpha) = \mathrm{Normal}(\boldsymbol{r}_n|f_m(\boldsymbol{u}_n), \alpha^{-1}\boldsymbol{I})$$

$$\boxed{r_{nm}} = f\left(\boxed{\boldsymbol{u}_n}\right)$$



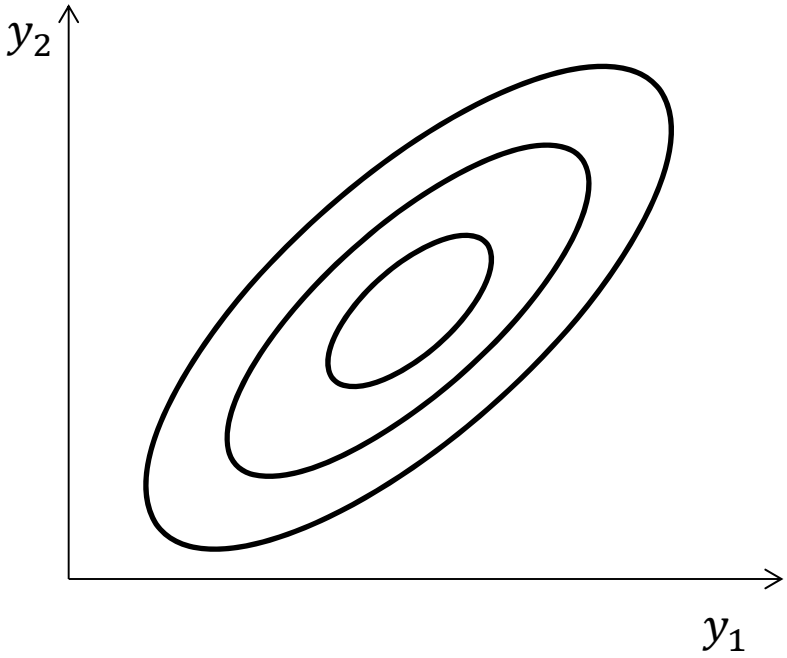
ガウス過程

- 無限次元の正規分布
- 関数の事前分布

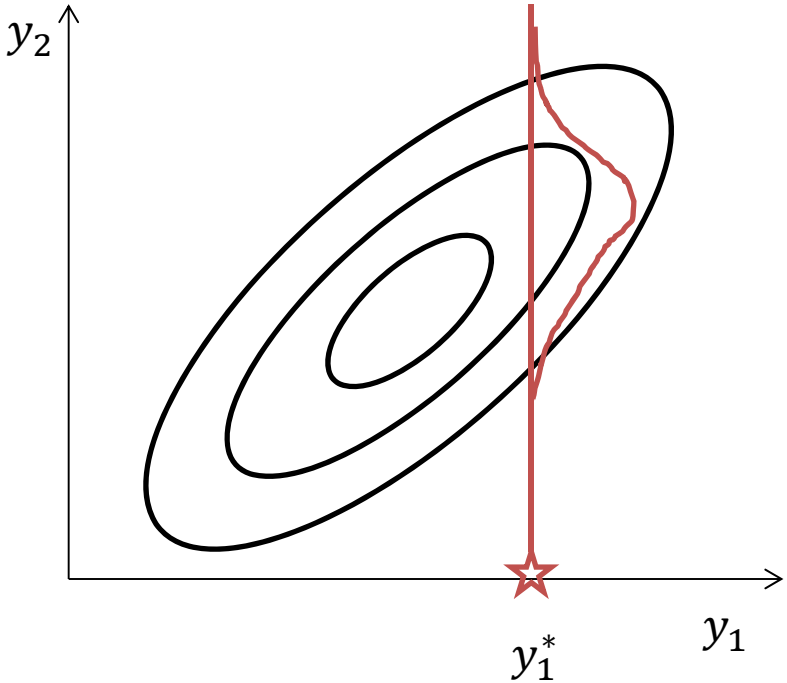


(a), prior

Gaussian Processes for Machine Learning
Carl Edward Rasmussen and Christopher K. I. Williams,MIT Press, 2006



D. Mackay, videolectures.net/gpip06_mackay_gpb



同時分布

$$p(x_1, x_2) = \mathrm{Normal}(\mu, \Sigma)$$

条件付分布

$$p(x_1|x_2) = \mathrm{Normal}(\bar{\mu}, \bar{\Sigma})$$

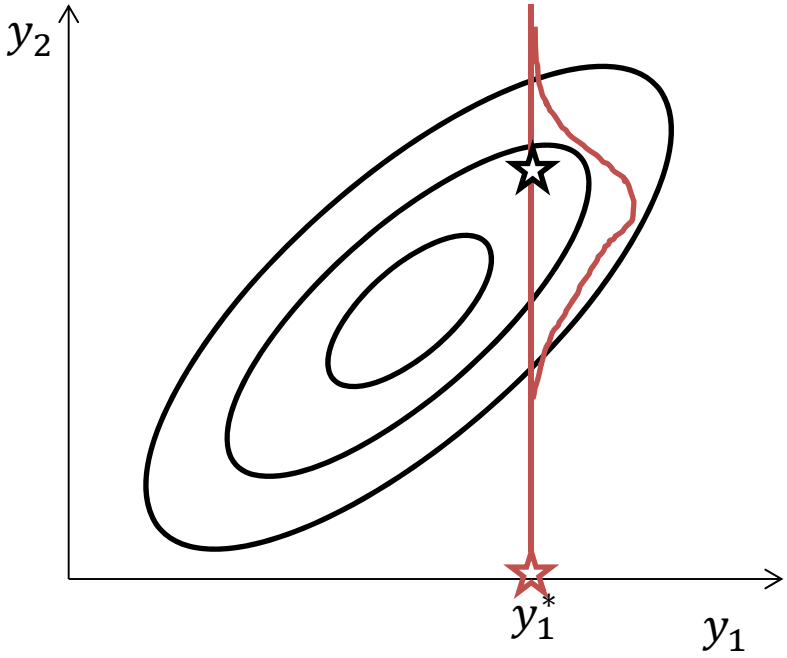
$$\bar{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

