次元削減・行列分解

Fukuta Keisuke

Agenda

1. 線形代数復習

2. PCA・SVD

3. NMF

次元削減

「高次元データを低次元空間に射影すること」

目的

- 次元の呪いを回避する
- データ構造を理解する、可視化する
- 汎化性能を上げる

次元削減 (機械学習的には)

- 機械学習では基本すべての特徴を等価に扱う

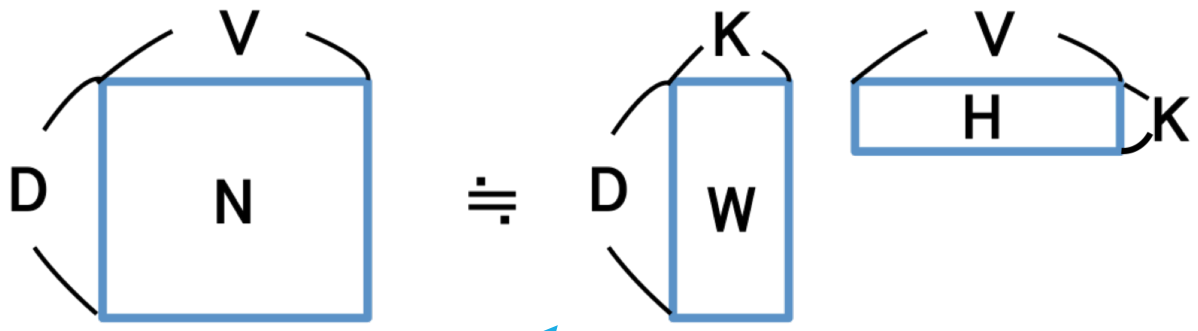
- 例 : 製品の複数の特徴から製造所を予想
  - 特徴の中にはいくつも相関があるものがある。
    - Ex. 長さと重さは恐らく相関がある

    -> 長さ、密度なら相関なさそう
    - こういうのを機械に自動的にやってほしい

長さ	幅	重さ	…	製造所ID
3.0	5.1	100.2	…	A
2.9	5.3	100.1	…	A
3.4	5.2	98.2	…	B
2.9	4.8	101.0	…	C
⋮	⋮	⋮	⋮	⋮

次元削減としての行列分解

- 行列を二つ以上の行列の積に分解して近似すること (行列の低ランク近似)
  - PCA, SVD
  - NMF

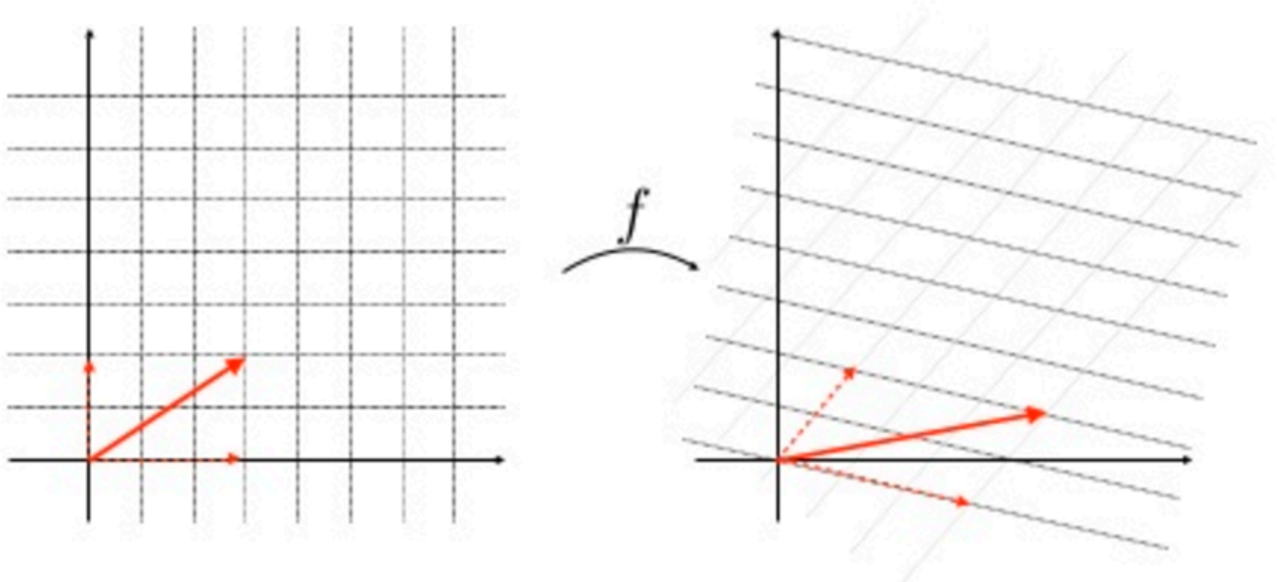


ここに元の特徴が現れるようにすれば、次元削減としても使える

線形代数復習

- 行列 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$をベクトル$\mathrm{v} = \begin{pmatrix} x \\ y \end{pmatrix}$に作用させる操作 $f(\mathrm{v}) = \mathrm{Av}$

