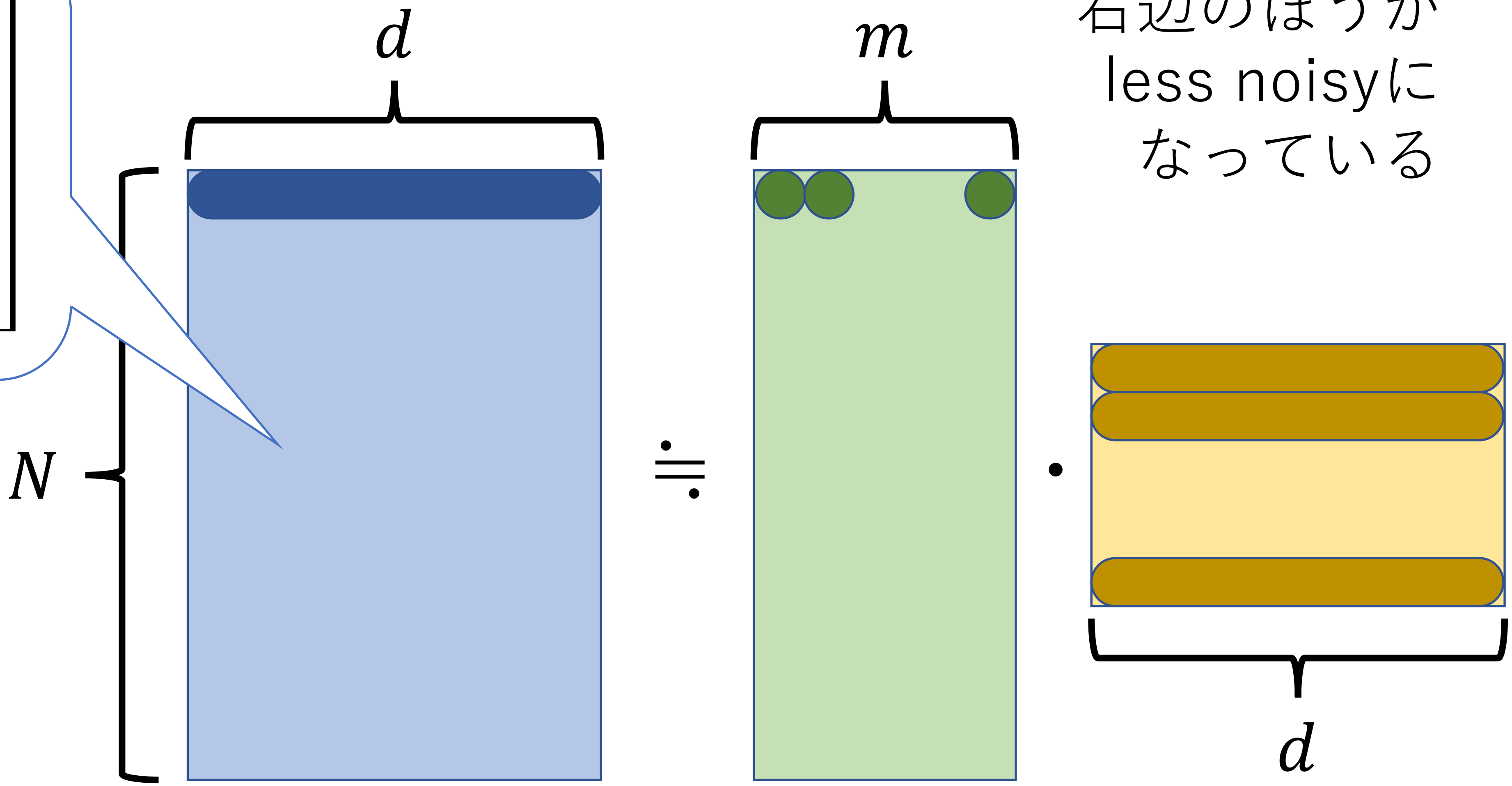
dimensionality reduction

正田 備也

masada@rikkyo.ac.jp

PCAで計画行列をless noisyにする

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,1} & x_{i,2} & \cdots & x_{i,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,d} \end{bmatrix}$$



