

# dimensionality reduction

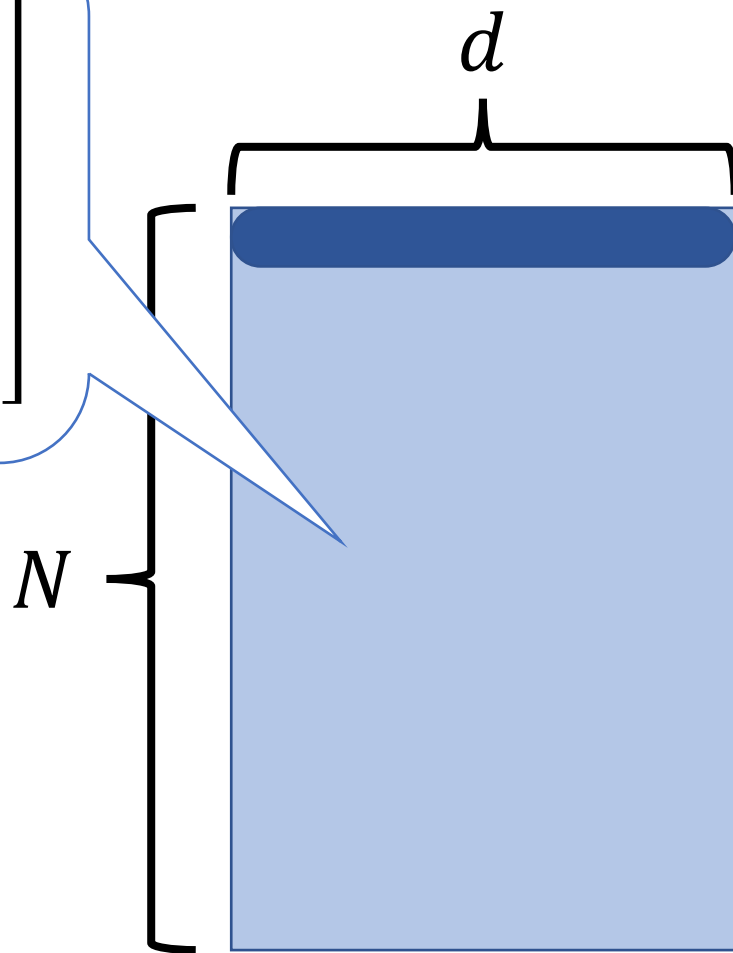
正田 備也

[masada@rikkyo.ac.jp](mailto:masada@rikkyo.ac.jp)