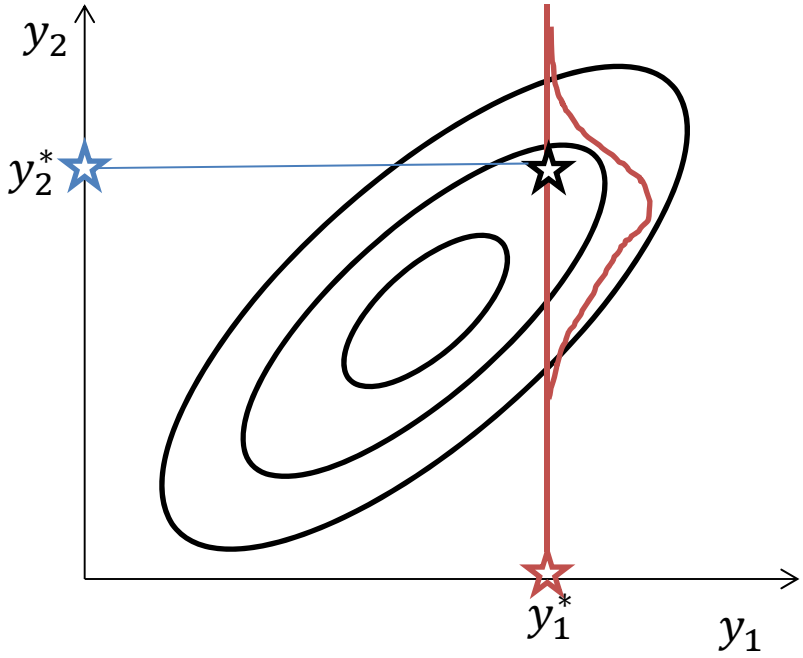
周辺分布

$$p(x_1) = \mathrm{Normal}(\mu_1, \Sigma_1)$$

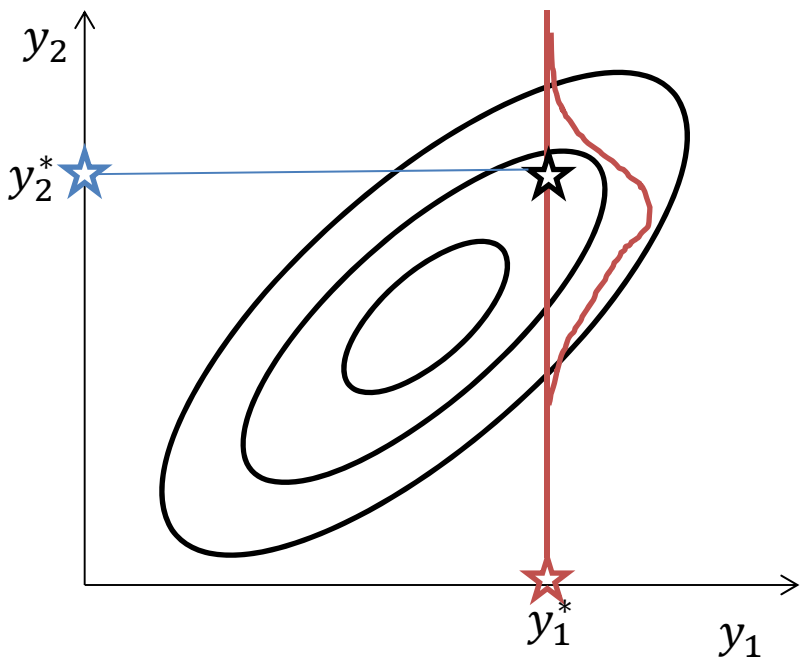
D. Mackay, videolectures.net/gpip06_mackay_gpb