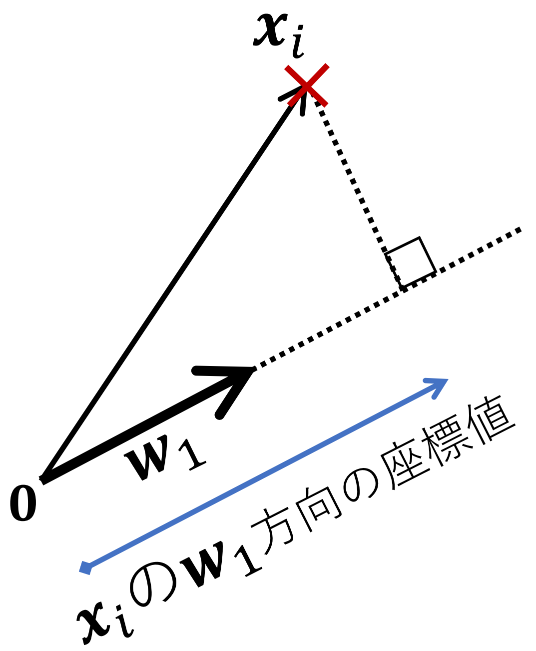
右辺のほうが
less noisyに
なっている

左辺は
元々の
計画行列

PCAの考え方

- 高次元空間に散らばったデータ点の集合が与えられている
    - データ点はベクトルです（原点から発してその点まで達する矢印のイメージ）
  - データの重心が原点になるように、全体の位置をシフトさせておく。
- データ点が最も広く散らばっている方向を見つける
  - この方向を表すベクトルが、主成分
- その方向と直交する超平面上に、全てのデータ点を押しつぶす
  - データ点の散らばっている空間の次元が、ひとつだけ下がる
- 次元を下げた空間で、また同じことをする
  - つまり、データ点が最も広く散らばっている方向を見つける