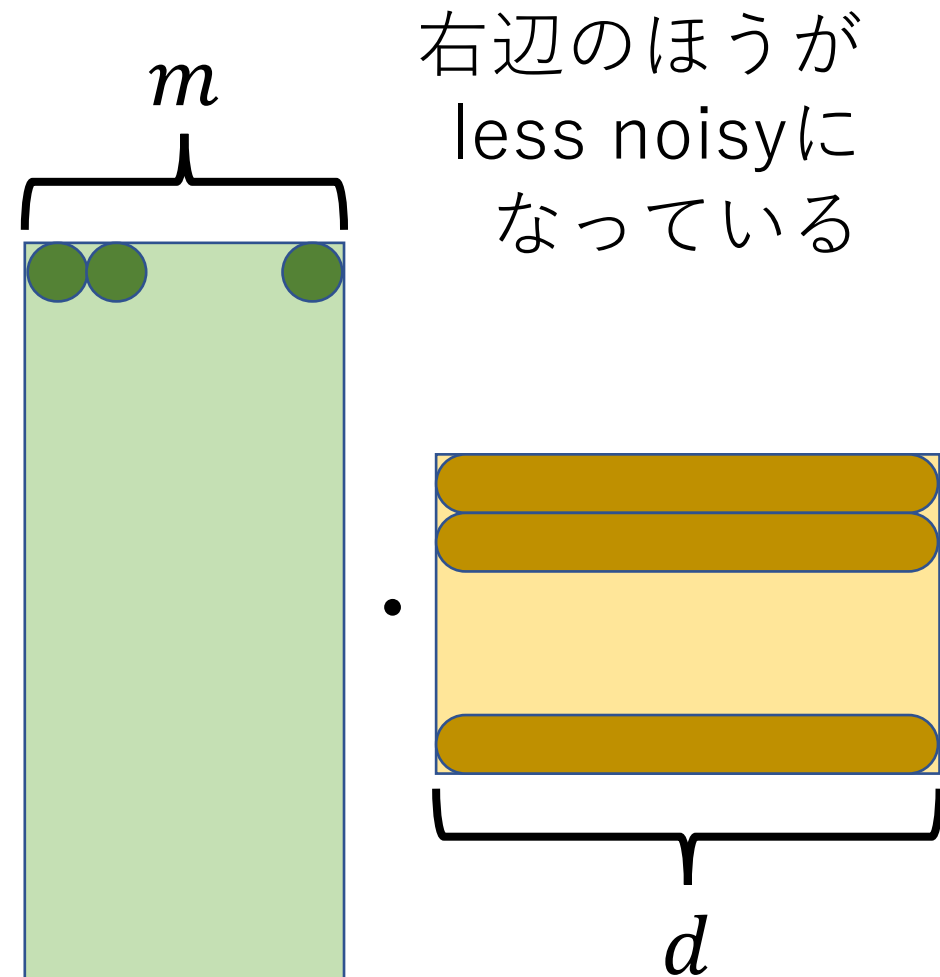
# PCAで計画行列をless noisyにする

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

左辺は  
元々の  
計画行列



$\hat{=}$



右辺のほうが  
less noisyに  
なっている

# PCAの考え方

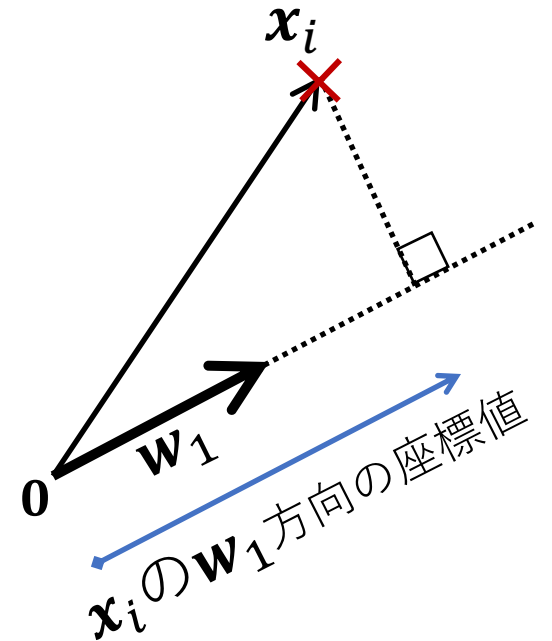
- 高次元空間に散らばったデータ点の集合が与えられている
  - データ点はベクトルです（原点から発してその点まで達する矢印のイメージ）
  - データの重心が原点になるように、全体の位置をシフトさせておく。
- データ点が最も広く散らばっている方向を見つける
  - この方向を表すベクトルが、主成分
- その方向と直交する超平面上に、全てのデータ点を押しつぶす
  - データ点の散らばっている空間の次元が、ひとつだけ下がる
- 次元を下げた空間で、また同じことをする
  - つまり、データ点が最も広く散らばっている方向を見つける