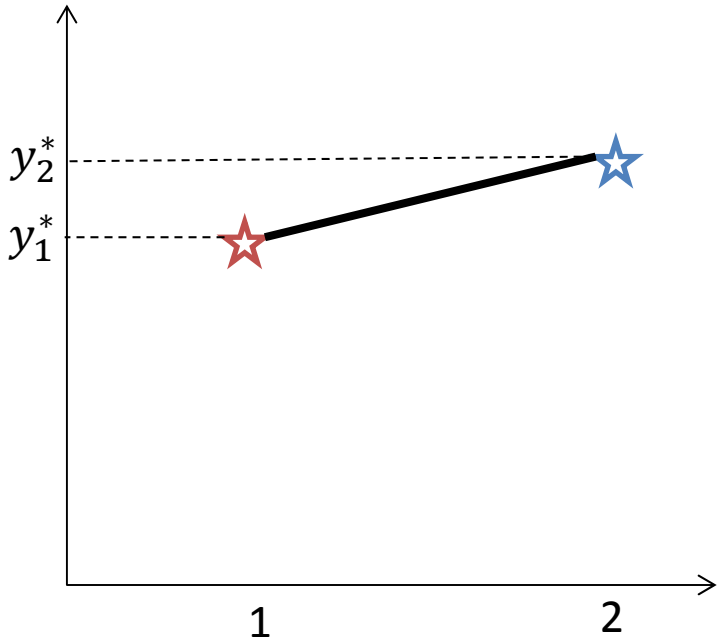


D. Mackay, videolectures.net/gpip06_mackay_gpb

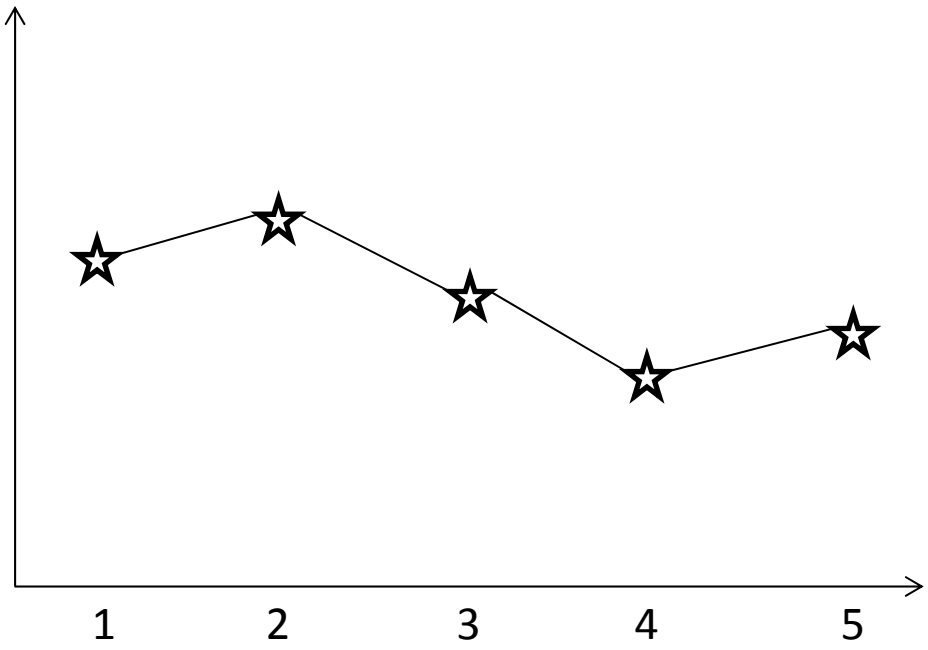


D. Mackay, videolectures.net/gpip06_mackay_gpb

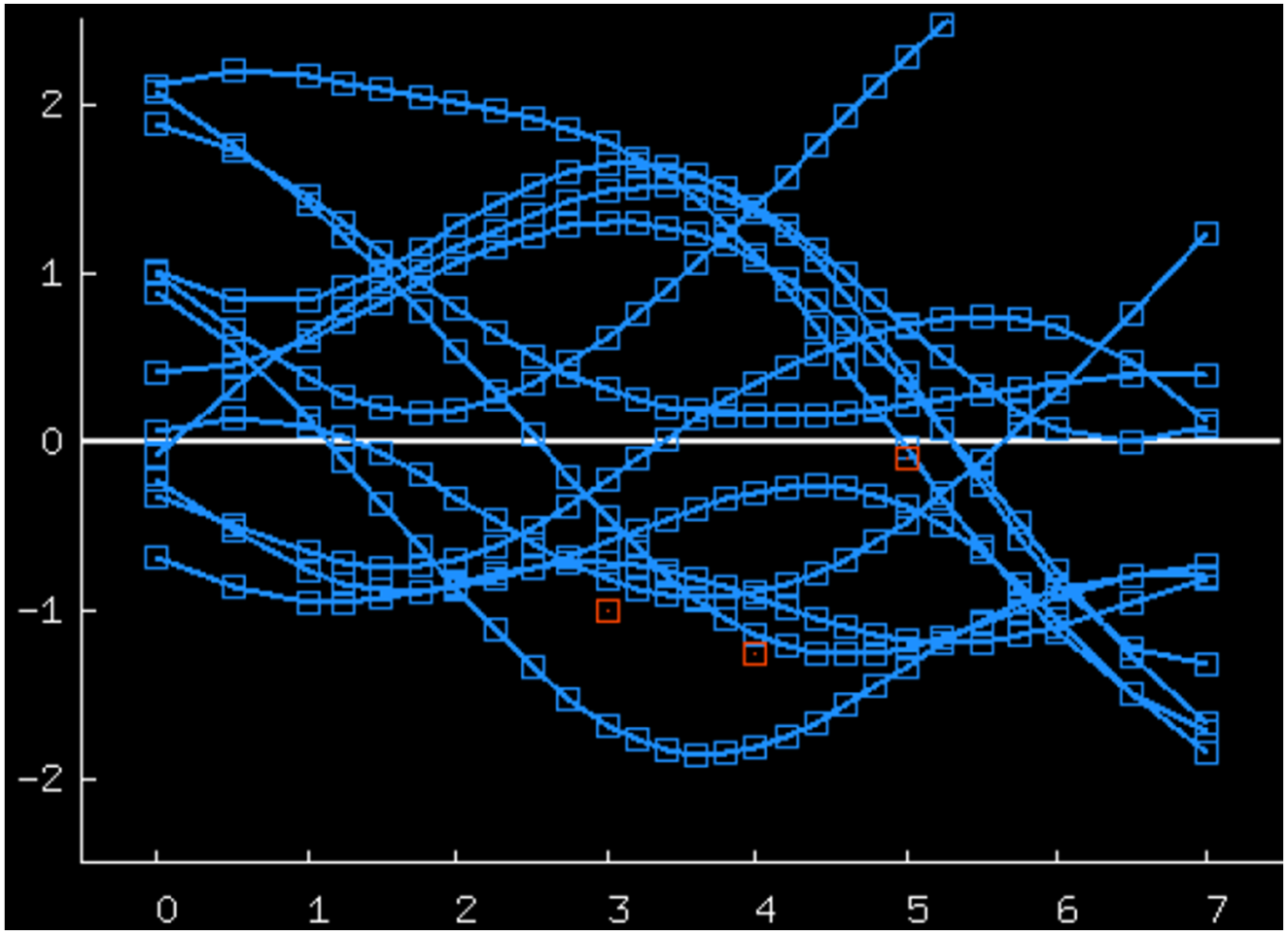




D. Mackay, videolectures.net/gpip06_mackay_gpb



D. Mackay, videolectures.net/gpip06_mackay_gpb



D. Mackay, videolectures.net/gpip06_mackay_gpb

ガウス過程

- 無限次元の正規分布
- 関数の事前分布

$$f(\boldsymbol{x}) \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$$