➔ ベクトルの線形変換

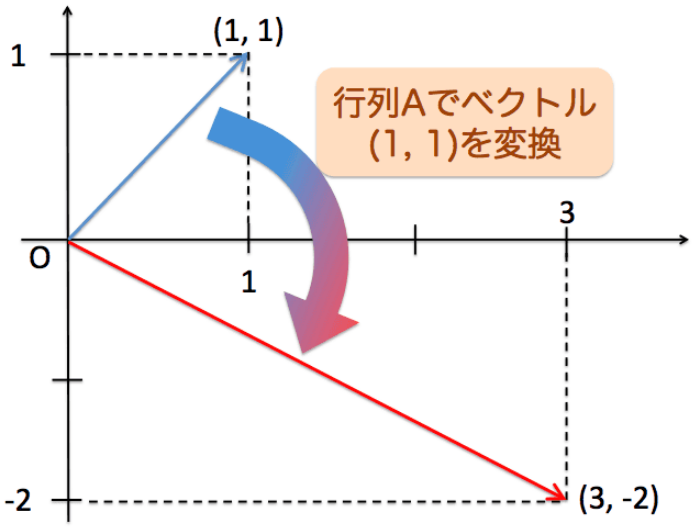


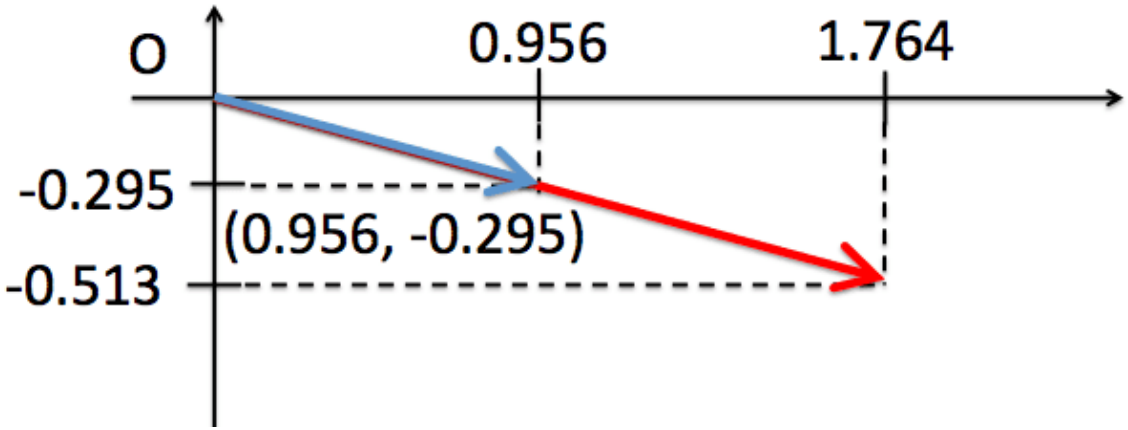
ベクトルの**線形性**（平行と比率）の保たれる変換

線形代数復習

$Av = \lambda v$ となるようなλを**固有値**、xを**固有ベクトル**と呼ぶ

- 線形変換Aに対して向きが変わらないベクトルが固有ベクトル
- 固有ベクトルが何倍されるっていうのが、固有値





線形代数復習

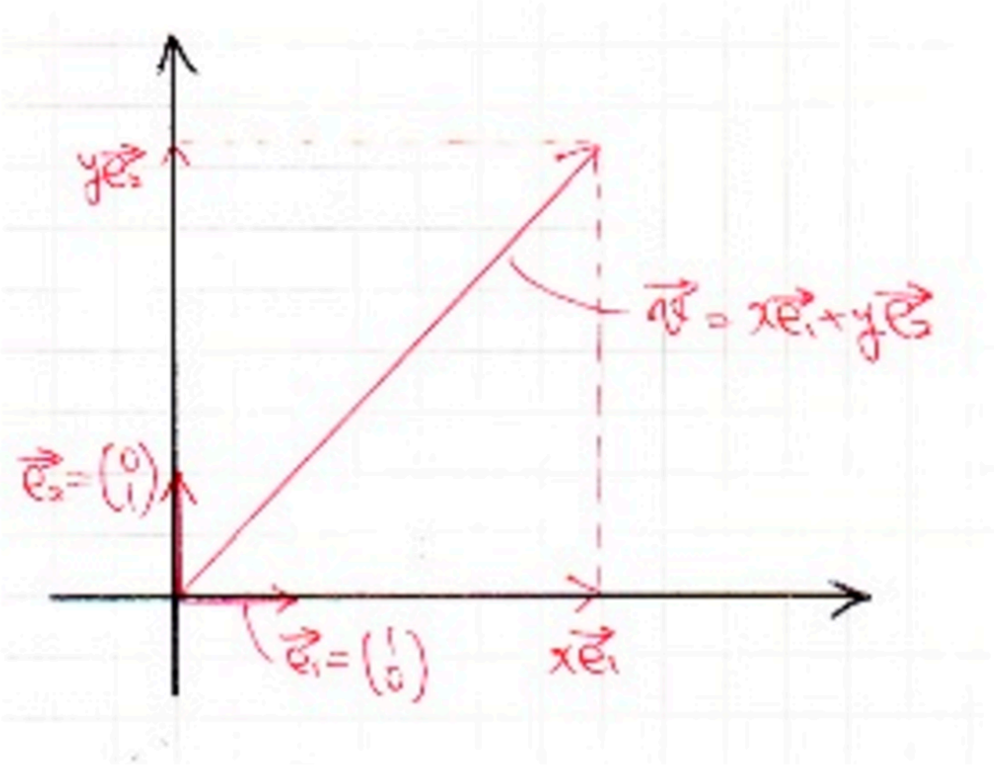
- 基底変換

普通$\mathrm{v} = \begin{pmatrix} x \\ y \end{pmatrix}$って言うときは

$\mathrm{v} = \mathrm{x} \cdot e_1 + y \cdot e_2$のことを指していた

このときe_1, e_2を基底と呼ぶ

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



線形代数復習

- 基底変換

e_1, e_2を基底に使わないといけないわけじゃなくて、

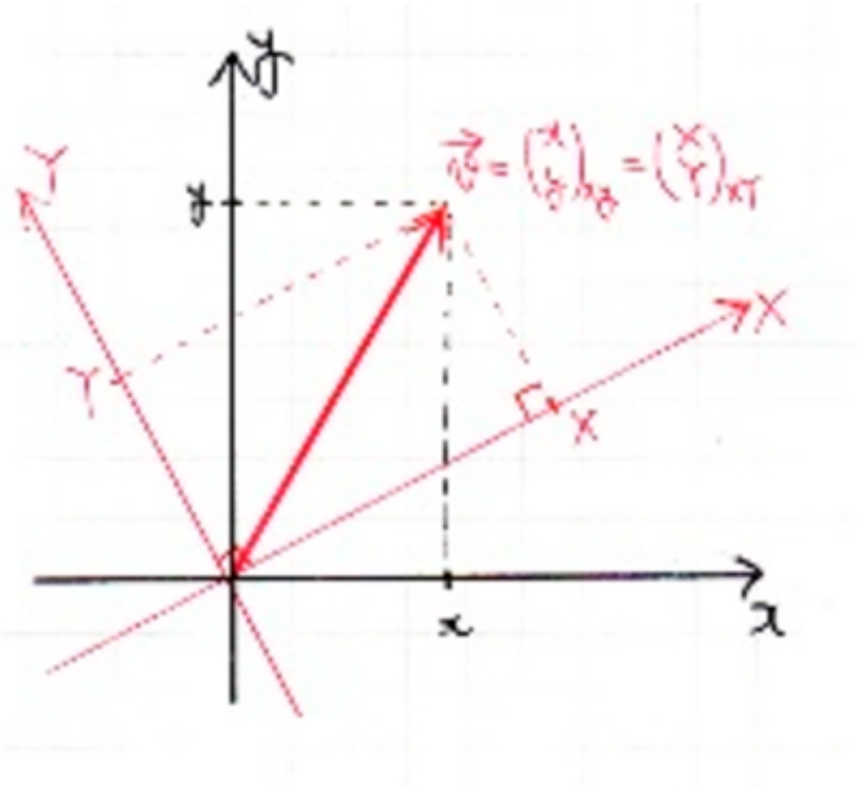
こんな回転された基底のほうが考えやすいことがある

新しい基底をそれぞれu_1, u_2とすると

$$\begin{pmatrix} x \\ y \end{pmatrix} = X \cdot u_1 + Y \cdot u_2$$

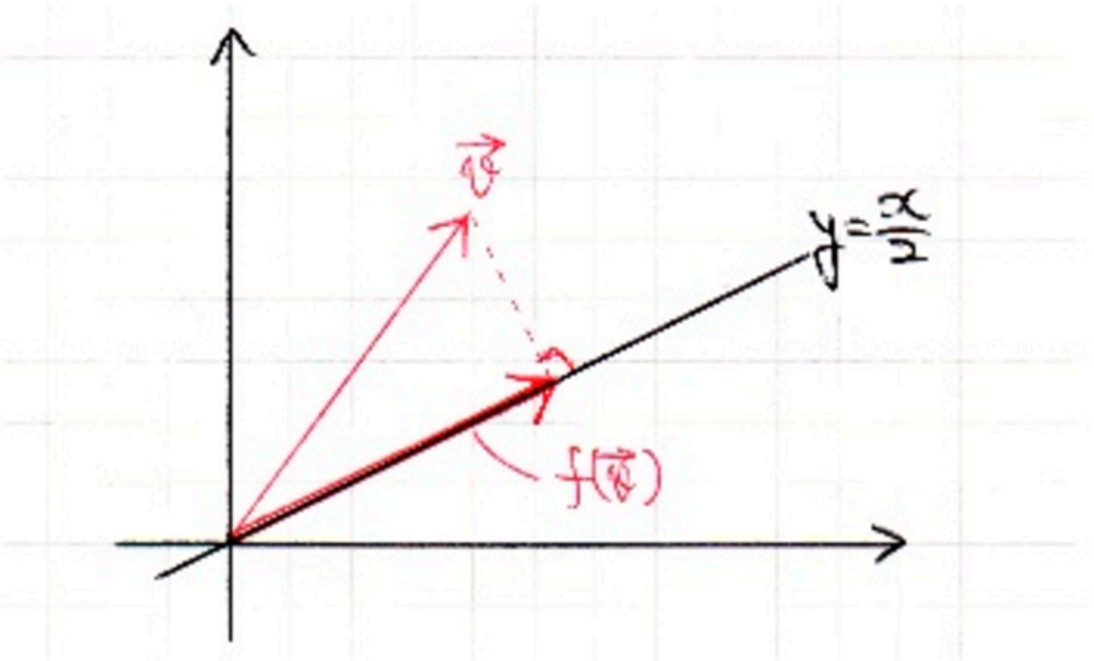
$$\begin{pmatrix} x \\ y \end{pmatrix} = U \begin{pmatrix} X \\ Y \end{pmatrix} \quad (U = (u_1\ u_2)\text{と置く})$$

$$\begin{pmatrix} X \\ Y \end{pmatrix} = U^{-1} \begin{pmatrix} x \\ y \end{pmatrix}$$ <- 基底変換の基本式

