授業計画

Prologue: データを読み解くとは何なのか
(多変量解析とデータサイエンスと統計学とパターン認識と機械学習とデータマイニング)

DAY-1 6/16 (01)(02) **単回帰: 点群への直線当てはめを"真剣に"考える**
(見えない世界へようこそ)
DAY-2 6/23 (03)(04) **重回帰と線形代数: 回帰の行列計算とその意味**
(データの計算とデータの解釈)
DAY-3 6/30 (05)(06) **重回帰と確率統計: なぜ回帰に確率が必要?**
(推測統計入門:データの向こう側について語るための代償)
DAY-4 7/07 (07)(08) **多変量正規分布: 多次元の正規分布と線形代数**
(ゼロから理解する正規分布)
DAY-5 7/14 (09)(10) **マハラノビス距離と判別分析: 線形代数を使う1**
(最適な判別とは)
DAY-6 7/28 (11)(12) **固有値分解と主成分分析: 線形代数を使う2**
(高次元データがかかえる大問題)
DAY-7 8/04 (13)(14) **特異値分解と数量化: 線形代数を使う3**
(数値じゃない対象に統計を効かすには)

Epilogue: 基礎の上に在る世界(話したことと話さなかったこと)

先週配布したレポートを提出すること！

提出先:　8/4の授業日 or 6-16瀧川居室ドアポスト or takigawa＠ist.hokudai.ac.jpへメール提出

〆切： 8/5(金) 17:00まで (難しい場合は事前に相談してください！)

注意：レポートには必ず名前と学生番号を記載すること！！

【講義Webサイト】

いままでの講義スライドや下の表1のデータファイルを置いています。

注: httpじゃなくhttpsでないと資料にアクセスできません。(たぶん)

https://art.ist.hokudai.ac.jp/~takigawa/course/da2016
(閲覧ID da2016 パスワード pokemongo)

データ解析： 8/4(木) 本日の内容

- 数量化理論とは?
- 質的データと量的データ
- 尺度とダミー変数/ダミーコーディング
- 「数量化III類」
  - 林知己夫によって提案(1955年ごろ)
  - 1935年のHirschfeldの論文をBenzécriが発展させた「対応分析」(1962年ごろ)と同等
  - 西里静彦の「双対尺度法」(1980年代)も同等
  - 統計ソフトでは「等質性分析」「多重応答分析」等
- 特異値分解 (SVD)

数量化理論とは？

- **質的データ(カテゴリカルデータ)**を扱う統計手法群
- 最近の統計学の本にはたぶんあまり載っていない。
  - オーソドックスな統計手法に帰着可能だから
    - **数量化I類 → ダミー変数を用いた回帰分析**
    - **数量化II類 → ダミー変数を用いた判別分析**
    - **数量化III類 → 対応分析/コレスポンデンス分析(CA)**
    - **数量化IV類～VI類 → 多次元尺度構成法(MDS)**
  - 日本独自の呼び名であり、上の理由により国際的に名前の認知度は低い…

数量化理論：定性情報の数量化とは？

林 知己夫 (1918-2002)

1974-1986年

統計数理研究所七代所長

- 林知己夫を中心として発展した多変量解析理論
- 「質的データに対して数値はアプリオリに与えるべきではない」
- 刑期1/3で累犯のおそれのないものを仮釈放する→累犯者予測
  裁判の態度、刑務所生活の勤勉度がそのまま累犯の説明にならず
- 「定性情報(裁判の態度)を1,2,3,4…」等と数量化するとダメ
- 数量は釈放の成功・不成功が最も当てはまるようにデータから定めるべき、という立場が生まれる。(数量化理論の動機付け)

1.3 数量化1類とは

表1.4は大学卒業時の総合成績 y，線形代数の成績 x_1 およびサークル所属の有無 x_2 のデータである．x_1 と x_2 は共に質的変数である．

表 1.4 成績のデータ

サンプル No.	線形代数 x_1	サークル x_2	総合成績 y
1	優	所属	96
2	優	所属	88
3	優	無所属	77
4	優	無所属	89
5	良	所属	80
6	良	無所属	71
7	良	無所属	77
8	可	所属	78
9	可	所属	70
10	可	無所属	62

このデータに基づいて知りたいことは次の通りである．

(1) 総合成績は線形代数の成績とサークル所属の有無より予測できるか．

(2) 予測できるとすればその精度はどのくらいか．

(3) 例えば，線形代数が優でサークルに無所属の学生の総合成績はどのように予測されるか．

1.5 数量化2類とは

表 1.6 は健常者とある疾病にかかっている患者に対する吐き気の程度 x_1（質的変数）と頭痛の程度 x_2（質的変数）のデータである．

表 1.6　健常者・患者の症状のデータ

サンプル No.	健常者・患者	吐き気 x_1	頭痛 x_2
1	健常者	無	少
2	健常者	少	無
3	健常者	無	無
4	健常者	無	無
5	健常者	無	無
6	患者	少	多
7	患者	多	無
8	患者	少	少
9	患者	少	多
10	患者	多	少

このデータに基づいて知りたいことは次の通りである．

(1) 疾病にかかっているか否かを吐き気と頭痛より判別できるか．

(2) 判別できるとすればその精度はどのくらいか．

(3) 例えば，吐き気が無く，頭痛が多い人はどのように判別されるか．

1.7 数量化3類とは

表1.8は，児童10人の得意科目のデータである．各科目は，各サンプルに対して○印が「あるか」「ないか」のいずれかなので，質的変数と考える．

表1.8 児童の得意科目のデータ（○印が得意科目）

児童 No.	国語	社会	算数	理科	音楽	図工	体育
1	○			○		○	
2			○		○	○	
3	○						○
4	○	○	○	○			
5		○					○
6				○	○	○	
7			○	○	○		
8	○	○		○			○
9			○			○	○
10	○	○	○	○			