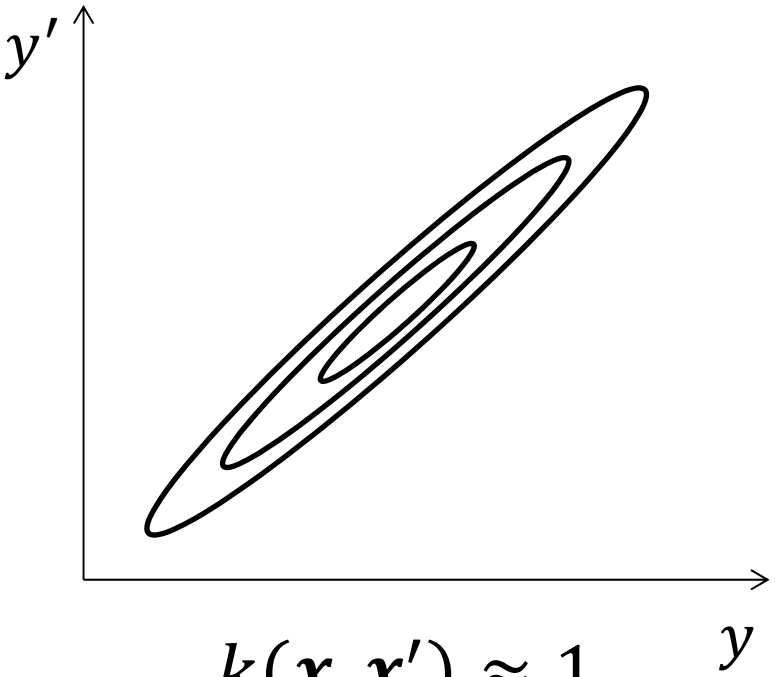
平均ベクトル（0ベクトルとすることが多い）

共分散関数（カーネル）

カーネルの例:ガウスカーネル

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\frac{1}{2\sigma^2} \parallel \boldsymbol{x} - \boldsymbol{x}' \parallel^2)$$

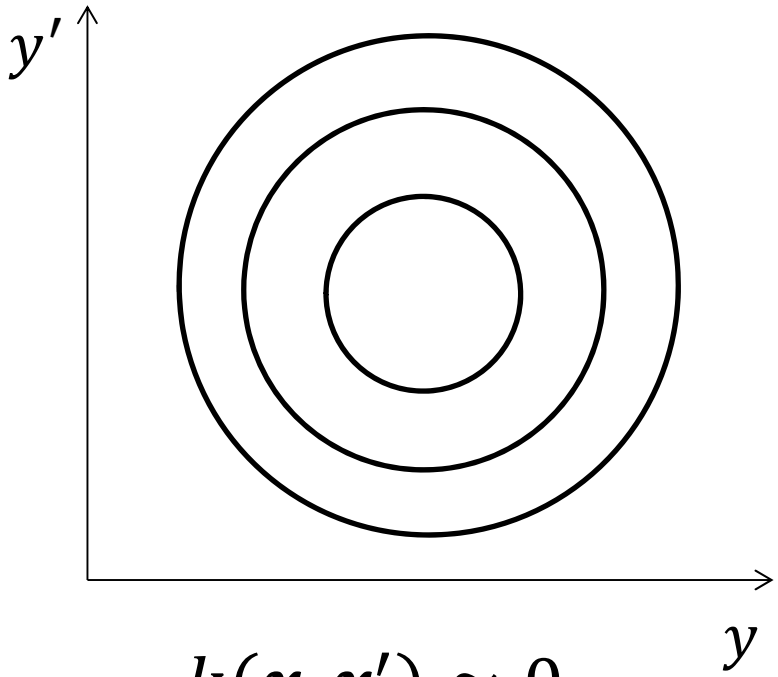
xとx'が近い場合



$k(\boldsymbol{x}, \boldsymbol{x}') \approx 1$

$\rightarrow \boldsymbol{y} \approx \boldsymbol{y}'$

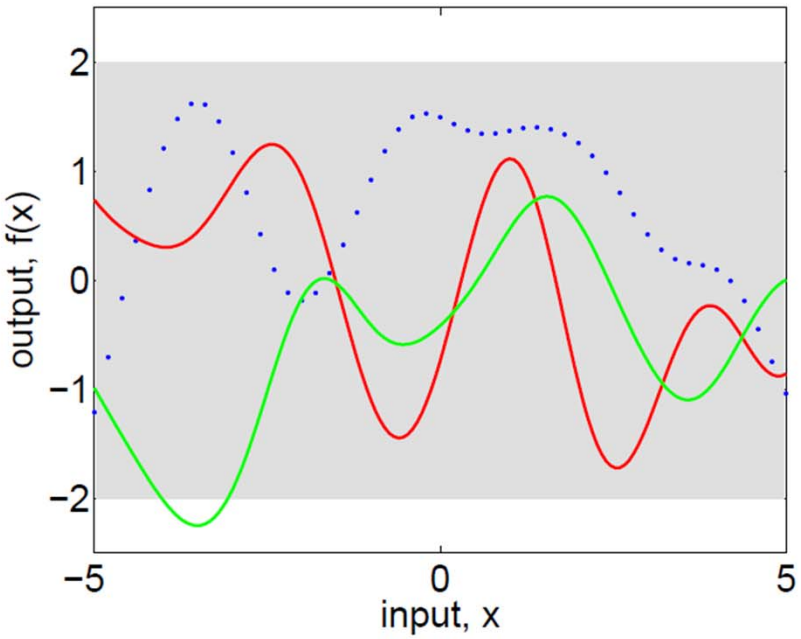
xとx'が遠い場合



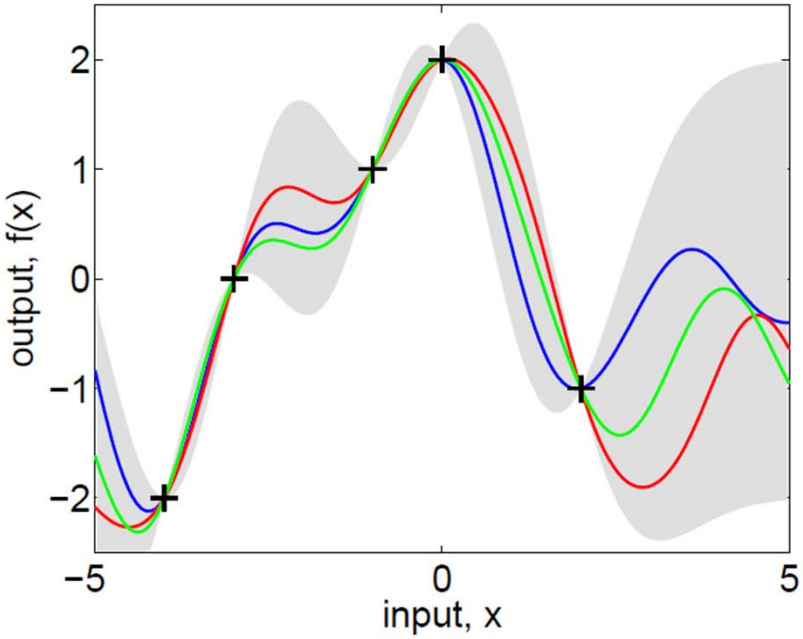
$k(\boldsymbol{x}, \boldsymbol{x}') \approx 0$