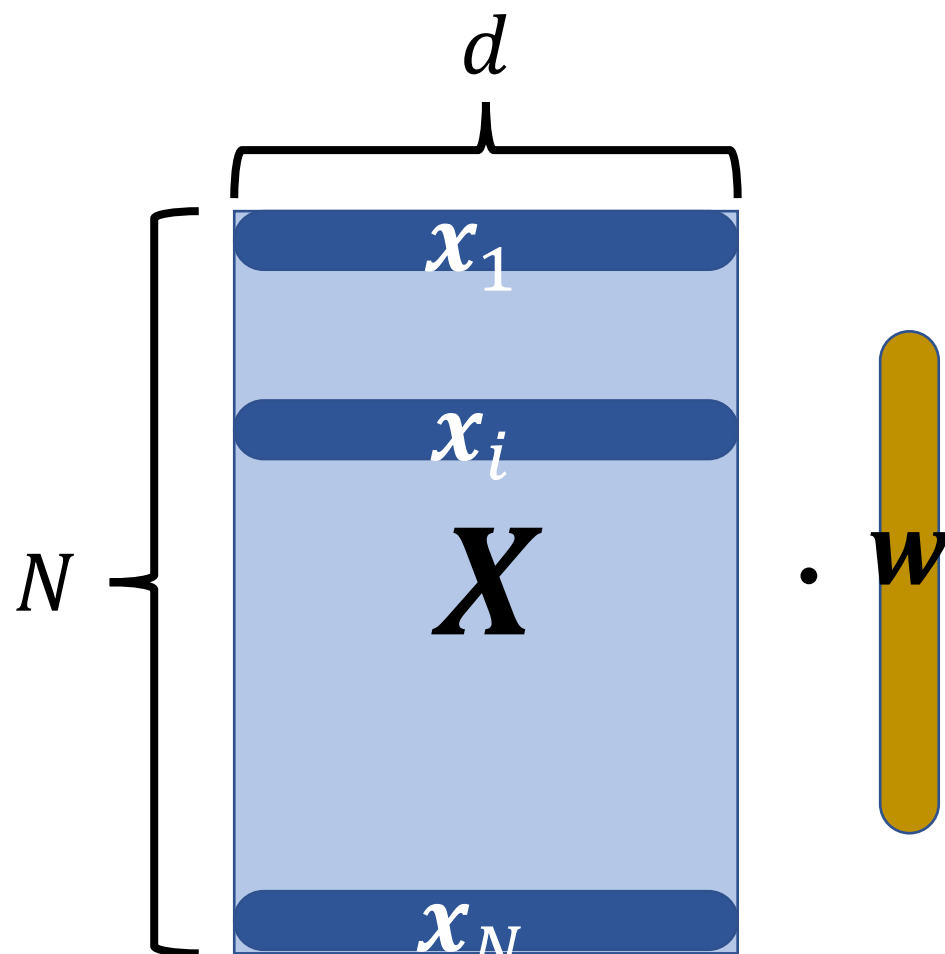
# 「最も広く散らばっている方向」

- 式で書くと以下の通り

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} \|\mathbf{X}\mathbf{w}\|^2$$

- $\mathbf{X}$ は計画行列（ただし重心を原点へ移動した後のもの）
- $\mathbf{X}\mathbf{w}$ は列ベクトル
  - $\mathbf{X}$ の各行ベクトル $\mathbf{x}_i$ と $\mathbf{w}$ 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、 $\mathbf{X}$ の各行ベクトル $\mathbf{x}_i$ の、 $\mathbf{w}$ の方向への座標値。
    - つまり、 $\mathbf{w}$ の方向への+か-の符号がついた長さ