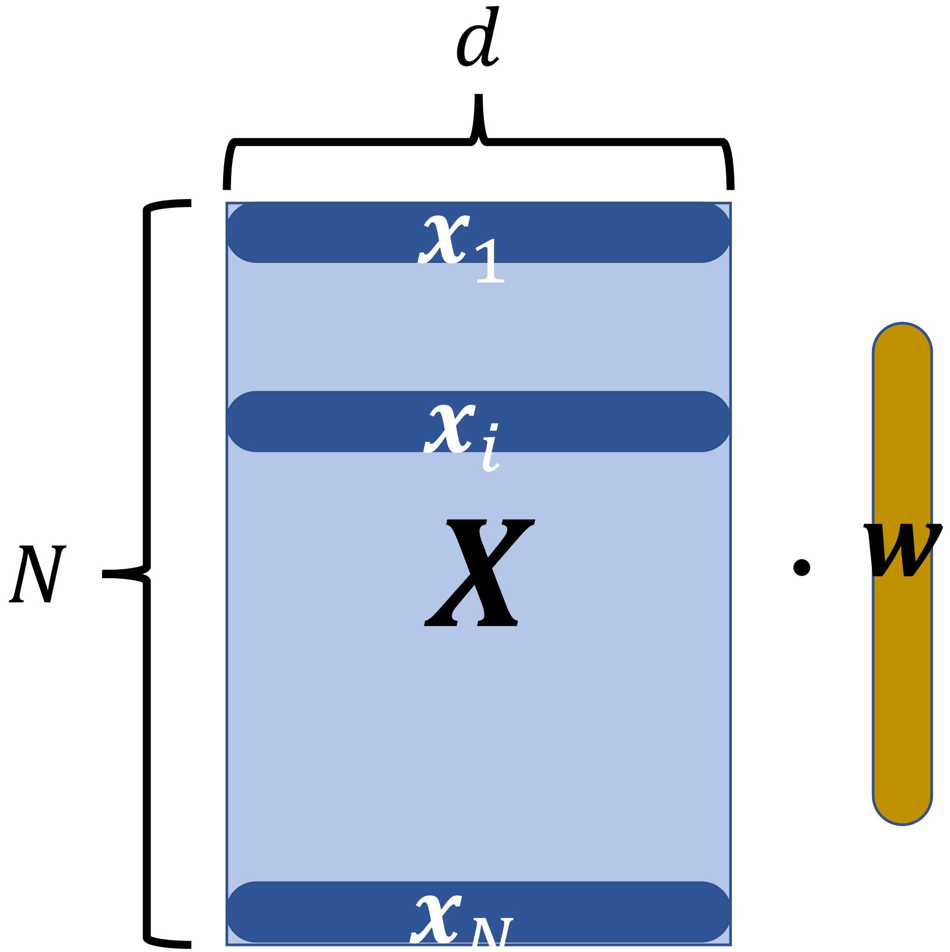
「最も広く散らばっている方向」



- 式で書くと以下の通り

$$\boldsymbol{w}_1 = \underset{\|\boldsymbol{w}\|=1}{\arg\max} \|\boldsymbol{X}\boldsymbol{w}\|^2$$

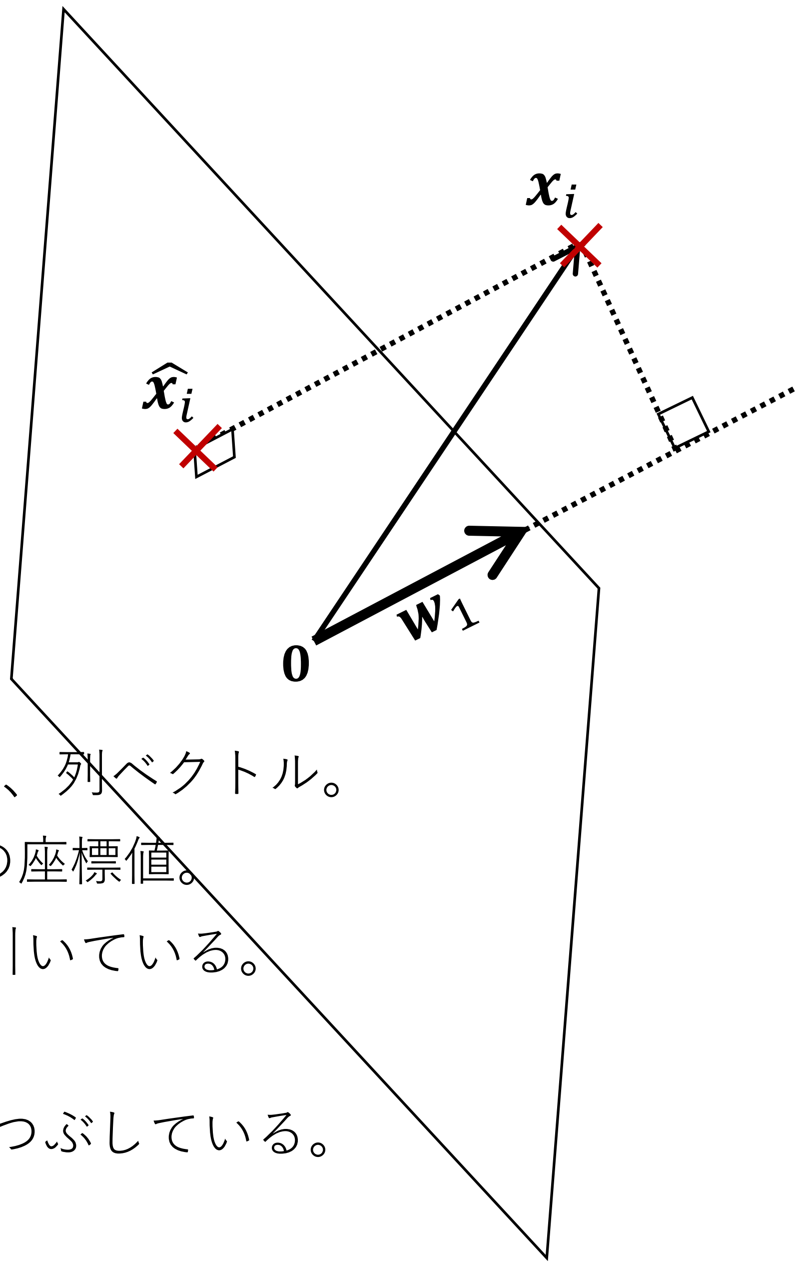
- $\boldsymbol{X}$は計画行列（ただし重心を原点へ移動した後のもの）
- $\boldsymbol{X}\boldsymbol{w}$は列ベクトル
  - $\boldsymbol{X}$の各行ベクトル$\boldsymbol{x}_i$と$\boldsymbol{w}$との内積の値が要素として並ぶ、列ベクトル。
  - この内積の値は、 $\boldsymbol{X}$の各行ベクトル$\boldsymbol{x}_i$の、$\boldsymbol{w}$の方向への座標値。
    - つまり、 $\boldsymbol{w}$の方向への＋か－の符号がついた長さ