$\rightarrow \boldsymbol{y}$と$\boldsymbol{y}'$は相関しない

事前分布と事後分布



事前分布



事後分布

Gaussian Processes for Machine Learning
Carl Edward Rasmussen and Christopher K. I. Williams,MIT Press, 2006

非線形関数を積分消去

非線形関数fの事前分布がガウス過程，観測が正規分布に従うとき，fは積分消去できる

$$p(\boldsymbol{r}_m|\,\alpha) = \mathrm{Normal}(\boldsymbol{r}_m|\boldsymbol{0}, \boldsymbol{K} + \alpha^{-1}\boldsymbol{I})$$

カーネル行列
$\boldsymbol{K}_{nn'} = k(\boldsymbol{u}_n, \boldsymbol{u}_{n'})$

潜在ベクトルuが似ていれば，相関が高くなる
→評点が近くなる

類似度ベースの手法との関連

- 非線形行列分解は類似度に基づく手法の1種ともみなせる
  - 類似度はデータから学習
- 予測評点は他のユーザの重み付和

$$\hat{r}_{nm} = \sum_{n'=1}^{N} \underbrace{(\boldsymbol{K} + \alpha^{-1}\boldsymbol{I})^{-1}\boldsymbol{K}_{nn'}}_{\text{ユーザnとn'の類似度}} r_{n'm}$$

商品潜在ベクトルを積分消去

- 事前分布

$$p(\boldsymbol{V}|\boldsymbol{\lambda}) = \mathrm{Normal}(\boldsymbol{v}_n|\boldsymbol{0}, \lambda^{-1}\boldsymbol{I})$$

- 事前分布をかけてVを積分消去

$$p(\boldsymbol{R}|\boldsymbol{U}, \alpha, \lambda) = \int p(\boldsymbol{R}|\boldsymbol{U}, \boldsymbol{V}, \alpha) p(\boldsymbol{V}|\lambda) d\boldsymbol{V}$$

$$= \prod_{m=1}^{M} \mathrm{Normal}(\boldsymbol{r}_m|\boldsymbol{0}, \lambda^{-1}\boldsymbol{U}^\top\boldsymbol{U} + \alpha^{-1}\boldsymbol{I})$$

Uの共分散をカーネル行列$\boldsymbol{K}$に置き換えるとガウス過程行列分解になる

学習

対数尤度

$$\log \prod_{m=1}^{M} \mathrm{Normal}(\boldsymbol{r}_m | \mathbf{0}, \boldsymbol{K} + \alpha^{-1}\boldsymbol{I})$$

が最大になるようにユーザ潜在ベクトルUを学習

比較実験

- NMAE: normalized mean absolute error

URP	0.4341 ± 0.0023
Attitude	0.4320 ± 0.0055
MMMF	0.4156 ± 0.0037
Item	0.4096 ± 0.0029
E-MMMF	0.4029 ± 0.0027
Ours linear	0.4052 ± 0.0011
Ours RBF	$\mathbf{0.4026 \pm 0.0020}$

URP: user rating profile
MMMF: max margin matrix factorization
E-MMMF: ensembles of MMMF