## $Xw$ のイメージ

各データ点 $x_i$ と $w$ との内積 $x_i^T w$ を、まとめて書いているだけ。

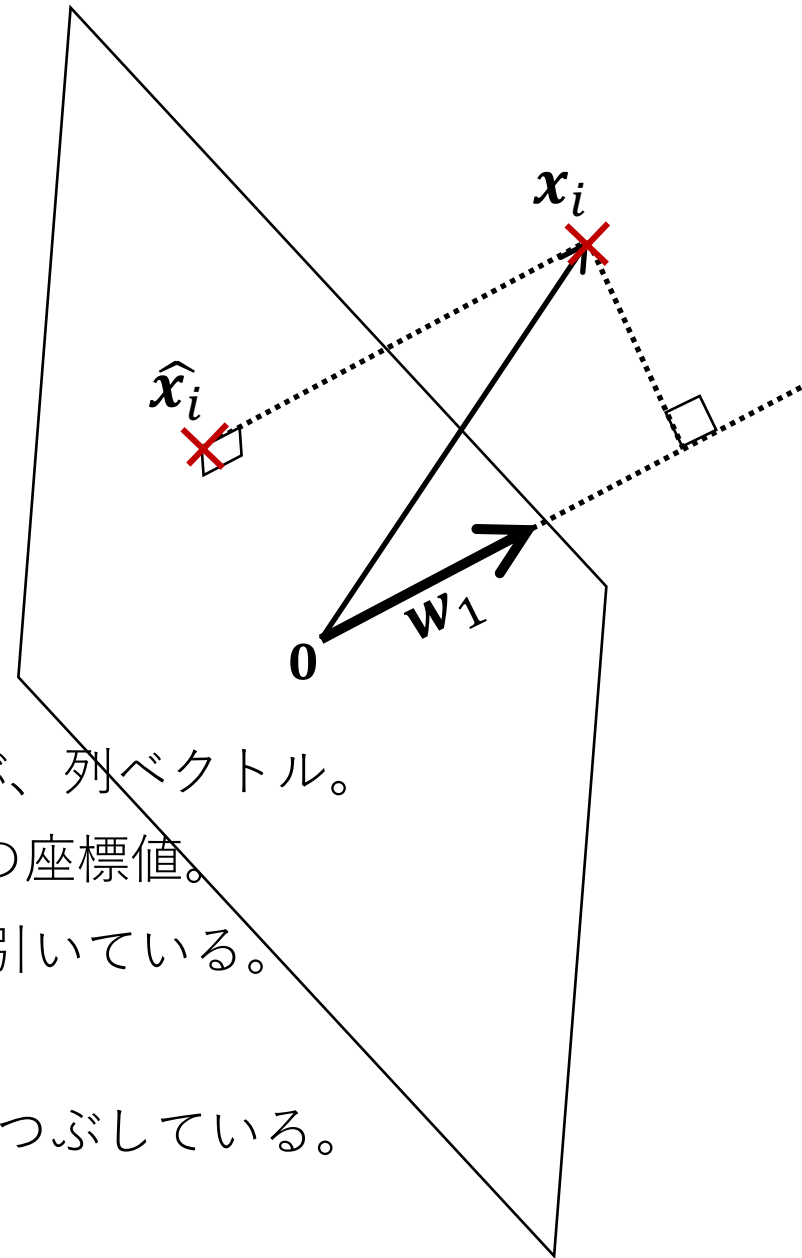
# 「データ点を押しつぶす」

- 式で書くと以下の通り

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

- $\mathbf{X}\mathbf{w}$ は列ベクトル

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

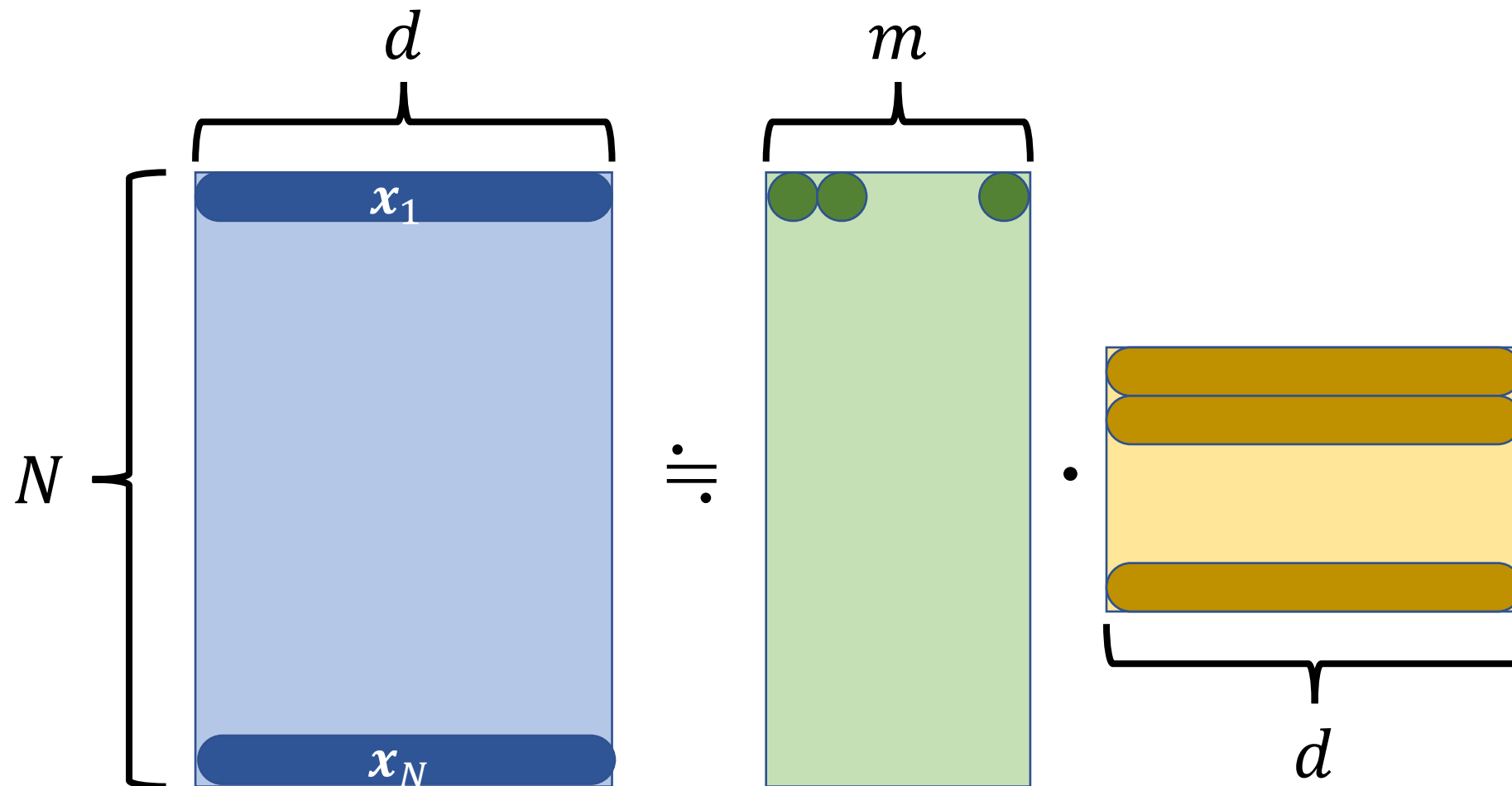