$\boldsymbol{Xw}$のイメージ

各データ点$\boldsymbol{x}_i$と$\boldsymbol{w}$との内積$\boldsymbol{x}_i{}^T\boldsymbol{w}$を、まとめて書いているだけ。

「データ点を押しつぶす」



- 式で書くと以下の通り

$$\hat{\boldsymbol{X}} = \boldsymbol{X} - (\boldsymbol{X}\boldsymbol{w}_1)\boldsymbol{w}_1^T$$

- $\boldsymbol{X}\boldsymbol{w}$は列ベクトル
  - $\boldsymbol{X}$の各行ベクトル$\boldsymbol{x}_i$と$\boldsymbol{w}$との内積の値が要素として並ぶ、列ベクトル。
  - この内積の値は、$\boldsymbol{X}$の各行ベクトル$\boldsymbol{x}_i$の、$\boldsymbol{w}$の方向への座標値。
  - その座標値に$\boldsymbol{w}_1$をかけて、元の$\boldsymbol{X}$の行ベクトル$\boldsymbol{x}_i$から引いている。
  - こうすると、$\boldsymbol{x}_i$から$\boldsymbol{w}_1$方向の成分を消せる。
  - つまり、$\boldsymbol{x}_i$を、$\boldsymbol{w}_1$と直交し原点を通る超平面上へ押しつぶしている。