このデータに基づいて知りたいことは次の通りである．

(1) 科目と児童に数量を与え，低い次元でデータを解釈できないか．

(2) そのような数量化によって説明力はどれくらいあるか．

(3) 科目や児童の特徴付けおよび分類をどのようにできるか．

1.6 主成分分析とは

表 1.7 は，4 教科の試験の成績である．すべて量的変数と考える．

表 1.7 試験の成績のデータ

生徒 No.	国語 x_1	英語 x_2	数学 x_3	理科 x_4
1	86	79	67	68
2	71	75	78	84
3	42	43	39	44
4	62	58	98	95
5	96	97	61	63
6	39	33	45	50
7	50	53	64	72
8	78	66	52	47
9	51	44	76	72
10	89	92	93	91

この（4 次元）データに基づいて知りたいことは次の通りである．

(1) 主成分の構成により低い次元でデータを解釈できないか．

(2) それぞれの主成分の説明力はどれくらいか．

(3) 科目や生徒の特徴付けおよび分類をどのようにできるか．

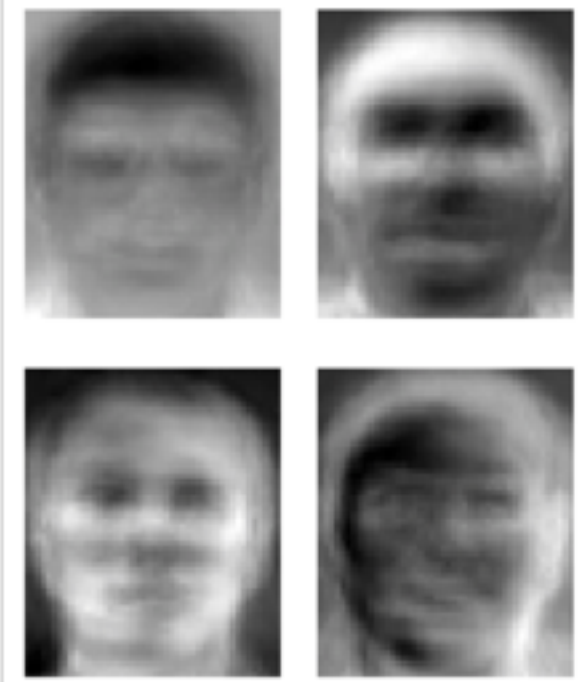
固有顔 (Eigenface): 顔画像の主成分分析

【例 6.13】 図 6.9 の上段左端は 50 人の顔画像（32×32 画素）をそれぞれ 24 通りの照明条件で撮影した，合計 1200 枚の顔画像の平均画像 $\boldsymbol{g}_0$ である．この 1200 枚の画像から計算した固有ベクトルから作られる基底の最初の 4 個 $\boldsymbol{g}_1,\dots,\boldsymbol{g}_4$ をその右に順に並べてある．

図 6.9 上段: 平均顔画像と第 1, 2, 3, 4 基底画像．中段と下段: 入力画像とその第 1, 2, 3, 4 近似．

固有顔

固有顔（英: **Eigenface**）とは、顔認識システムというコンピュータビジョンの応用で使われる固有ベクトルの集合である。固有顔を利用した顔認識は1987年、Matthew Turk と Alex Pentland が開発した。



固有顔の例。AT&T研究所より

この固有ベクトル群は、「人間の考えられる顔」の高次元ベクトル空間の確率分布についての共分散行列から得られる。

目次 [非表示]

1 固有顔生成
2 顔認識における利用
3 関連項目
4 参考文献
5 外部リンク

固有顔生成 [編集]

固有顔集合を生成するには、人間の顔のデジタル化された画像（同じ照明条件でなければならない）を多数集積し、目と鼻の位置を合わせる。そして、解像度を合わせるために再標本化する。このようなデータから主成分分析 (PCA) と呼ばれる手段で固有顔を抽出する。顔画像を固有顔に変換する過程は以下の通りである。

カテゴリカルデータ分析

今まで暗黙のうちに変数は「数量」としてきた。
… 平均値や分散を計算したりできる

現実で解析したいデータはしばしば数量ではない。

（例えば社会調査におけるアンケート分析を考えよ）

「カテゴリカルデータ」
性別、Yes/No、国、色、車種、銘柄、社名、都道府県、etc.

… {赤,赤,赤,青,黄,紫,紫}の平均値や分散って何！？

数量化

{赤,赤,赤,青,黄,紫,紫}の平均値や分散って何！？

カラーコード 赤=1,青=2,黄=3,紫=4 を付与してみる

$$\{赤,赤,赤,青,黄,紫,紫\} \xrightarrow[\text{encode}]{} \{1,1,1,2,3,4,4\}$$

標本平均 2.285714

標本分散 1.904762 (標準偏差 1.380131)

この値にどういう意味が?? 平均は青と黄の間!?

意味のわからない結果になった原因

カラーコード 赤=1,青=2,黄=3,紫=4 を付与してみる

→この「数字」にしたとき、変な序列ができてしまった…

コード Yes=1, No=0 を付与して

データ{Yes, Yes, No, No, Yes, No, Yes, Yes}を分析

→ {1, 1, 0, 0, 1, 0, 1, 1}

→標本平均 0.625 は回答者集団のYesの率を意味する

(この場合、問題が特に起こらない)

One-Hotエンコーディング

デジタル回路では電圧High→1, Low→0で2値化を考える

{赤,赤,赤,青,黄,紫,紫}

	X1	X2	X3	X4
赤	1	0	0	0
赤	1	0	0	0
赤	1	0	0	0
青	0	1	0	0
黄	0	0	1	0
紫	0	0	0	1
紫	0	0	0	1

one-hot(ワン・ホット)は
1つだけHigh(1)であり、
他はLow(0)であるような
ビット列のことを指す

参考) 逆の一つだけ0で
他が1の場合はone-cold

One-Hot Encodingと多重共線形性

One-Hot Encoding ⟶ $X1 + X2 + X3 + X4 = 1$

$\Leftrightarrow X4 = 1 - (X1 + X2 + X3)$

多重共線形性(Multicolinearity)

	X1	X2	X3	X4
赤	1	0	0	0
赤	1	0	0	0
赤	1	0	0	0
青	0	1	0	0
黄	0	0	1	0
紫	0	0	0	1
紫	0	0	0	1