# PCAの（実際の）アルゴリズム

1. データを中心化する（平均を引く）
  2. 共分散行列 $\mathbf{C} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}$ を計算する
  3. 共分散行列 $\mathbf{C}$ 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。

# 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





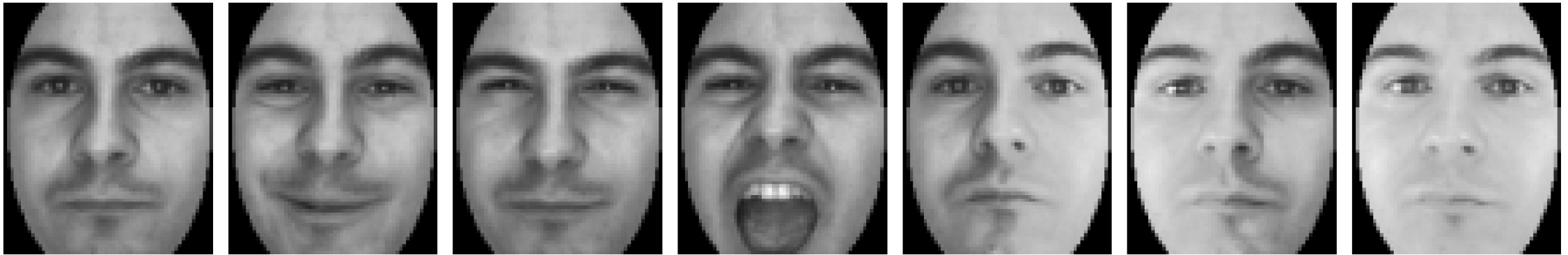
# NMF (Nonnegative Matrix Factorization)