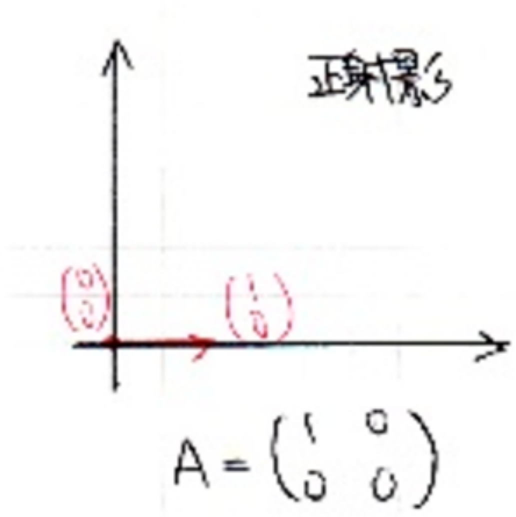


線形代数復習

(例1) 直線 $y = x/2$ への正射影



この変換Aを求めてみたい



x軸への射影だったら超簡単だけど、これはどうすれば？

線形代数復習

XY座標系での f の行列表示は $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ で、

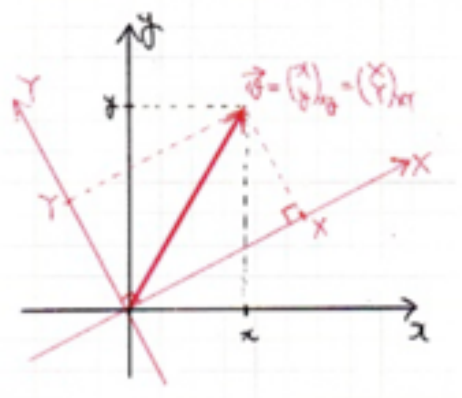
$$f(\vec{v}) = f\left(\begin{pmatrix} x \\ y \end{pmatrix}_{xy}\right) = \left(A\begin{pmatrix} x \\ y \end{pmatrix}\right)_{xy}$$ … xy座標系での表現

$$= f\left(\begin{pmatrix} X \\ Y \end{pmatrix}_{XY}\right) = \left(B\begin{pmatrix} X \\ Y \end{pmatrix}\right)_{XY}$$ … XY座標系での表現

$$= \left(BU^{-1}\begin{pmatrix} x \\ y \end{pmatrix}\right)_{XY}$$

$$= \left(UBU^{-1}\begin{pmatrix} x \\ y \end{pmatrix}\right)_{xy}$$ … xy座標系での表現

より、xy 座標系での f の行列表示は、$A = UBU^{-1}$



$$\begin{tikzcd}
\begin{pmatrix} X \\ Y \end{pmatrix} \arrow[r, "B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}"] & B\begin{pmatrix} X \\ Y \end{pmatrix} \arrow[d, "U = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}"] \\
\begin{pmatrix} x \\ y \end{pmatrix} \arrow[u, "U^{-1} = \frac{1}{5}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}"] \arrow[r, Rightarrow, "A"] & A\begin{pmatrix} x \\ y \end{pmatrix} = UBU^{-1}\begin{pmatrix} x \\ y \end{pmatrix}
\end{tikzcd}$$

線形代数復習

- 行列の対角化

いい感じのUという行列を取ると$U^{-1}AU = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$ 対角行列にすることができる

変形すれば、$A = U\mathrm{BU}^{-1}$

実は、、

- Uは固有ベクトルによる基底行列　$Ex.\ U = (u_1\ u_2)$
- Bは対角成分に固有値が並んだ対角行列　$Ex.\ B = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$