時間情報を考慮

- パラメータを時間依存にする

$$\hat{r}_{nm} = \mu + b_n(t) + b_m(t) + u_n(t)^\top v_m$$

全体の評点平均

ユーザnの時刻tの評点平均

商品mの時刻tの評点平均

ユーザmの時刻tの潜在ベクトル

ユーザの評点基準が変化

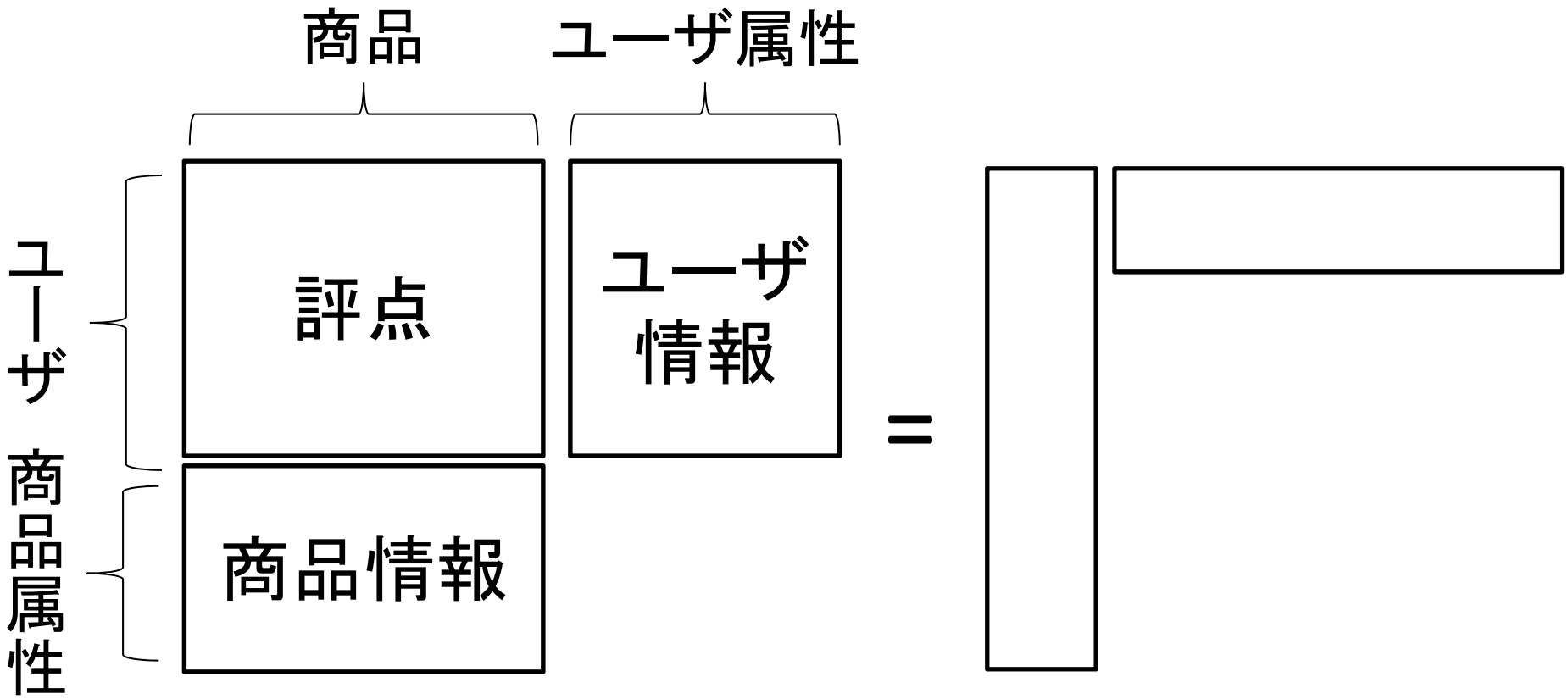
商品の人気が変化

ユーザの興味が変化

Collaborative filtering with temporal dynamics,
Yehuda Koren, KDD2009

他の情報も利用

- 商品情報，ユーザ情報なども利用する
- 複合行列分解（collective matrix factorization）

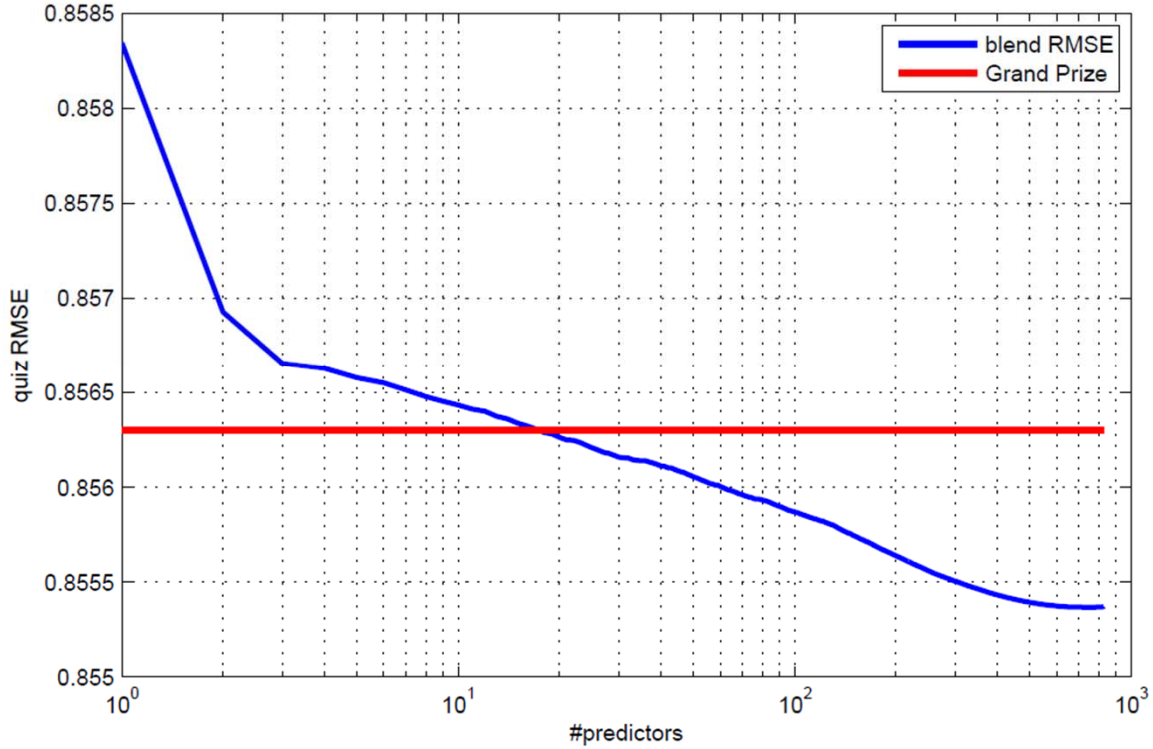


Relational learning via collective matrix factorization, AP Singh, GJ Gordon, KDD2008

複数の手法をまとめる

- Netflix prizeでは複数の異なる手法をまとめることにより精度向上

誤差



使用した手法数

The BigChaos Solution to the Netix Grand Prize,
Andreas Toscher , Michael Jahrer, Robert M. Bell, 2009

目的関数

- 評点誤差を下げるのが目的？
- 推薦しなくても買う商品を推薦しても仕方がない
- 推薦しなかったら買わなかった商品を推薦したい
- 推薦したときに他の商品も波及的に買うようになる商品を推薦したい

顧客生涯価値を高めるための推薦

Tomoharu Iwata, Kazumi Saito, Takeshi Yamada,
"Recommendation Method for Improving Customer Lifetime Value,"
IEEE Transactions on Knowledge and Data Engineering, Vol.20, No.9, 1254-1263, 2008

既存のリコメンデーションの問題点

- 興味が似ている人が買っている商品を推薦（短期的）

- すぐやめてしまうなど、非優良顧客の購入商品を推薦する可能性もある
  ⇒長期的な収益が上がるとは限らない

長期的利益を高める推薦

- 長期的利益を顧客生涯価値で定量化
  - あるユーザから将来にわたって得られる利益の現在価値（購買数）

- 推薦する商品の例
  - 優良顧客が多く購入している商品（プロの利用者を見習う）
  - 顧客の満足度を高める商品
  - 将来の購入確率を高める商品（周辺機器の多い機械など）