X1～X4を説明変数とする回帰モデル

$$Y \sim \beta0 + \beta1\ X1 + \beta2\ X2 + \beta3\ X3 + \beta4\ X4$$

$$= \beta0 + \beta1\ X1 + \beta2\ X2 + \beta3\ X3 + \beta4 - \beta4\ X1 - \beta4\ X2 - \beta4\ X3$$

$$= (\beta0-\beta4) + (\beta1-\beta4)\ X1 + (\beta2-\beta4)\ X2 + (\beta3-\beta4)\ X3$$

変数3つ、未知数4つなので一意に決まらない！

ダミー変数 / ダミーコーディング

One-Hot Encoding

	X1	X2	X3	X4
赤	1	0	0	0
赤	1	0	0	0
赤	1	0	0	0
青	0	1	0	0
黄	0	0	1	0
紫	0	0	0	1
紫	0	0	0	1

$$X1 + X2 + X3 + X4 = 1$$

ダミー変数

	X1	X2	X3
赤	1	0	0
赤	1	0	0
赤	1	0	0
青	0	1	0
黄	0	0	1
紫	0	0	0
紫	0	0	0

(X4をリファレンスとして)

X1 … (紫と比べた)赤の貢献

X2 … (紫と比べた)青の貢献

X3 … (紫と比べた)黄の貢献

コード Yes=1, No=0 を付与して

データ{Yes, Yes, No, No, Yes, No, Yes, Yes}を分析

→ {1, 1, 0, 0, 1, 0, 1, 1}

One-Hot Encoding

	X1	X2
1	1	0
2	1	0
3	0	1
4	0	1
5	1	0
6	0	1
7	1	0
8	1	0

ダミー変数

	X1
1	1
2	1
3	0
4	0
5	1
6	0
7	1
8	1

Noを基準とした
1ダミーで済む

「男 or 女」などの場合、
解釈するとき基準に注意！
(どちらを1としたか?)

表5　平日・休日自主学習時間の規定要因（重回帰分析）

独立変数	所有財得点：上位層		所有財得点：下位層	
	偏回帰係数	標準化偏回帰係数	偏回帰係数	標準化偏回帰係数
男子ダミー	34.161	0.047	−14.898	−0.024
蔵書数	16.977	0.070 *	21.834	0.087 **
小5クラス内成績	9.117	0.026	8.266	0.025
通塾ダミー	−0.666	−0.001	37.378	0.059 +
母親によるしっかり勉強指導ダミー	128.034	0.171 ***	63.773	0.099 **
母親による多様な体験付与ダミー	73.136	0.082 **	−9.861	−0.015
議論授業	−0.282	−0.001	4.627	0.012
レポート授業	23.348	0.055	4.132	0.011
グループ授業	14.320	0.034	−0.200	−0.001
ドリル授業	27.622	0.074 *	−4.018	−0.012
テスト授業	−9.014	−0.025	−1.786	−0.006
教え方が上手な先生が多いダミー	−6.208	−0.009	46.753	0.072 *
意欲・態度得点	40.358	0.192 ***	31.410	0.168 ***
（定数）	−345.404		−87.663	
決定係数	0.107		0.061	
調整済み決定係数	0.096		0.048	
モデル適合度	p=0.000		p=0.000	
N	1009		918	

注：+：p<0.10、*：p<0.05、**：p<0.01、***：p<0.001。

http://berd.benesse.jp/berd/center/open/report/kanagawa_kouritsu/2010/hon2_2_05.html

表2：「大学進学希望」の規定要因（ロジスティック回帰分析）

独立変数		専門高校		普通科進路多様校	
		偏回帰係数	オッズ比	偏回帰係数	オッズ比
性別(基準：女子)	男子ダミー	−0.277	0.758	1.171	3.225***
出身階層(基準：親非大卒)	親大卒ダミー	0.418	1.518*	0.546	1.726+
中学成績	中２時成績	0.014	1.014	0.147	1.159
高校入学理由	大学進学の実績がよいから	0.501	1.650***	0.216	1.241
高校成績	現在の校内成績	0.139	1.149+	0.394	1.483**
重要な他者との進路相談	保護者との進路相談	0.134	1.144	0.165	1.179
	先生との進路相談	0.172	1.188+	0.585	1.796**
学科（基準：商業科）	工業科ダミー	−0.225	0.798		
	農業科ダミー	−0.730	0.482*		
	定数	−3.130	0.044***	−4.148	0.016***
	Nagelkerke 決定係数	0.129		0.220	
	モデル適合度	p=0.000		p=0.000	
	N	1,007		252	

注）+：p<0.10、*：p<0.05、**：p<0.01、***：p<0.001。

ロジスティック回帰 = 判別を目的変数が2値で回帰分析的に解く (詳細略)

http://berd.benesse.jp/berd/center/open/report/toritsu_senmon/2009/hon4_4_06.html

1.1 多変量データ

n 個のサンプルのそれぞれに対して p 個の変数（変量とも呼ぶ）$x_1, x_2, \cdots, x_p$ の値が観測されているとしよう．これは表 1.1 の形式にまとめることができる．この (サンプル)×(変数) の形式のデータを**多変量データ**と呼ぶ．

表 1.1　多変量データ

サンプル No.	変数（変量） x_1	x_2	$\cdots$	x_p
1	x_{11}	x_{12}	$\cdots$	x_{1p}
2	x_{21}	x_{22}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
i	x_{i1}	x_{i2}	$\cdots$	x_{ip}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	x_{n1}	x_{n2}	$\cdots$	x_{np}

多変量解析法で扱う変数は，離散的な値をとるなら**質的変数**，連続的な値をとるなら**量的変数**と区別する．さらに，これらは，表 1.2 に示すように**名義尺度**，**順序尺度**，**間隔尺度**，**比率尺度**に区分される．

表 1.2　データの区分

質・量	名称	特徴
質的	名義尺度	性別や職業などのようにカテゴリーの違いだけを表す
質的	順序尺度	優・良・可・不可のように順序に意味があるが，カテゴリー間の差は同じでない
量的	間隔尺度	温度のように，順序も間隔も意味があるが，原点の位置はどこでもよい
量的	比率尺度	長さや重さのように，間隔尺度であり，そして原点が定まっている