- 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

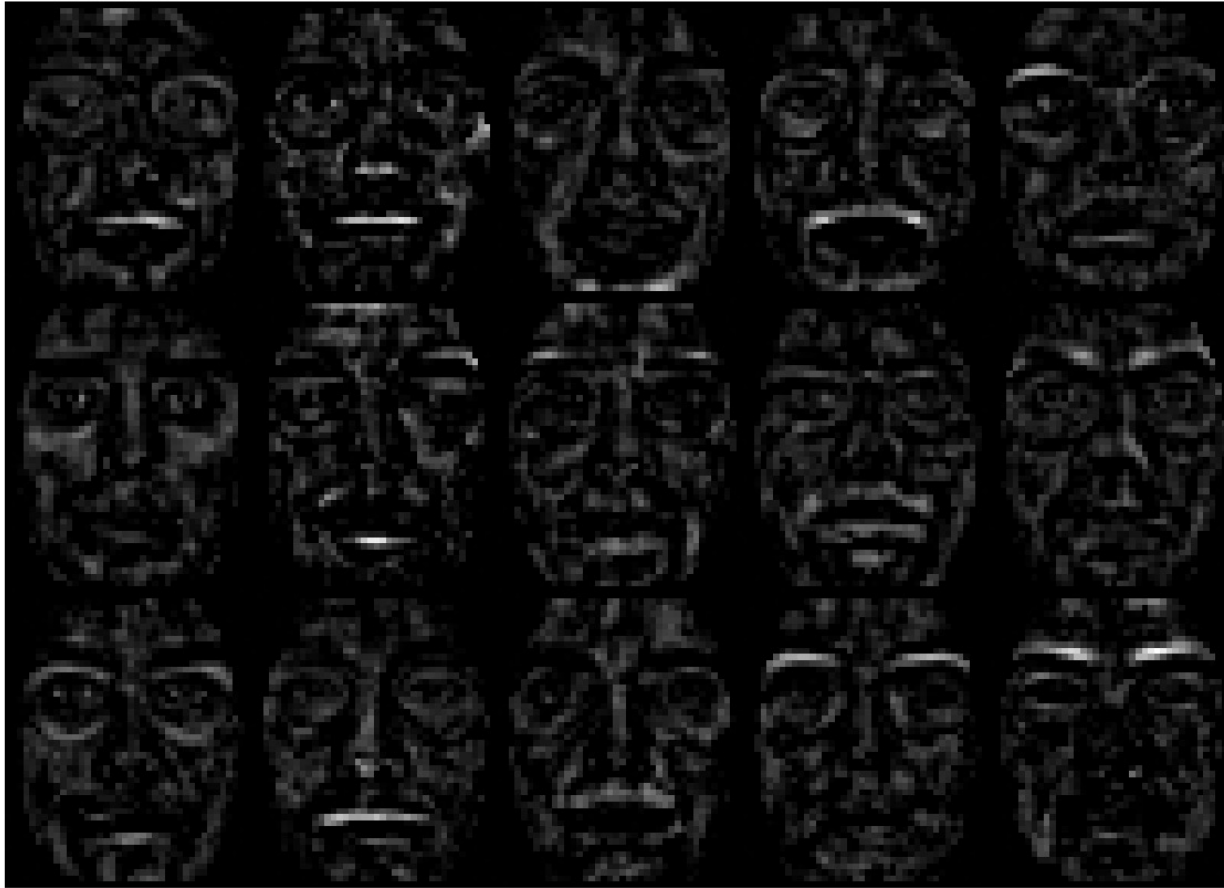
- PCAでいう主成分に相当するベクトルが非負ベクトルとして得られる
- 得られる非負ベクトルは、お互いに直交しないこともありうる
  - ただし、疎なベクトルになるので、ほぼ直交しているともみなせる。