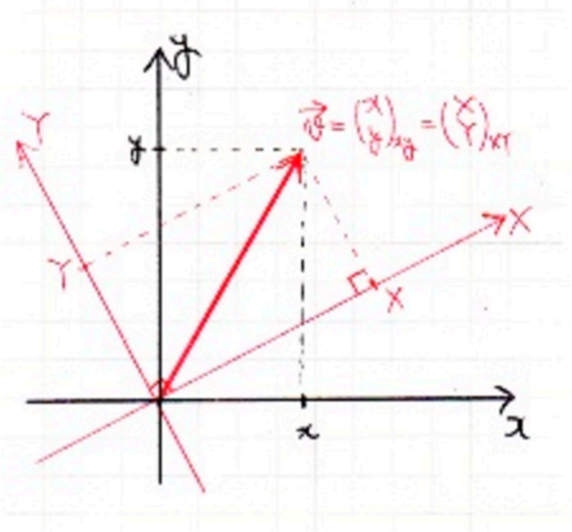
線形代数復習

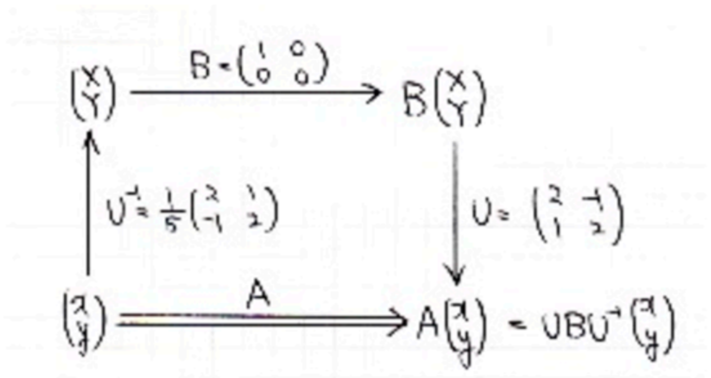
$$A = U\mathrm{B}\mathrm{U}^{-1}$$

解釈すると、

固有ベクトルを基底とする基底変換を施すと、行列Aによる変換は、それぞれのそれぞれの方向に固有値倍したものだとみなせる。

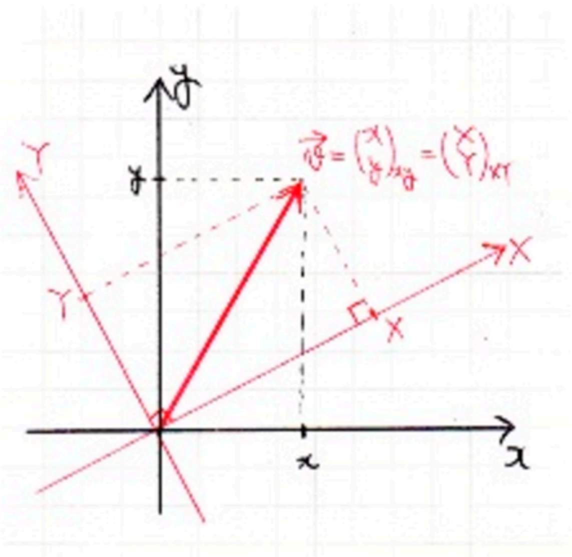
更にその後また元の基底に戻せば行列Aの変換と同じ。

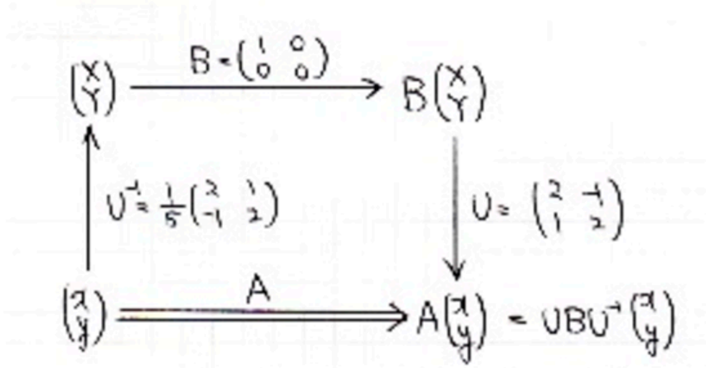




線形代数復習

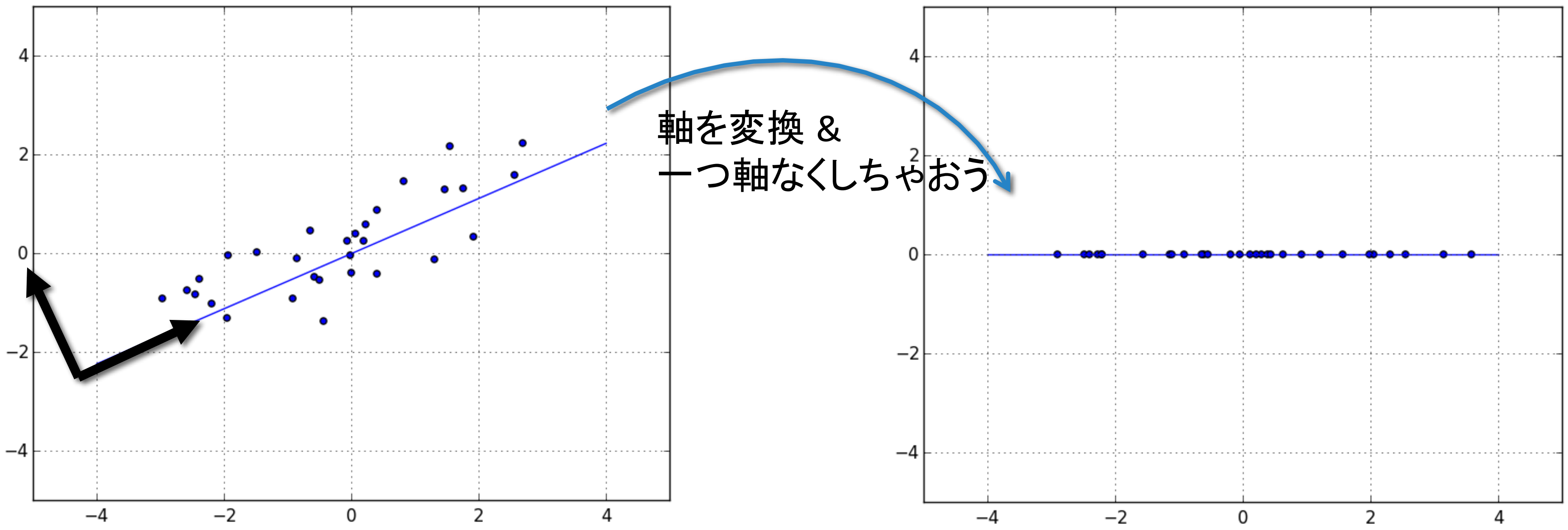
何が言いたいか.. -> うまく基底を変換すれば見やすくできることがある！





主成分分析 (Principal Component Analysis)

- 可視化、次元削減によく使われる！
- データ間に含まれる相関をなくしたい。
- できるだけ情報を落とさずにより小さい次元で表現したい



主成分分析 (Principal Component Analysis)

- データを行列Xで表現 ($X \in R^{n \times m}$)

$X =$

n

長さ	幅	重さ	…
3.0	5.1	100.2	…
2.9	5.3	100.1	…
3.4	5.2	98.2	…
2.9	4.8	101.0	…
⋮	⋮	⋮	⋮

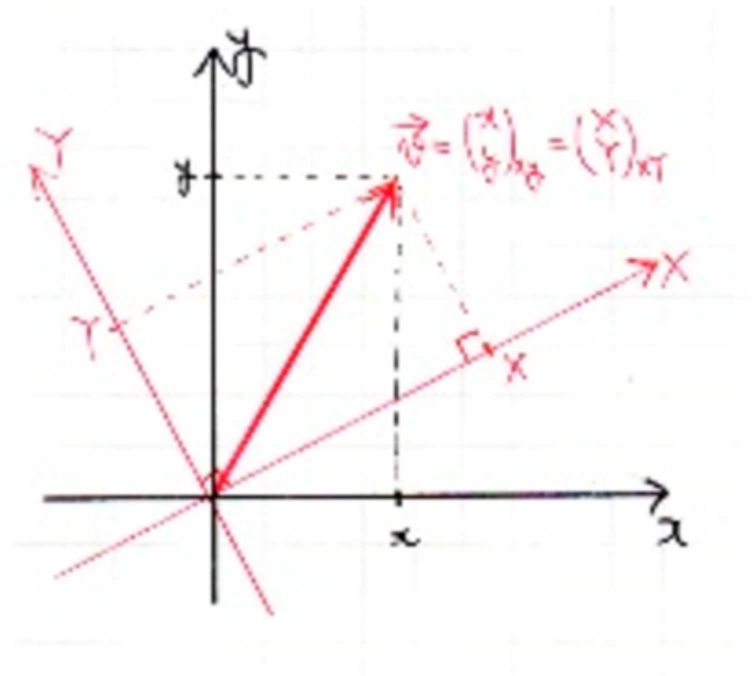
m

主成分分析 (Principal Component Analysis)

- 元の特徴量の線形結合によって新しい、いい感じの軸を作る

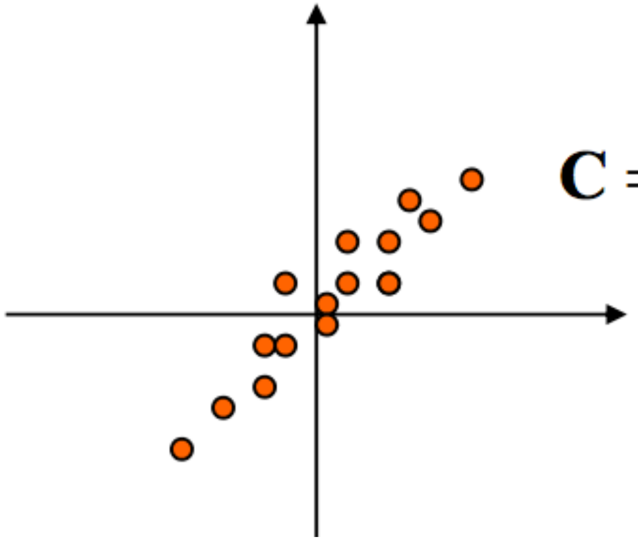
➔ 基底変換

- 基底変換は$\begin{pmatrix} X \\ Y \end{pmatrix} = U^T \begin{pmatrix} x \\ y \end{pmatrix}$で表せた。
- いい感じの$U = (\mathrm{u}_1\ \mathrm{u}_2)$を求めたい！！