主な多変量解析手法の分類

外的基準(目的変数)あり　“教師あり”

目的変数=**量的** 説明変数=**量的**　　回帰分析

目的変数=**質的** 説明変数=**量的**　　判別分析

ロジスティック回帰分析

目的変数=**量的** 説明変数=**質的**　　数量化I類

目的変数=**質的** 説明変数=**質的**　　数量化II類

外的基準(目的変数)なし　“教師なし”

変数=**量的**　主成分分析, クラスター分析, 多次元尺度構成

変数=**質的**　数量化III類, IV類〜VI類, クロス表(分割表)分析

数量化理論とは？

- **質的データ(カテゴリカルデータ)**を扱う統計手法群
- 最近の統計学の本にはたぶんあまり載っていない。
  - オーソドックスな統計手法に帰着可能だから
    - **数量化I類 → ダミー変数を用いた回帰分析**
    - **数量化II類 → ダミー変数を用いた判別分析**
    - **数量化III類 → 対応分析/コレスポンデンス分析(CA)**
    - **数量化IV類～VI類 → 多次元尺度構成法(MDS)**
  - 日本独自の呼び名であり、上の理由により国際的に名前の認知度は低い…

数量化Ⅲ類 (カテゴリカルな主成分分析?)

表 10.1　児童の得意科目のデータ（○印が得意科目）

児童 No.	国語	社会	算数	理科	音楽	図工	体育
1	○			○		○	
2			○		○	○	
3	○						○
4	○	○	○	○			
5		○					○
6				○	○	○	
7			○	○	○		
8	○	○		○			○
9			○			○	○
10	○	○	○	○			

表 10.2　表 10.1 のデータの行と列の並び替え

児童 No.	音楽	図工	算数	理科	国語	社会	体育
2	○	○	○				
6	○	○		○			
7	○		○	○			
1		○		○	○		
9		○	○				○
4			○	○	○	○	
10			○	○	○	○	
8				○	○	○	○
3					○		○
5						○	○

各列と各行が似るように並び替え

低次元近似して
可視化

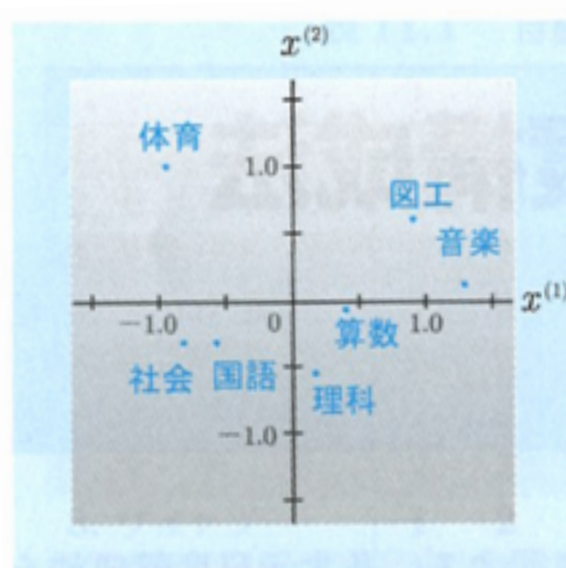


図 10.1　変数スコア散布図

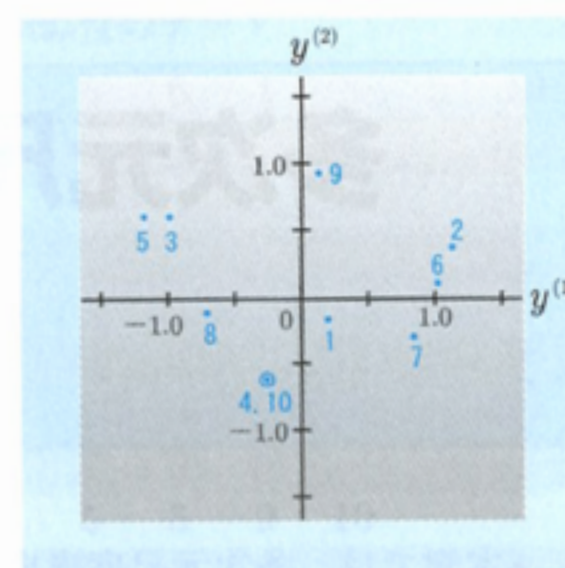


図 10.2　サンプルスコア散布図

カテゴリカルデータ分析の基本：クロス表(分割表)

今回は2値として扱うが対応分析は一般の二元クロス表に適用可

度数 列% 行%	1. かなり多い	2. 多いほう	3. ふつう	4. 少ない	4. 少ないほう	行　和 行%
1. たいへん気に入っている	166 54.43 31.68	239 27.19 45.61	86 18.49 16.41	7 10.14 1.34	26 11.40 4.96	524 26.93
2. まあ気に入っている	131 42.95 10.61	598 68.03 48.42	324 69.68 26.23	36 52.17 2.91	146 64.04 11.82	1,235 63.46
3. あまり気に入っていない	6 1.97 3.49	40 4.55 23.26	55 11.83 31.98	20 28.99 11.63	51 22.37 29.65	172 8.84
4. 気に入っていない	2 0.66 13.33	2 0.23 13.33	0 0.00 0.00	6 8.70 40.00	5 2.19 33.33	15 0.77
列　和 列%	305 15.67	879 45.17	465 23.90	69 3.55	228 11.72	1,946

http://www.wordminer.org/wp-content/uploads/2013/04/63_19.pdf

数量化Ⅲ類

行、列の各々に数量 x_i, y_j を割り付ける。

表 10.1　児童の得意科目のデータ（○印が得意科目）

児童 No.	国語	社会	算数	理科	音楽	図工	体育	
1	○			○		○		y_1
2			○		○	○		y_2
3	○						○	y_3
4	○	○	○	○				y_4
5		○					○	y_5
6				○	○	○		y_6
7			○	○	○			y_7
8	○	○		○			○	y_8
9			○			○	○	y_9
10	○	○	○	○				y_{10}
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	