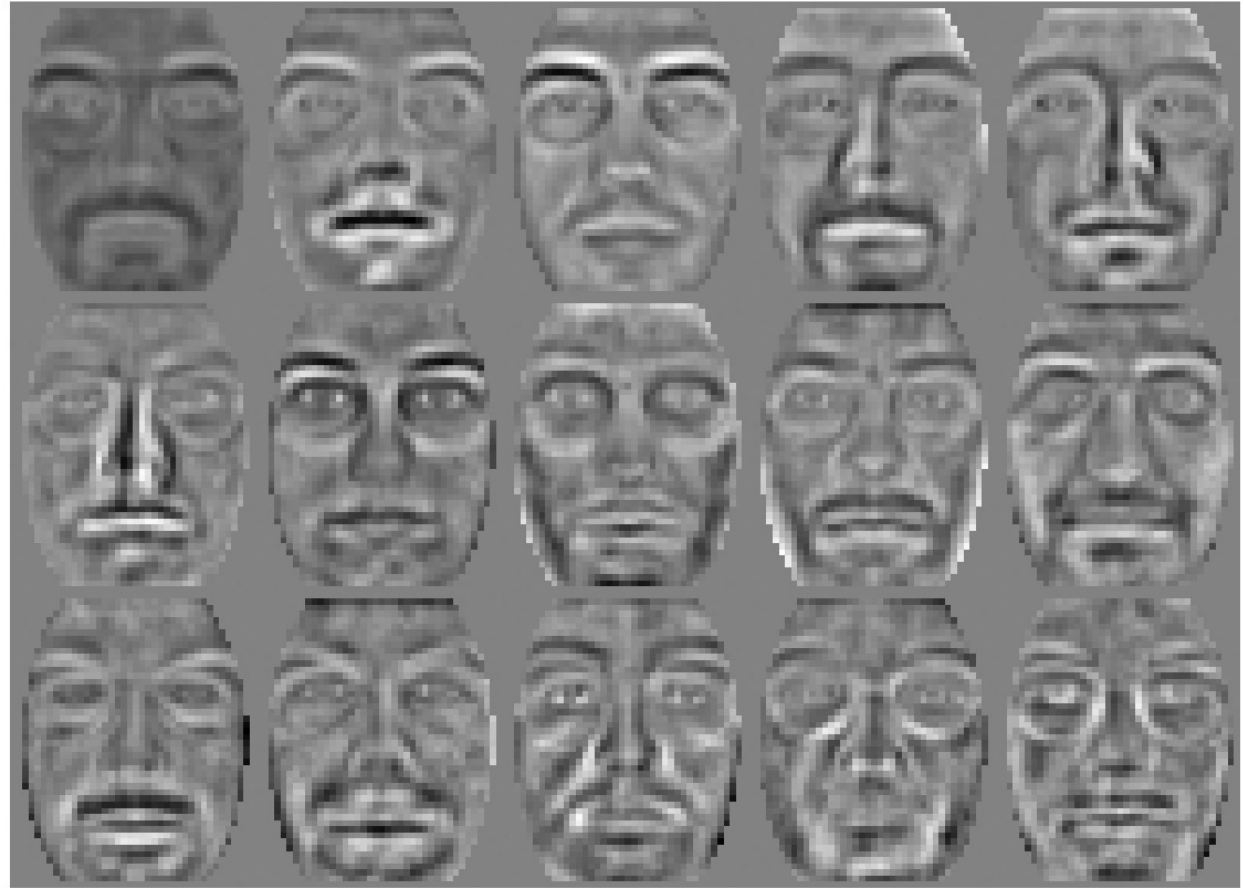
AR 01    AR 02    AR 03    AR 04    AR 05    AR 06    AR 07



AR 08    AR 09    AR 10    AR 11    AR 12    AR 13



(a) NMF bases.



(b) PCA bases.

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

# PLSI (probabilistic latent semantic indexing)

- NMFの確率モデル版
- 詳細は来年夏学期の「統計モデリング2」で
- EMアルゴリズムで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
- 詳細は来年夏学期の「統計モデリング2」で
- 変分ベイズ推定で事後分布のパラメータを推定する

# 今日の課題

- 手話画像を再構成してみる
- inverse\_transformメソッドを持つクラスを使う
  - <https://scikit-learn.org/stable/modules/classes.html#module-sklearn.decomposition>
- 手法ごとの特徴のようなものはあるか？
  - [https://scikit-learn.org/stable/auto\\_examples/decomposition/plot\\_faces\\_decomposition.html](https://scikit-learn.org/stable/auto_examples/decomposition/plot_faces_decomposition.html)