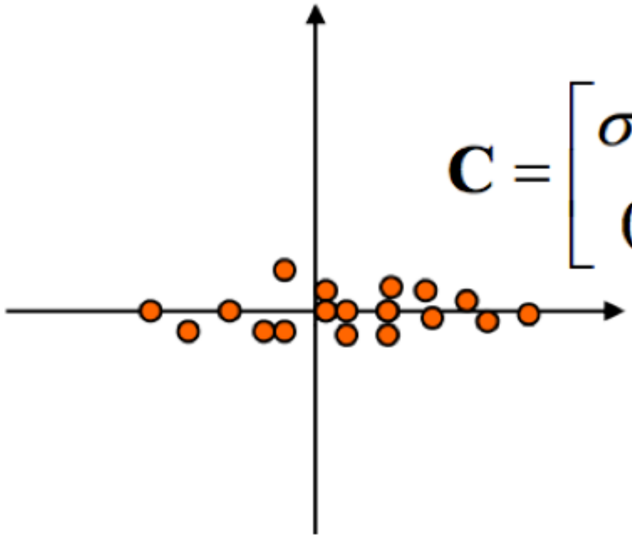


主成分分析 (Principal Component Analysis)

- Idea :「共分散行列の非対角成分が0になるような変換を施せばいいのでは？」
- 共分散行列 $\sum = XX^T$
- 基底変換後のYの共分散行列 $\rightarrow U^T \sum U$



$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$



$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

主成分分析 (Principal Component Analysis)

- 結局、$U^T \sum U$ が対角行列になればよい。

- さっき、行列の対角化やりましたよね！

➔ 行列は固有値ベクトルを並べた行列によって対角化できた！！

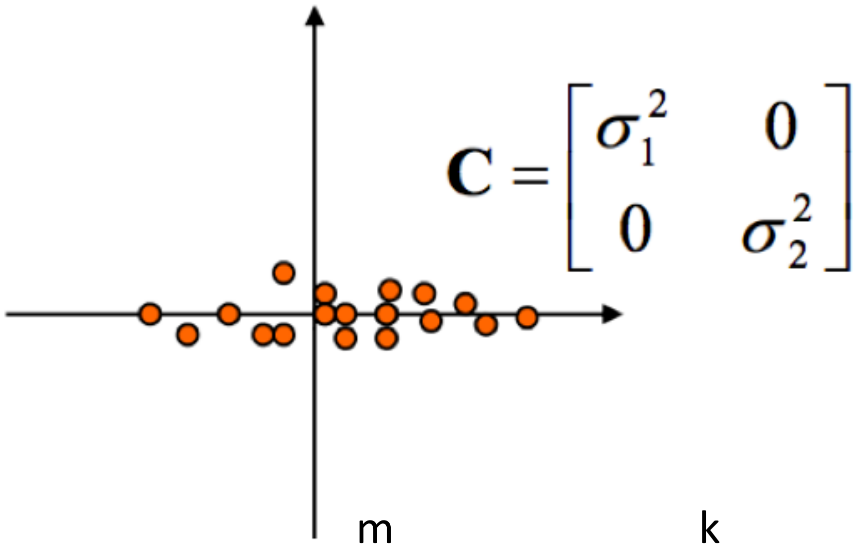
- 結局データ行列Xの共分散行列$\sum = XX^T$の固有ベクトルを求める問題に帰着

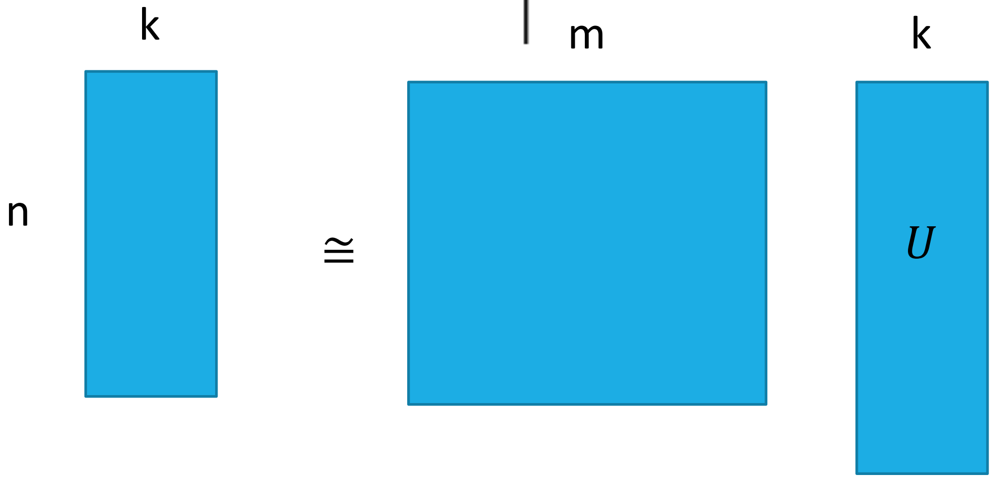
主成分分析 (Principal Component Analysis)

- どうやって次元削減するか
- 求めた基底変換$Y = U^T X$だと次元が減ってない
- 嬉しいことに、各固有値が、その軸の分散を表す

➔ 分散が小さい軸は使いたくない

➔ 固有値が大きい順に数個使えば良い

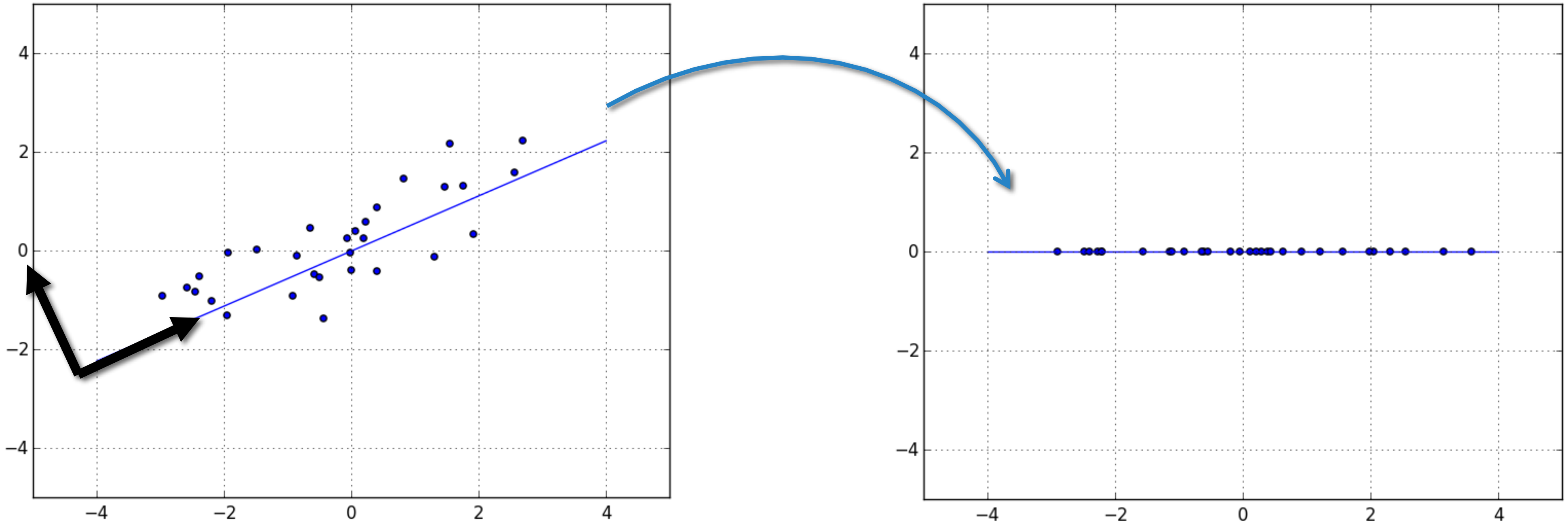




主成分分析 (Principal Component Analysis)