例えば、数量 x_i, x_j は対応する列が似ていたら近い値にしたい。 y_i, y_j も似ている行は近い値にしたい。

復習：分散共分散行列の固有構造 (スペクトル分解)

$$\Sigma \quad = \quad P \quad \Lambda \quad P'$$

$$\begin{bmatrix} \sigma_{x_1,x_1} & \sigma_{x_2,x_1} & \cdots & \sigma_{x_p,x_1} \\ \sigma_{x_1,x_2} & \sigma_{x_2,x_2} & \cdots & \sigma_{x_p,x_2} \\ \vdots & & & \\ \sigma_{x_1,x_p} & \sigma_{x_2,x_p} & \cdots & \sigma_{x_p,x_p} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1p} \\ p_{21} & p_{22} & \cdots & p_{p2} \\ \vdots & & & \\ p_{p1} & p_{p2} & \cdots & p_{pp} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{p1} \\ p_{12} & p_{22} & \cdots & p_{p2} \\ \vdots & & & \\ p_{1p} & p_{2p} & \cdots & p_{pp} \end{bmatrix}$$

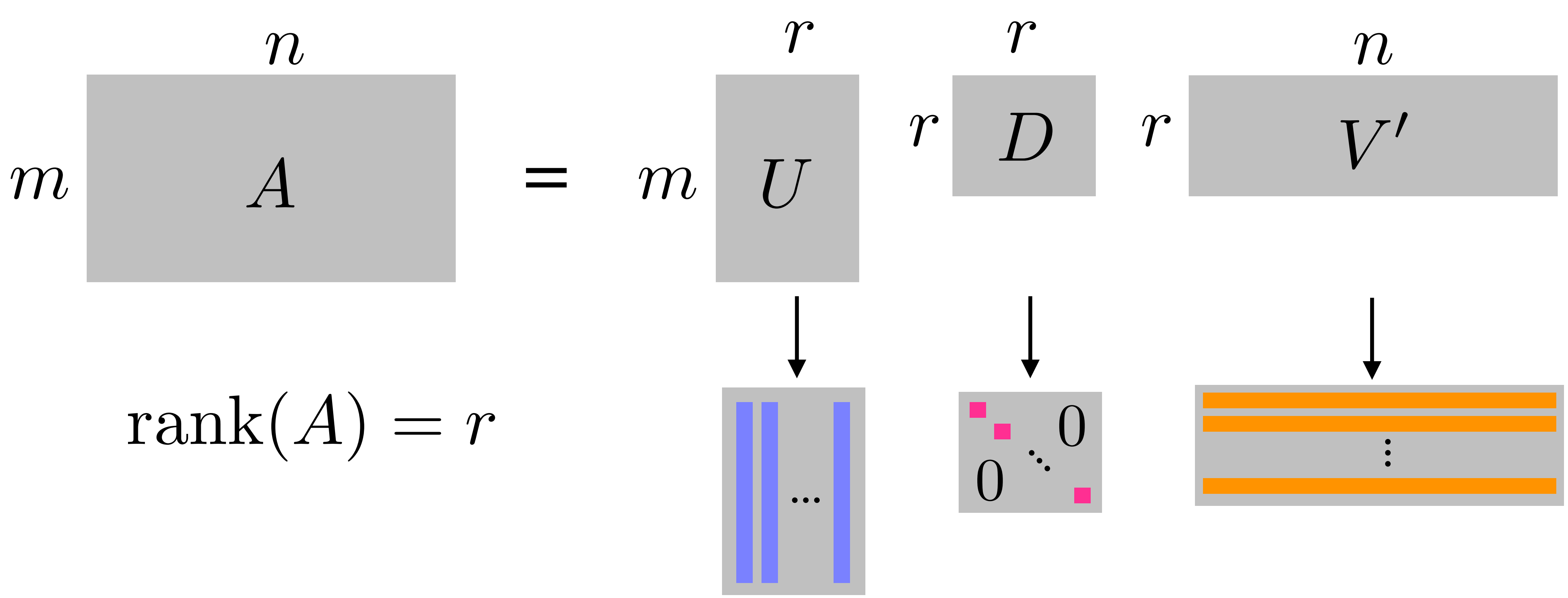
固有ベクトル　**固有値**　**固有ベクトル(転置)**

$$= \lambda_1 \begin{bmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{p1} \end{bmatrix} \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{p1} \end{bmatrix} + \lambda_2 \begin{bmatrix} p_{12} \\ p_{22} \\ \vdots \\ p_{p2} \end{bmatrix} \begin{bmatrix} p_{12} & p_{22} & \cdots & p_{p2} \end{bmatrix} + \cdots + \lambda_p \begin{bmatrix} p_{1p} \\ p_{2p} \\ \vdots \\ p_{pp} \end{bmatrix} \begin{bmatrix} p_{1p} & p_{2p} & \cdots & p_{pp} \end{bmatrix}$$

- 対称行列なので固有ベクトルは直交する
- 正定値行列なので正定値な平方根行列が唯一存在

特異値分解 (SVD, Singular Value Decomposition)