アプローチ

1

Cox比例ハザードモデル
（生存時間解析）を用い
顧客生涯価値の高いユーザに
特徴的な購買パターンを抽出

顧客生涯価値の高いユーザ
の購買履歴

スターウォーズ
ハリーポッター
バットマン
⋮

スーパーマン
ホームアローン
トイストーリー
⋮

リング
スターウォーズ
ハリーポッター
⋮

スターウォーズ → ハリーポッター
ホームアローン → トイストーリー
バットマン → スターウォーズ
⋮

最大エントロピーモデル
を用いユーザの興味を推定

興味：アニメ

トイストーリー
を推薦

興味：SF

スターウォーズ
を推薦

2

興味を考慮しつつ
顧客生涯価値
の高いユーザ
と同じような購買パターン
になるように推薦

3

定式化

推薦されたときLTV（LifeTime Value; 顧客生涯価値）が上昇する確率が最も高い商品を提示

商品mをユーザnに推薦したときにLTVが上昇する確率

$$\hat{m} = \underset{m \in M}{\operatorname{argmax}}\, P(\ell | n, m)$$

推薦商品

商品集合

LTV上昇

ユーザ

商品

条件付独立を仮定して分解

商品mを推薦したときに
LTVが上昇する確率

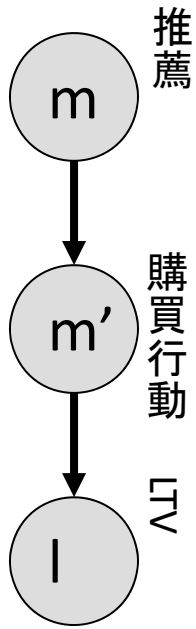
商品m'を購入

$$p(\ell|n,m) = \sum_{m' \in \boldsymbol{M}} P(\ell, m'|n,m)$$

$$= \sum_{m' \in \boldsymbol{M}} Q(\ell|m',n,m)\, R(m'|n,m)$$

$$= \sum_{m' \in \boldsymbol{M}} Q(\ell|m',n,)\, R(m'|n,m)$$

条件付独立



商品m'を購入したときに
LTVが上昇する確率

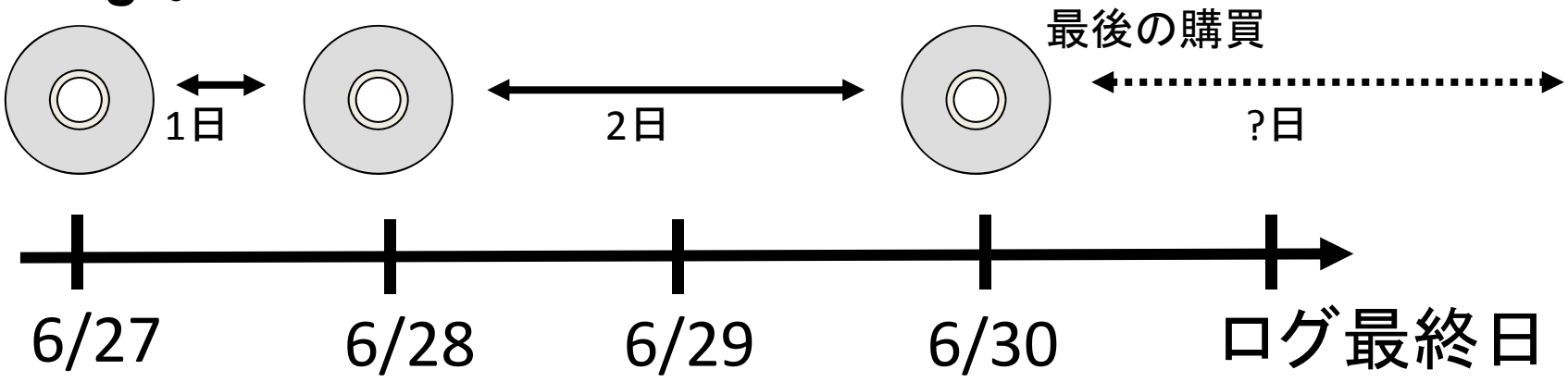
1 生存時間解析を用いて推定

商品mを推薦したときに
商品m'を購入する確率

2 最大エントロピーモデルで推定

生存時間解析

- あるイベントが起こるまでの時間を解析
  - 死ぬまでの時間
  - 機械が故障するまでの時間
  - 次の購買までの時間（＝購買間隔）
- 打切りデータも扱える
  - 最後の購買の場合、真の購買間隔は分からない



Cox比例ハザードモデル

- ハザード関数：購買間隔がt以上のユーザのうち購買間隔がtのユーザの割合
- ハザード関数の対数が購買パターン***x***に比例
  - λが高い特徴量は購買間隔が短い（＝LTVが高い）ユーザに特徴的な購買パターン