- どうやって次元削減するか

軸を変換 &
一つ軸なくしちゃおう



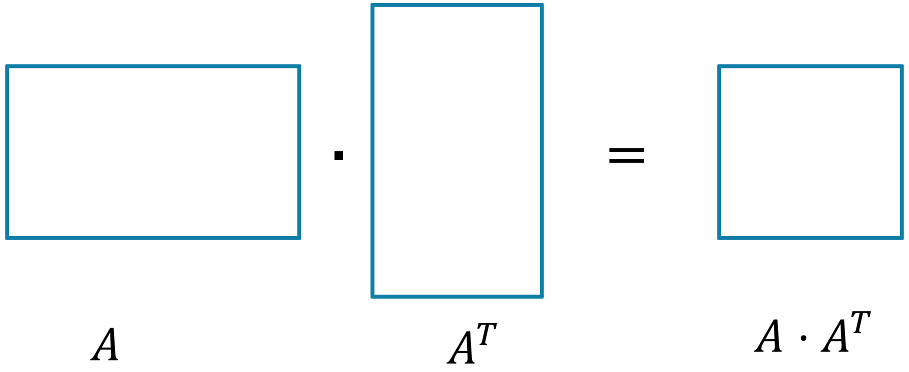
特異値分解 (Singular Value Decomposition)

特異値って何？？

- 固有値は正方行列じゃないと定義できない

➔ 正方でない場合にも拡張

- $\sigma(A) = \sqrt{\lambda(A \cdot A^T)}$ (行列Aの特異値 = 行列$A \cdot A^T$の固有値の平方根)



こんな感じで無理やり正方にして固有値計算

特異値分解 (Singular Value Decomposition)

行列$\mathrm{M} \in R^{m \times n}$に対して

$\mathrm{M} = U\,\Sigma\,V^T$ と分解すること

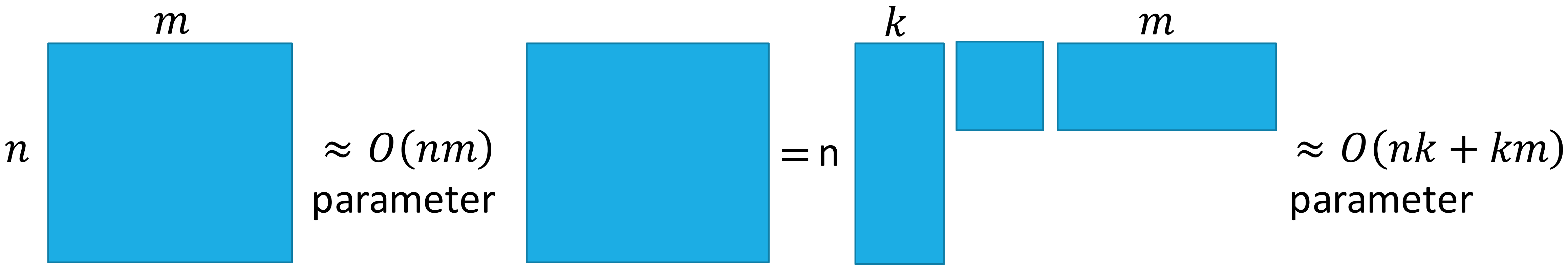
- U、Vはユニタリ行列(直交行列の複素数版)

- Σは右図のように、対角成分に特異値が並びそれ以外0

$$\Sigma = \left[\begin{array}{c|c} \begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{matrix} & O \\ \hline O & O \end{array}\right]$$

特異値分解 (Singular Value Decomposition)

- SVDによる低ランク近似
  - さっきのPCAのように、大きい特異値しか使わずに行列を表現
  - 特徴抽出だけでなく省メモリという意味合いも



$$\widetilde{M} = U\,\widetilde{\Sigma}\,\mathrm{V}^{\mathrm{T}}$$

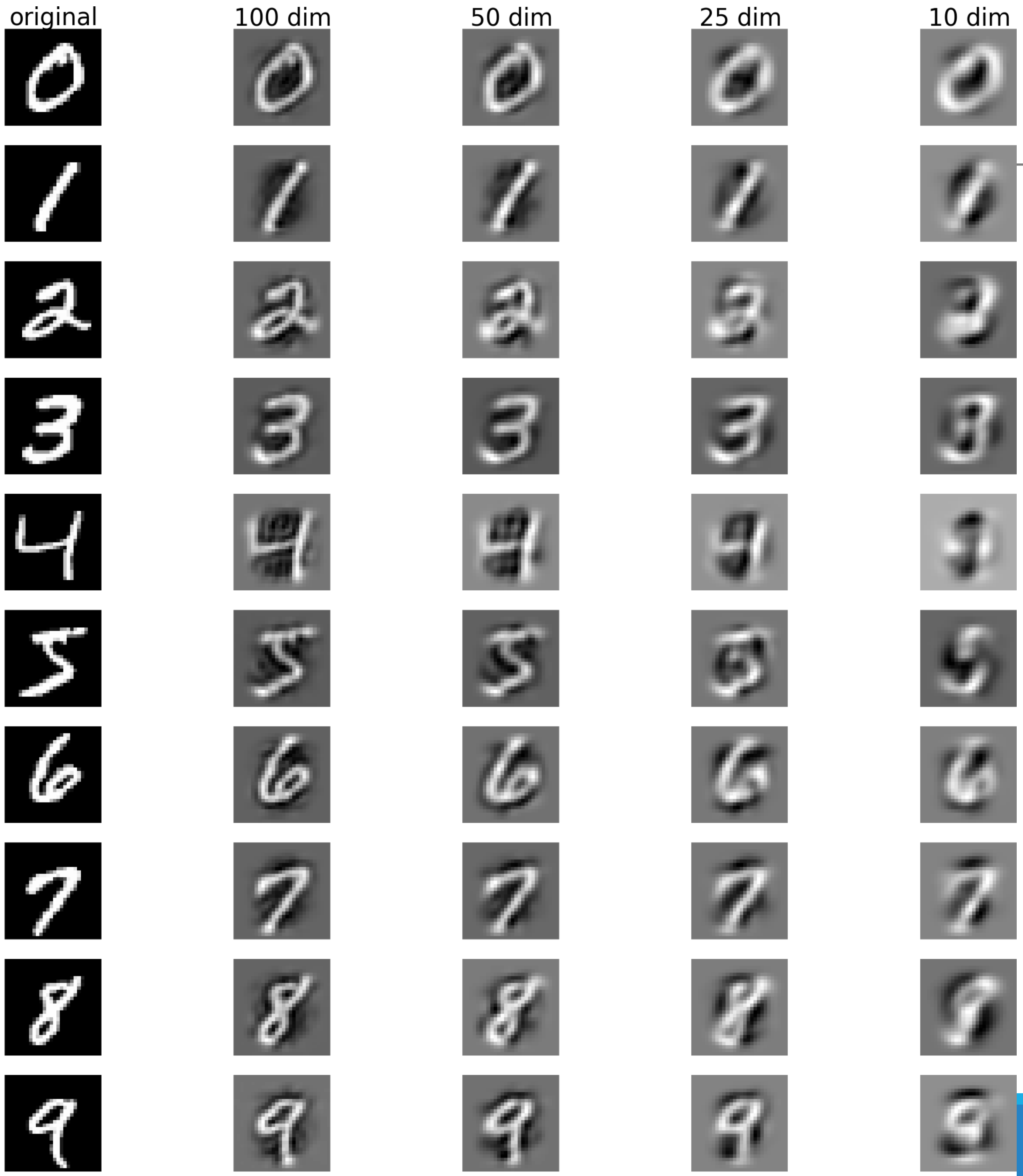
SVDとPCA

- 実はPCAとSVDはほぼ等価
- どちらも行列Xに対してXX^Tの固有値問題を解いている
- SVDの$\mathrm{M} = U\,\Sigma\,V^T$のU, Vは$XX^T, X^T X$のそれぞれの固有ベクトル
- ちょっと固有ベクトルのところが違う（たぶん）