任意の行列 A は積形 U D V’ に分解できる。



$$\mathrm{rank}(A) = r$$

U, V: 直交行列

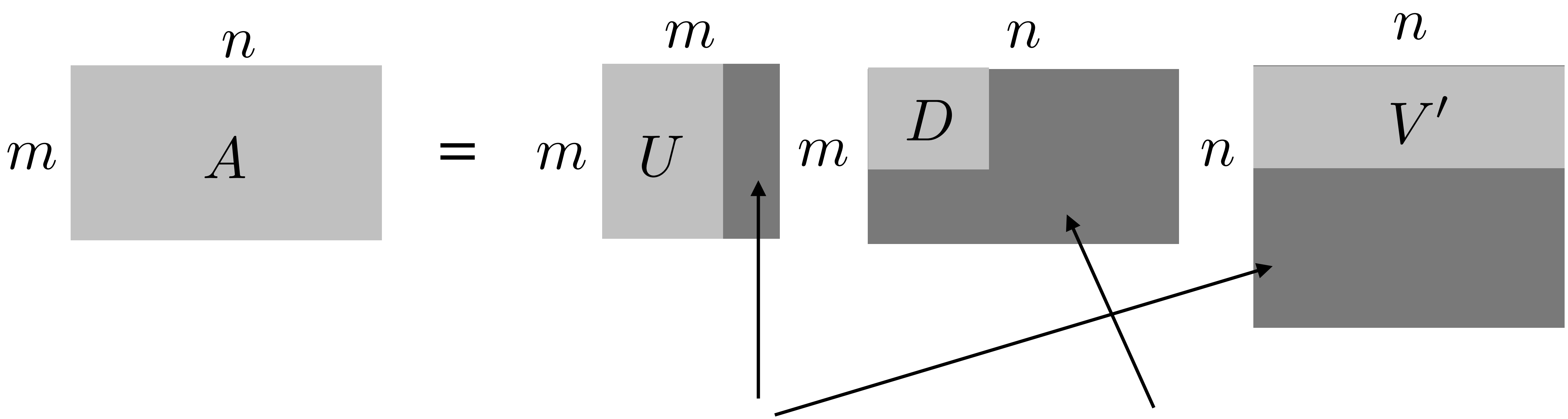
つまり、列ベクトルは正規直交

$$U'U = V'V = I$$

D: 対角行列 (全成分は正)

拡張特異値分解 (利便性のため)

数値計算パッケージだとこう返ってくる場合が多いかも？



正規直交基底を増やす

ここは全部ゼロ

ここがゼロなので
元の特異値分解からこれはすぐ言える。

例その１：vs 非正方行列

※表示上、有効数字3桁まで

$$
\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 & 6 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix}
$$

$$
= \begin{bmatrix} -0.214 & -0.674 \\ -0.953 & 0.303 \\ 0.214 & 0.674 \end{bmatrix}
\begin{bmatrix} 9.95 & 0 \\ 0 & 1.01 \end{bmatrix}
\begin{bmatrix} -0.235 & -0.330 & -0.426 & -0.522 & -0.618 \\ -0.738 & -0.437 & -0.135 & 0.166 & 0.467 \end{bmatrix}
$$

正規直交
(各大きさ1,内積0)

対角成分は
常に正の値

正規直交
(各大きさ1,内積0)

例その２：vs 固有値分解 (非対称正方行列)

※表示上、有効数字で打ち切り

特異値分解

$$\begin{pmatrix} 0.1 & 0.2 & -0.1 \\ -0.2 & 1.2 & 0.3 \\ 0.4 & 0.5 & 1. \end{pmatrix} = \begin{pmatrix} 0.0739 & 0.152 & -0.986 \\ 0.742 & 0.652 & 0.156 \\ 0.667 & -0.742 & -0.0643 \end{pmatrix} \begin{pmatrix} 1.53 & 0. & 0. \\ 0. & 0.825 & 0. \\ 0. & 0. & 0.18 \end{pmatrix} \begin{pmatrix} 0.0825 & 0.812 & 0.578 \\ -0.5 & 0.536 & -0.681 \\ -0.862 & -0.233 & 0.45 \end{pmatrix}$$

正

固有値分解

$$\begin{pmatrix} 0.1 & 0.2 & -0.1 \\ -0.2 & 1.2 & 0.3 \\ 0.4 & 0.5 & 1. \end{pmatrix} = \begin{pmatrix} 0.0466 & 0.432 & -0.664 \\ 0.689 & 0.451 & -0.368 \\ 0.724 & -0.781 & 0.651 \end{pmatrix} \begin{pmatrix} 1.5 & 0. & 0. \\ 0. & 0.49 & 0. \\ 0. & 0. & 0.309 \end{pmatrix} \begin{pmatrix} 0.0466 & 0.432 & -0.664 \\ 0.689 & 0.451 & -0.368 \\ 0.724 & -0.781 & 0.651 \end{pmatrix}^{-1}$$

固有ベクトルが直交しないので逆行列

正とは限らないし一般には複素数にもなり得る。

例その３：vs 固有値分解 (対称非正定値行列)

※表示上、有効数字3桁まで

特異値分解

$$\begin{pmatrix} 0.1 & 0.2 & -0.1 \\ 0.2 & -1.2 & 0.5 \\ -0.1 & 0.5 & 1. \end{pmatrix} = \begin{pmatrix} 0.149 & -0.0558 & 0.987 \\ -0.966 & 0.206 & 0.157 \\ 0.213 & 0.977 & 0.0231 \end{pmatrix} \begin{pmatrix} 1.34 & 0. & 0. \\ 0. & 1.11 & 0. \\ 0. & 0. & 0.13 \end{pmatrix} \begin{pmatrix} -0.149 & 0.966 & -0.213 \\ -0.0558 & 0.206 & 0.977 \\ 0.987 & 0.157 & 0.0231 \end{pmatrix}$$

絶対値は等しい

固有値分解

$$\begin{pmatrix} 0.1 & 0.2 & -0.1 \\ 0.2 & -1.2 & 0.5 \\ -0.1 & 0.5 & 1. \end{pmatrix} = \begin{pmatrix} 0.149 & -0.0558 & 0.987 \\ -0.966 & 0.206 & 0.157 \\ 0.213 & 0.977 & 0.0231 \end{pmatrix} \begin{pmatrix} -1.34 & 0. & 0. \\ 0. & 1.11 & 0. \\ 0. & 0. & 0.13 \end{pmatrix} \begin{pmatrix} 0.149 & -0.966 & 0.213 \\ -0.0558 & 0.206 & 0.977 \\ 0.987 & 0.157 & 0.0231 \end{pmatrix}$$

固有ベクトルが直交するので、転置行列

正とは限らないが必ず実数にはなる。

例その４：vs 固有値分解 (正定値行列)