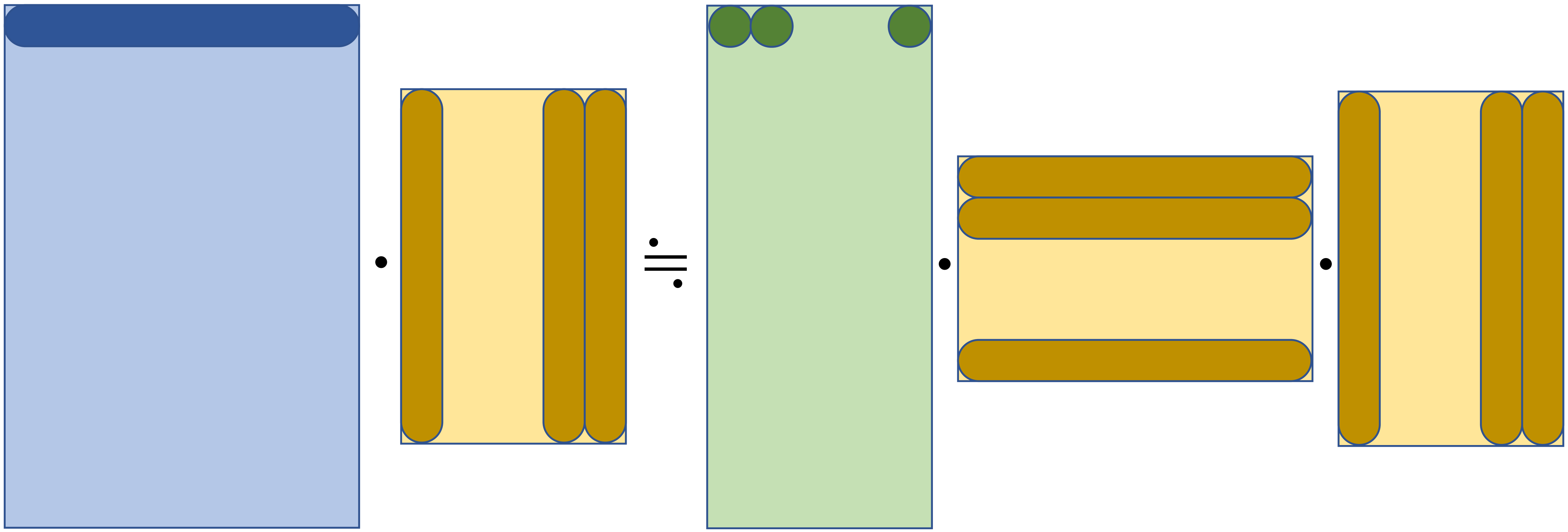
PCAの（実際の）アルゴリズム

1. データを中心化する（平均を引く）
2. 共分散行列$\boldsymbol{C} = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{X}$を計算する
3. 共分散行列$\boldsymbol{C}$の固有値と固有ベクトルを求める
4. 最も大きいm個の固有値に対応する固有ベクトルを選ぶ

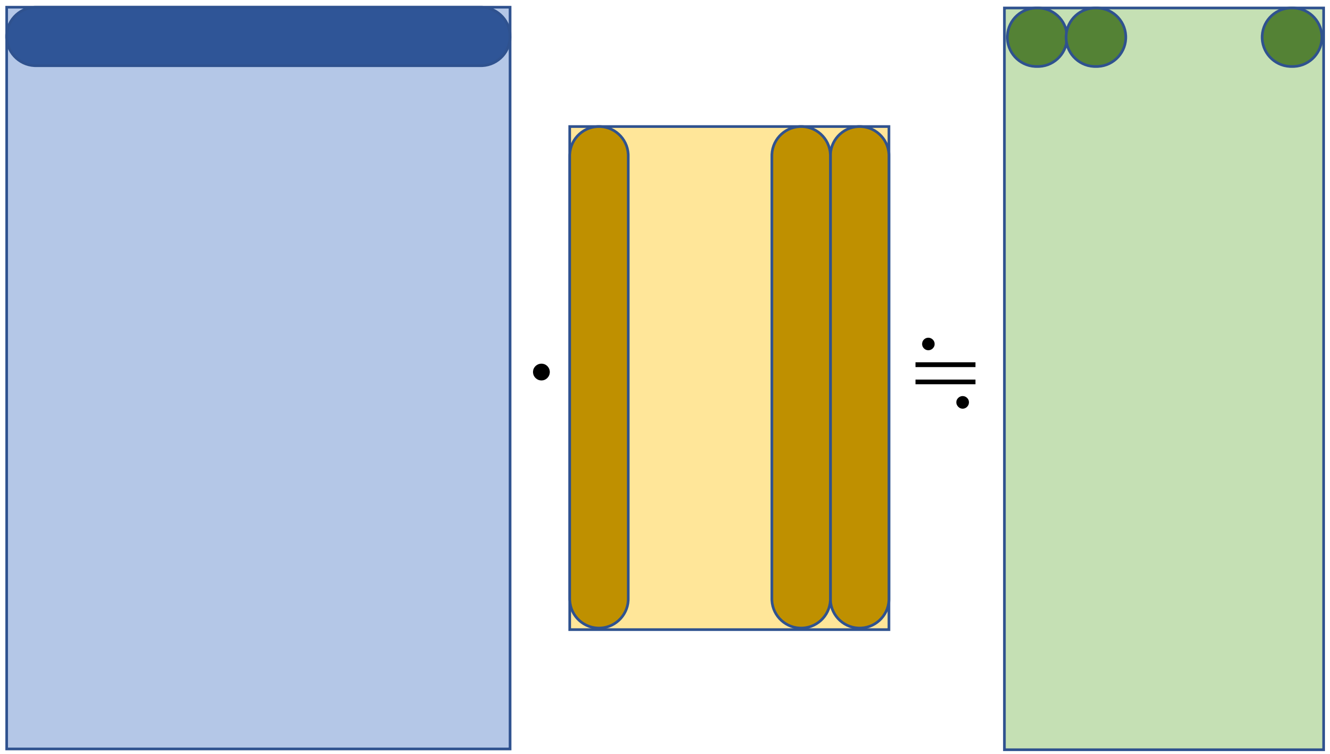
- このアルゴリズムで求まる固有ベクトルと、「最も広く散らばった方向」が同じになることは、ちゃんと証明できる
  - 等式制約の場合のラグランジュ未定乗数法を使う。