SVDとPCA

意義

- 特徴抽出
  - 特徴選択となるだけの場合も
- 情報を落とさずに次元削減
  - 次元の呪いを回避、高速に動作させる
  - 可視化 (t-sneと同じ、併用されることもある)


original
100 dim
50 dim
25 dim
10 dim

NMF

- 行列Xの要素が全部非負となるようなデータ
  - Ex. 勾配データ、遺伝子発現の有無など

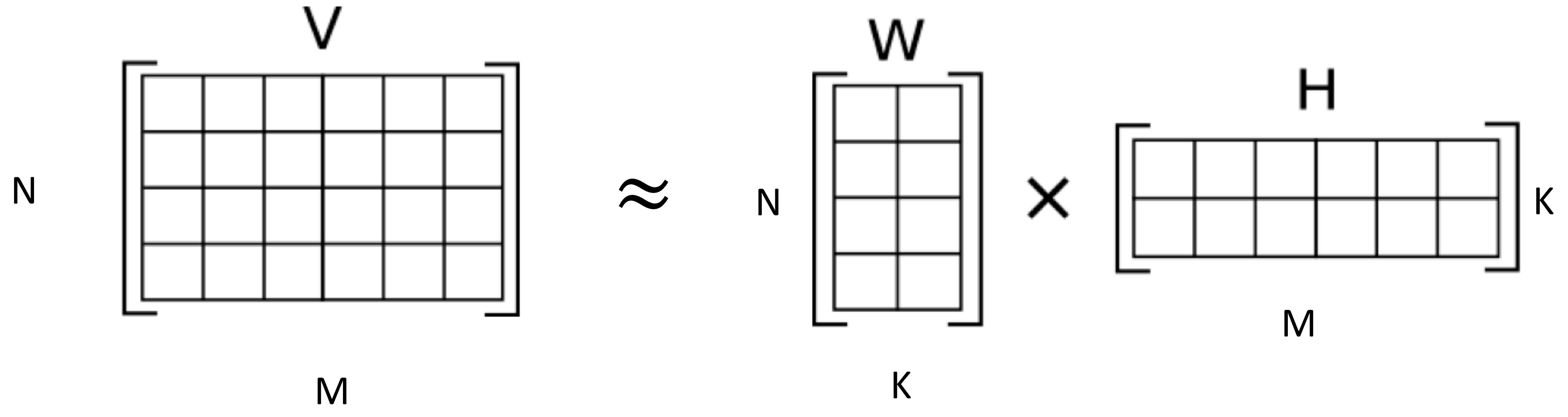
- 分解後(近似後)も非負のまま扱いたい
  - 解釈しやすい
  - また、負の数が使えない場合0を表現するには0しかないこともポイント (sparseになる)

NMF

NMFとは…

要素が全て非負の行列Vに対して、

$V \approx W \cdot H$ と分解すること (W, Hともに非負）



NMF

PCA、SVDのように厳密解ではなく数値解析的に解く

解くべき問題は

$$\min |V - WH|_F{}^2 \quad \mathrm{s.t.\,W,H} \geq 0$$

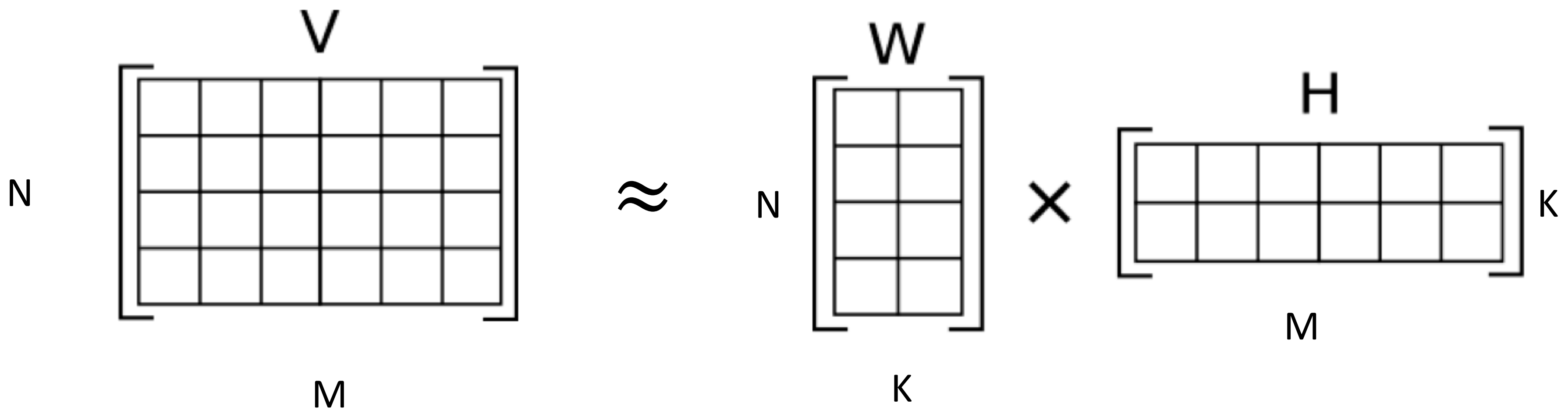
いくつか解法があるが、よく使われるのは乗法更新式と呼ばれるもの

適当にW, Hを初期化して、W, Hが収束するまで以下の更新式によって更新

$$W \leftarrow W \times \frac{\mathrm{VH^T}}{WHH^T} \quad ,\mathrm{H} \leftarrow \frac{W^TV}{W^TWH}$$