※表示上、有効数字3桁まで

特異値分解

$$\begin{pmatrix} 0.1 & 0.2 & -0.1 \\ 0.2 & 1.2 & 0.5 \\ -0.1 & 0.5 & 1. \end{pmatrix} = \begin{pmatrix} -0.0618 & -0.336 & 0.94 \\ -0.78 & -0.572 & -0.256 \\ -0.623 & 0.748 & 0.227 \end{pmatrix} \begin{pmatrix} 1.62 & 0. & 0. \\ 0. & 0.663 & 0. \\ 0. & 0. & 0.0215 \end{pmatrix} \begin{pmatrix} -0.0618 & -0.78 & -0.623 \\ -0.336 & -0.572 & 0.748 \\ 0.94 & -0.256 & 0.227 \end{pmatrix}$$

全く同じ！！

固有値分解

$$\begin{pmatrix} 0.1 & 0.2 & -0.1 \\ 0.2 & 1.2 & 0.5 \\ -0.1 & 0.5 & 1. \end{pmatrix} = \begin{pmatrix} -0.0618 & -0.336 & 0.94 \\ -0.78 & -0.572 & -0.256 \\ -0.623 & 0.748 & 0.227 \end{pmatrix} \begin{pmatrix} 1.62 & 0. & 0. \\ 0. & 0.663 & 0. \\ 0. & 0. & 0.0215 \end{pmatrix} \begin{pmatrix} -0.0618 & -0.78 & -0.623 \\ -0.336 & -0.572 & 0.748 \\ 0.94 & -0.256 & 0.227 \end{pmatrix}$$

前回のスライド参照

$$= \begin{pmatrix} -0.0785 & -0.274 & 0.138 \\ -0.991 & -0.466 & -0.0375 \\ -0.792 & 0.609 & 0.0332 \end{pmatrix} \begin{pmatrix} -0.0785 & -0.991 & -0.792 \\ -0.274 & -0.466 & 0.609 \\ 0.138 & -0.0375 & 0.0332 \end{pmatrix}$$

P’P　P:非対称

$$= \begin{pmatrix} 0.226 & 0.182 & -0.125 \\ 0.182 & 1.05 & 0.261 \\ -0.125 & 0.261 & 0.957 \end{pmatrix} \begin{pmatrix} 0.226 & 0.182 & -0.125 \\ 0.182 & 1.05 & 0.261 \\ -0.125 & 0.261 & 0.957 \end{pmatrix}$$

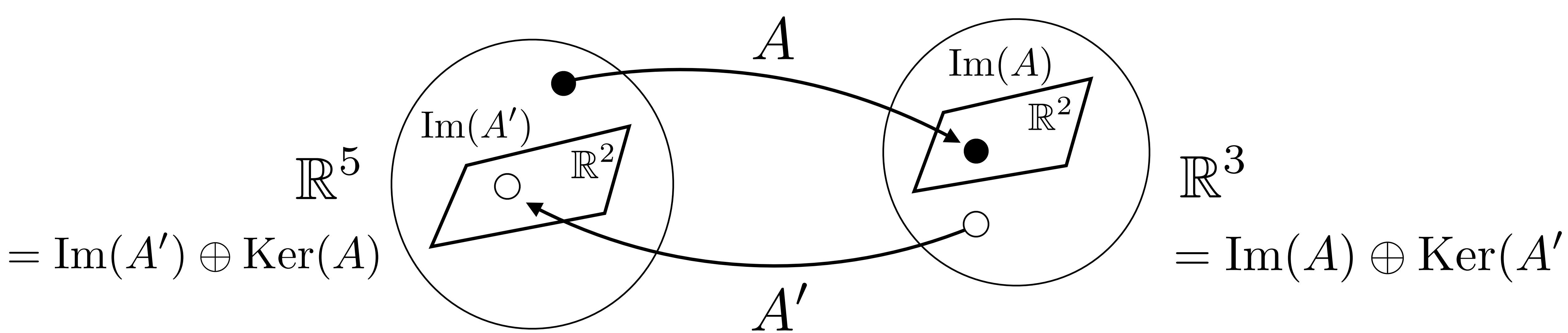
S S　S:対称

(平方根行列)

復習) 線形写像の表現としての行列

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 & 6 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}$$

この行列は5次元のベクトルを3次元のベクトルに移す
ある線形写像を表している。(rank=2)



固有値 vs 特異値

固有値

$$\begin{bmatrix} 0.1 & 0.2 & -0.1 \\ 0.2 & 1.2 & 0.5 \\ -0.1 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} -0.336 \\ -0.572 \\ 0.748 \end{bmatrix} = 0.663 \times \begin{bmatrix} -0.336 \\ -0.572 \\ 0.748 \end{bmatrix}$$

同じ

特異値

$$\begin{bmatrix} 0.1 & 0.2 & -0.1 & 0.3 \\ 0.2 & 1.2 & 0.5 & 1.1 \\ -0.1 & 0.5 & 1 & 3.1 \end{bmatrix} \begin{bmatrix} -0.0001 \\ 0.267 \\ 0.307 \\ 0.913 \end{bmatrix} = 3.6 \times \begin{bmatrix} 0.082 \\ 0.410 \\ 0.908 \end{bmatrix}$$

異なる

!?

特異値

$$\begin{bmatrix} 0.1 & 0.2 & -0.1 & 0.3 \\ 0.2 & 1.2 & 0.5 & 1.1 \\ -0.1 & 0.5 & 1 & 3.1 \end{bmatrix} \begin{bmatrix} -0.0001 \\ 0.267 \\ 0.307 \\ 0.913 \end{bmatrix} = 3.6 \times \begin{bmatrix} 0.082 \\ 0.410 \\ 0.908 \end{bmatrix}$$

異なる

異なって良ければ、定数を任意にくくり出せるので
なんだってアリなんじゃ…と感じたそこのアナタ。

ナイス着眼点！

２つ作るところが特異値分解の最大のミソ。

$$\begin{bmatrix} 0.1 & 0.2 & -0.1 & 0.3 \\ 0.2 & 1.2 & 0.5 & 1.1 \\ -0.1 & 0.5 & 1 & 3.1 \end{bmatrix} \begin{bmatrix} -0.0001 \\ 0.267 \\ 0.307 \\ 0.913 \end{bmatrix} = 3.6 \times \begin{bmatrix} 0.082 \\ 0.410 \\ 0.908 \end{bmatrix}$$