ベースラインハザード

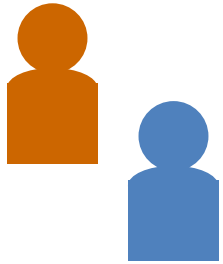
購買パターンの特徴量

$$h(t \mid \mathbf{x}) = h_0(t)\exp(\boldsymbol{\lambda}^T \mathbf{x})$$

- 部分尤度最大化で**λ**を推定
  - ベースラインハザードを推定する必要がない
  - 大域的最適解が求まる

商品m'が購入されたときの購買頻度上昇確率

n: 商品m'を購入しなかったユーザ
$n_{+m'}$: 商品m'を購入したユーザ



$Q(\ell|n,m')$
=先に購入するのが$n_{+m'}$である確率

$$=\frac{h(t|x_{+m'},n_{+m'})}{h(t|x_{+m'},n_{+m'})+h(t|x,n)}$$

$$=\frac{1}{1+\exp(-\lambda^{\top}(x_{+m'}-x))}$$

商品m'を購入した
ユーザのハザード関数

商品m'を購入しなかった
ユーザのハザード関数

商品mを推薦したときの購買確率

- 最大エントロピーモデル
- 未知パラメータα_cはログデータから学習

購入商品

推薦商品

推薦有無の特徴量

$$\mathrm{I}(m, m') = \begin{cases} 1 & m = m \\ 0 & m \neq m' \end{cases}$$

履歴（興味）の特徴量

$$R(m'|n, m)$$

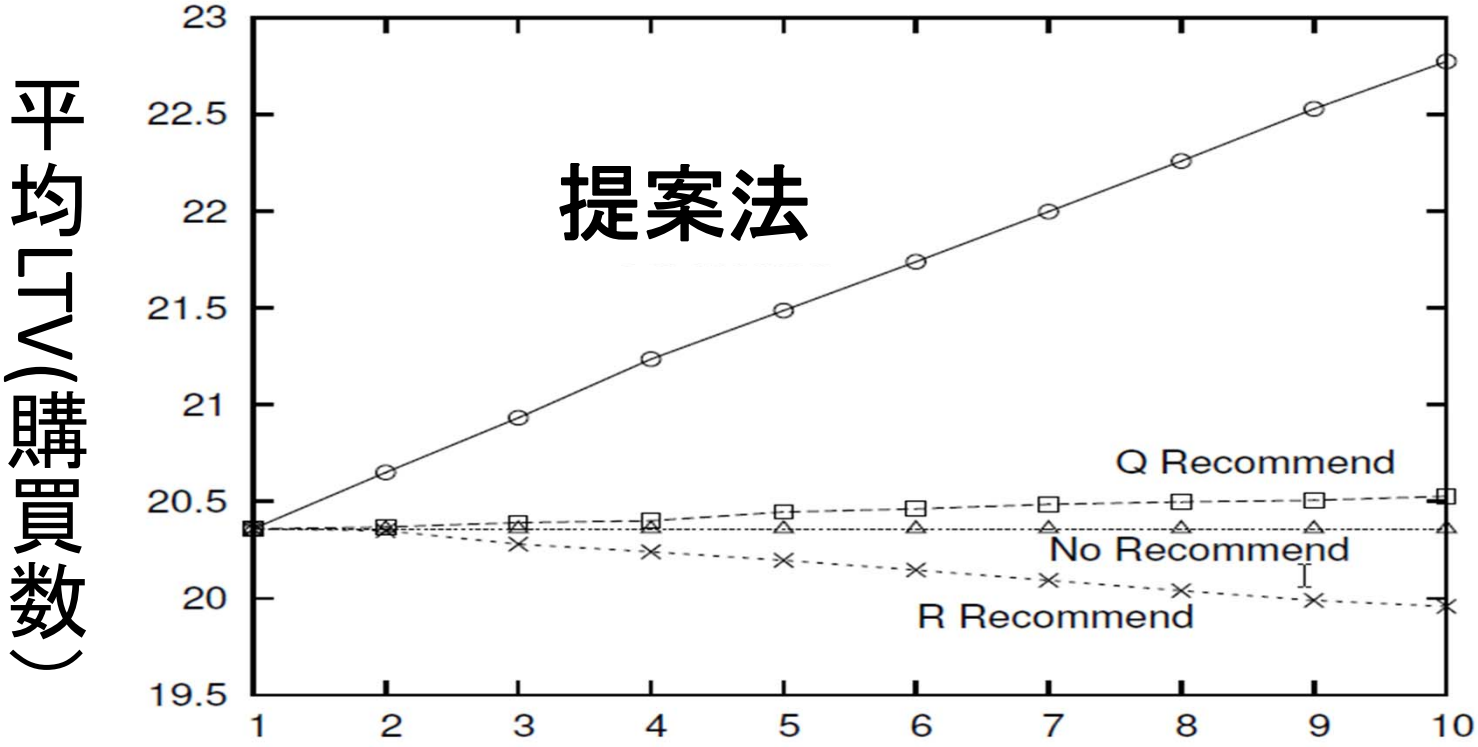
$$= \frac{1}{Z}\exp(\beta I(m, m') + \sum_c \alpha_c y_c(n, m'))$$

$$= \begin{cases} \frac{\exp(\beta)}{Z} R(m'|n) & m = m' \text{（推薦商品を購入）} \\ \frac{1}{Z} R(m'|n) & m \neq m' \text{（推薦商品を購入しない）} \end{cases}$$

推薦の影響度γ

推薦なしでの購買確率

シミュレーションによる試算



- Q Recommend:LTVのみ考慮
  購買確率が極めて低い商品も推薦してしまう
- R Recommend :購買確率のみ考慮
  LTVを高める推薦ではない

推薦リストの多様化

- 複数の商品を同時に推薦する場合，似たような商品を推薦されてもうれしくない
- 多様性が高い商品集合を推薦する
- ユーザ満足度が向上

Improving Recommendation Lists Through Topic Diversification, Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, Georg Lausen, WWW2005

まとめ

- 推薦
  - ユーザの興味に合わせておすすめのアイテムを提示
- 機械学習
  - 目的を達成するためにモデルを学習
- 行列分解
  - 評点，言語，ネットワーク，音楽
- ベイズ推定
  - パラメータを分布で推定; データ数が少ない場合に効果
- 非線形拡張
  - カーネル法を用いて表現能力を上げる

参考文献1

- 類似度ベース:
  - Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, Item-based collaborative filtering recommendation algorithms, WWW2001
- 行列分解:
  - Matrix Factorization Techniques for Recommender Systems, Yehuda Koren, Robert Bell and Chris Volinsky,, IEEE Computer 2009
- ベイズ行列分解:
  - Bayesian Probabilistic Matrix Factorization using MCMC. Ruslan Salakhutdinov and Andriy Mnih, ICML2008
- 非線形行列分解:
  - Non-linear matrix factorization with Gaussian processes, Neil Lawrence, and Rquel Urtasun, ICML 2009
- 顧客生涯価値:
  - Tomoharu Iwata, Kazumi Saito, Takeshi Yamada, "Recommendation Method for Improving Customer Lifetime Value,"IEEE Transactions on Knowledge and Data Engineering, Vol.20, No.9, 1254-1263, 2008

参考文献2

- 多様化:
  - Improving Recommendation Lists Through Topic Diversification, Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, Georg Lausen, WWW2005
- 手法の統合:
  - The BigChaos Solution to the Netix Grand Prize, Andreas Toscher , Michael Jahrer, Robert M. Bell, 2009
- 時間情報:
  - Collaborative filtering with temporal dynamics, Yehuda Koren, KDD2009
- 他の情報も利用する
  - Relational learning via collective matrix factorization, AP Singh, GJ Gordon, KDD2008
- サーベイ
  - 神嶌 敏弘 “推薦システムのアルゴリズム (1)-(3)” 人工知能学会誌, 2007-2008