https://dthul.de/misc/pca.html

PCAによる次元圧縮(1/2)

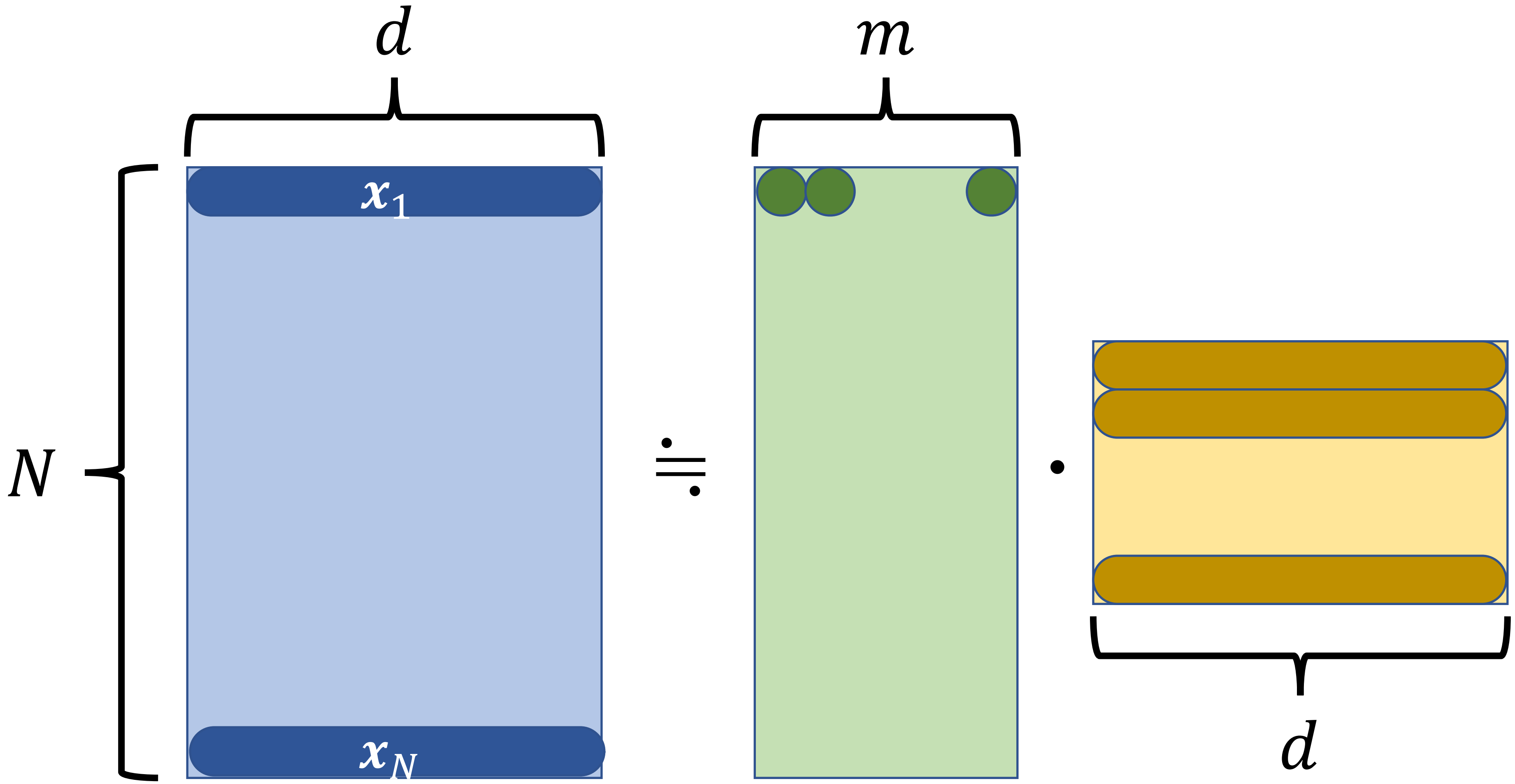


PCAによる次元圧縮(1/2)