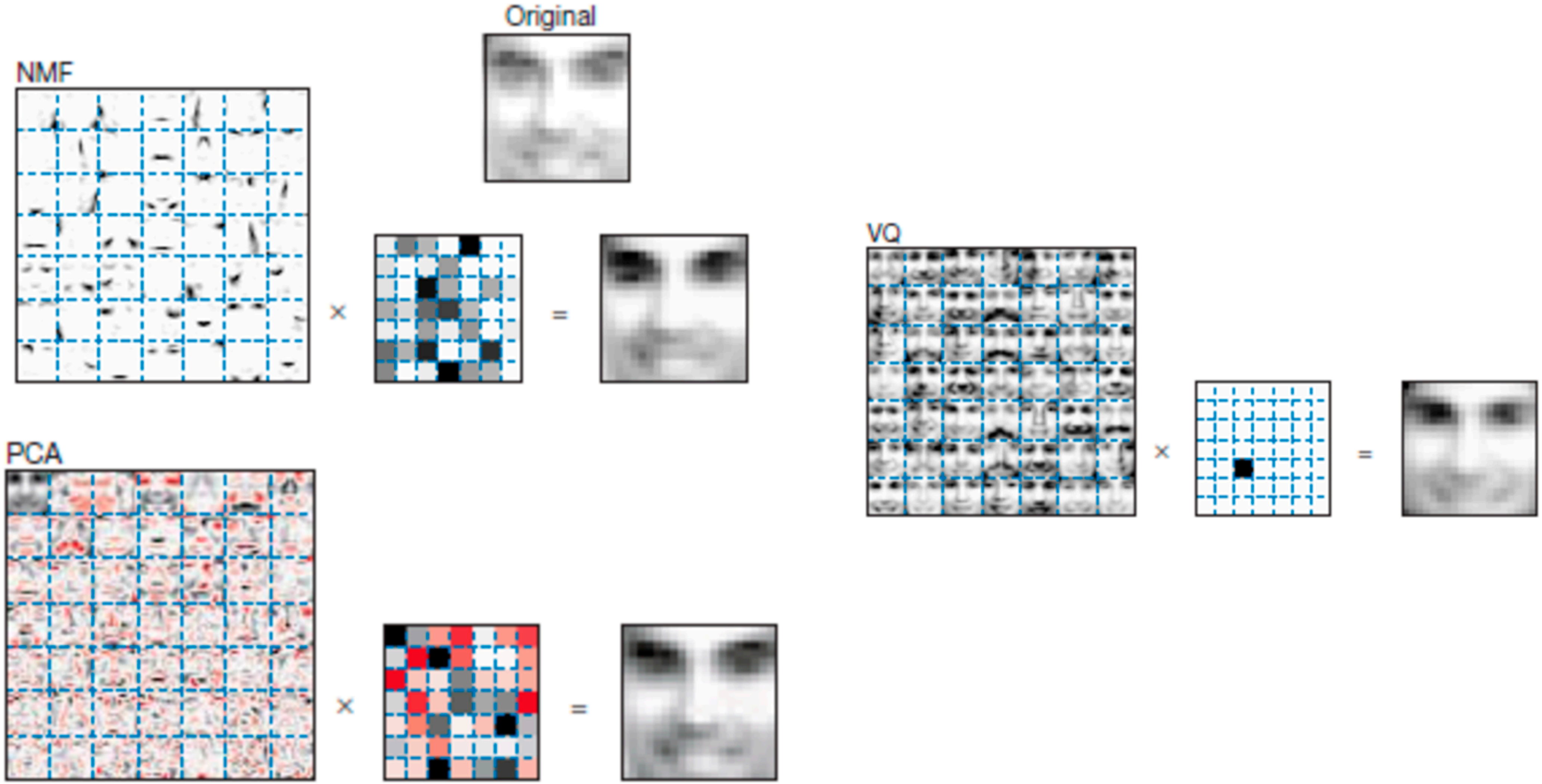
Learning the parts of objects by non-negative matrix factorization

- W, Hは何を表すのかの例
- m人の顔画像を表現する行列$(V: N\times M)$を$(W: N \times K,\ \ H: K \times M)$分解.



Wの各列は、基底画像
Hの各列は、それぞれの人は基底画像をどれだけの重みで結合して作られるか

Learning the parts of objects by non-negative matrix factorization



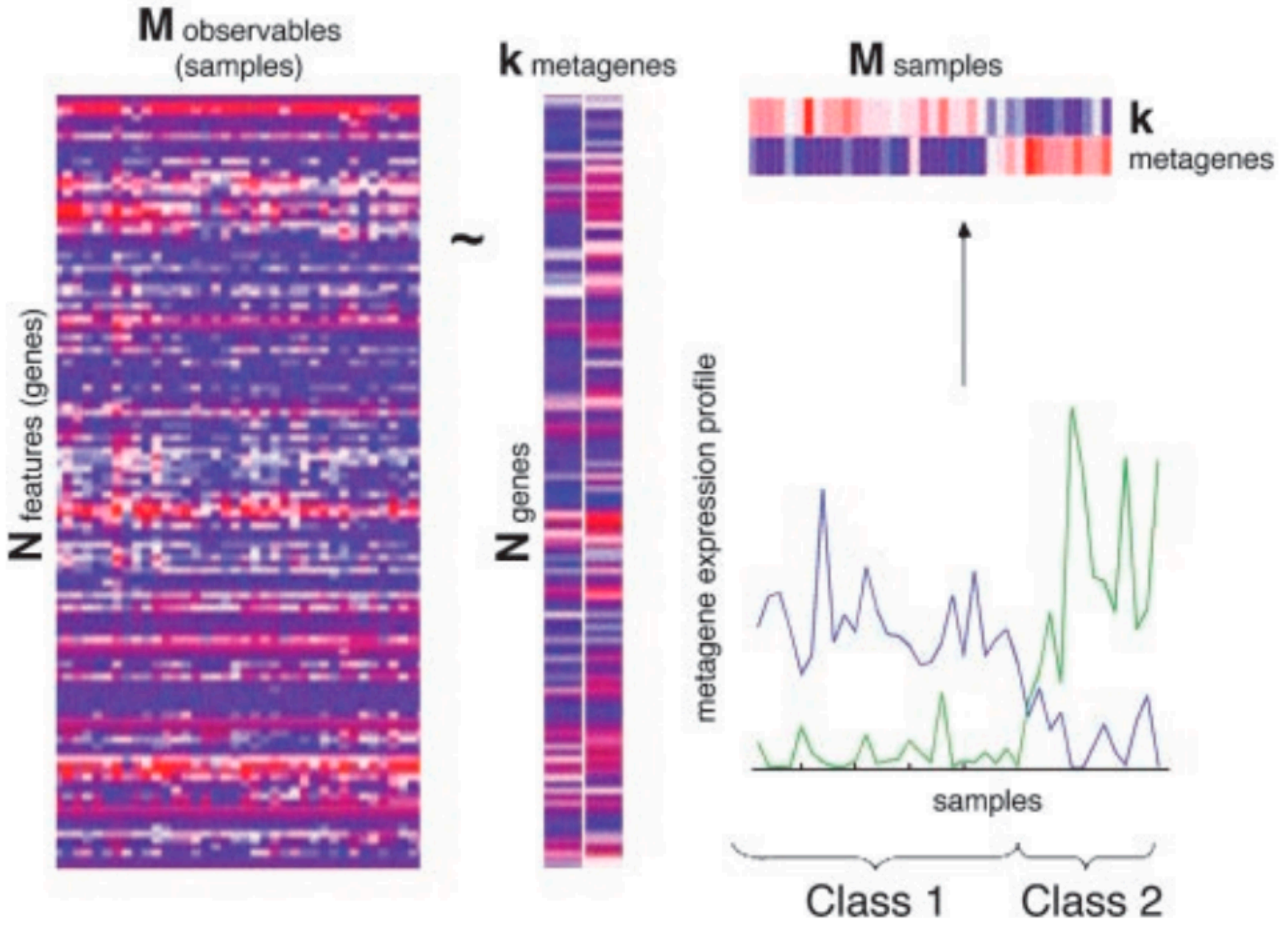
NMF

- 特徴
  - 基底が単純なものに (引き算ができないので)
  - 因子行列の方はスパースに
  - 結果として、解釈しやすい
  - データに欠損があっても使用できる
    - 欠損がないところだけ更新すれば良いので
    - SVDなどでは無理

遺伝子発現データをNMF

- N個の遺伝子表現
- Mサンプルのデータ

- NMFで二つの行列（rank 2)
- 因子行列の要素が明らかに
  二分された
- このとき、基底はなんらかの
  表現パターンを表すことがわかる



まとめ

- PCAはデータの共分散行列が対角になるように基底変換を行い
  分散の大きい軸だけ残す

- SVDはPCAとほぼ等価の行列の低ランク近似のための手法

- NMFは非負データの解析に利用される。
  解き方が単純なので欠損データがあっても柔軟に対応できる