PCA, NMF, PLSIなど

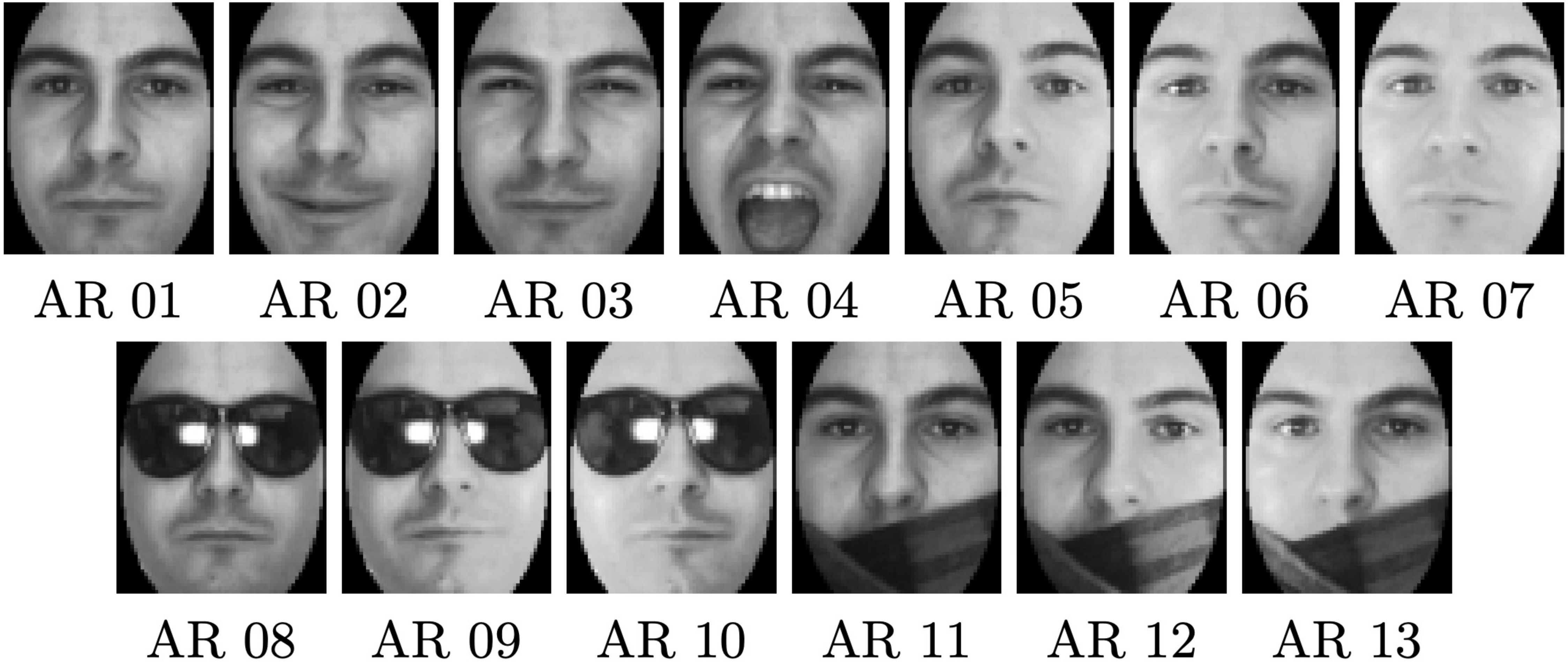
- 計画行列を、小さい2つの行列の積で表現し直す



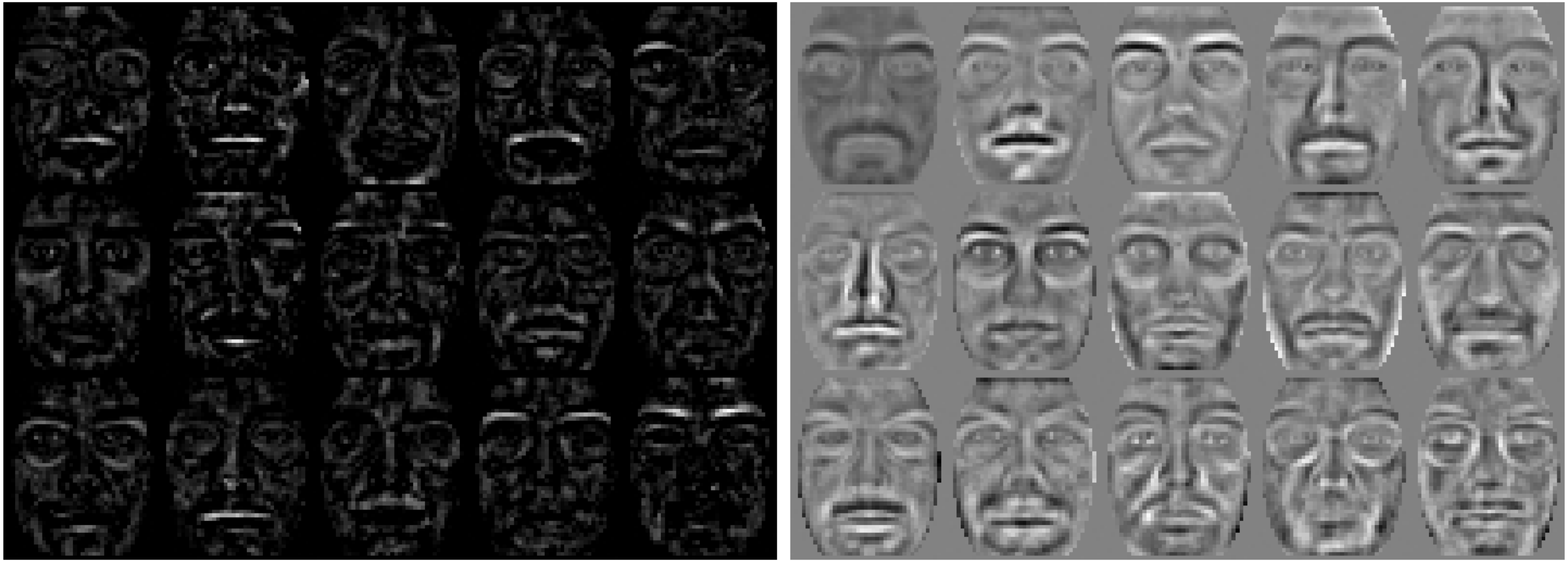
NMF (Nonnegative Matrix Factorization)

- PCAの問題点
  - 元々どの属性値も負の値を取り得ないデータには適用しにくい
  - 主成分は負の値を含みうるが、このような負の値は解釈が難しい

- NMFの特徴
  - PCAでいう主成分に相当するベクトルが非負ベクトルとして得られる
  - 得られる非負ベクトルは、お互いに直交しないこともありうる
    - ただし、疎なベクトルになるので、ほぼ直交しているともみなせる。



http://statweb.stanford.edu/~tibs/sta306bfiles/nnmf-faces.pdf



(a) NMF bases.   (b) PCA bases.

Fig. 2. Bases obtained by both techniques, PCA and NMF

http://statweb.stanford.edu/~tibs/sta306bfiles/nnmf-faces.pdf

PLSI (probabilistic latent semantic indexing)

- NMFの確率モデル版
- 詳細は秋学期の「統計モデリング１」で
- EMアルゴリズムでモデル・パラメータを推定する

LDA (latent Dirichlet allocation)

- PLSIをベイズ化した確率モデル
- 詳細は来年度夏学期の「統計モデリング２」で
- 変分推論(variational inference)で事後分布のパラメータを推定する

可視化のための次元削減

- UMAPとt-SNEに関するnotebookへ