転置

$$\begin{bmatrix} 0.1 & 0.2 & -0.1 \\ 0.2 & 1.2 & 0.5 \\ -0.1 & 0.5 & 1 \\ 0.3 & 1.1 & 3.1 \end{bmatrix} \begin{bmatrix} 0.082 \\ 0.410 \\ 0.908 \end{bmatrix} = 3.6 \times \begin{bmatrix} -0.0001 \\ 0.267 \\ 0.307 \\ 0.913 \end{bmatrix}$$

固有値 vs 特異値

A :正方対称

$$A\boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1$$
$$A\boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_2$$
$$\vdots$$
$$A\boldsymbol{v}_n = \lambda_n \boldsymbol{v}_n$$

$$+ \quad \boldsymbol{v}_i' \boldsymbol{v}_i = 1$$
$$\boldsymbol{v}_i' \boldsymbol{v}_j = 0 \quad (i \neq j)$$

行列形に
並べると

スペクトル分解 (直交行列による対角化)

$$A \underbrace{\begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \\ | & | & & | \end{bmatrix}}_{\text{正規直交}} = \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \lambda_v \end{bmatrix}$$

固有値 vs 特異値

A:任意(非正方も)

行列形に
並べると

$$A\boldsymbol{v}_1 = \lambda_1 \boldsymbol{u}_1$$
$$A\boldsymbol{v}_2 = \lambda_2 \boldsymbol{u}_2$$
$$\vdots$$
$$A\boldsymbol{v}_r = \lambda_r \boldsymbol{u}_r$$

$+$

$$\boldsymbol{v}_i'\boldsymbol{v}_i = 1$$
$$\boldsymbol{v}_i'\boldsymbol{v}_j = 0 \quad (i \neq j)$$
$$\boldsymbol{u}_i'\boldsymbol{u}_i = 1$$
$$\boldsymbol{u}_i'\boldsymbol{u}_j = 0 \quad (i \neq j)$$

$$\mathrm{rank}(A) = r$$

特異値分解

$$A \underbrace{\begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \\ | & | & & | \end{bmatrix}}_{\text{正規直交}} = \underbrace{\begin{bmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r \\ | & | & & | \end{bmatrix}}_{\text{正規直交}} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \lambda_r \end{bmatrix}$$

復習) 固有値分解の直感的意味

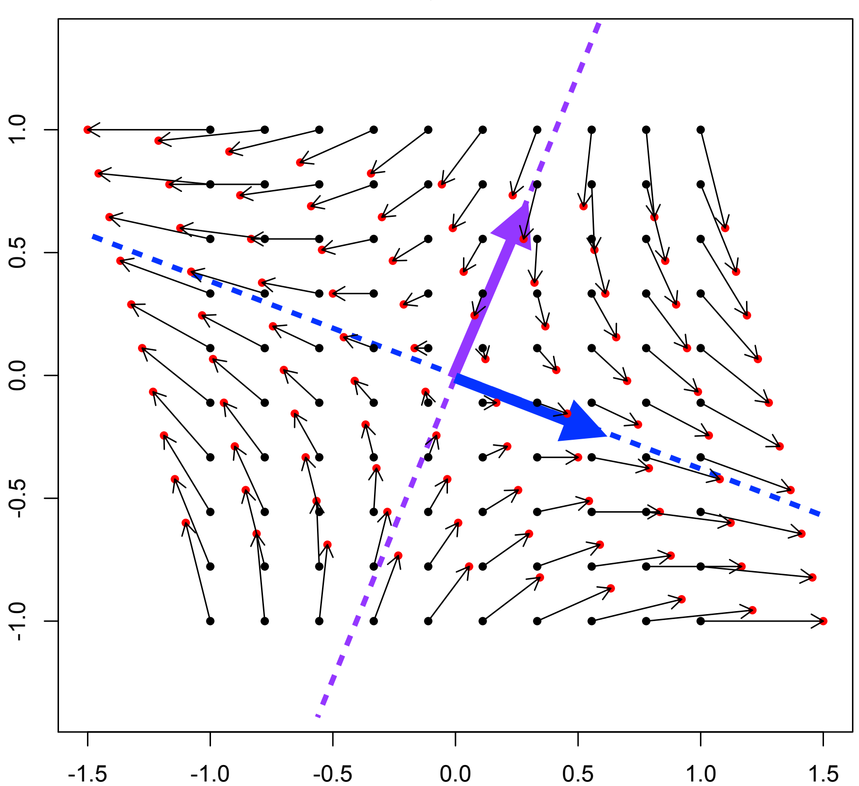
方向が変わらない

$$A\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda_1 \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

$$A\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \lambda_2 \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$A = \begin{bmatrix} v_1 & u_1 \\ v_2 & u_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} v_1 & u_1 \\ v_2 & u_2 \end{bmatrix}^{-1}$$

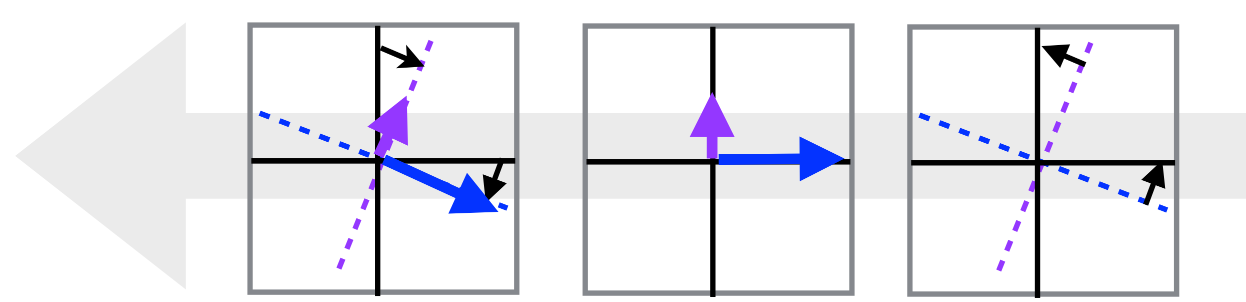
$$A\begin{bmatrix} v_1 & u_1 \\ v_2 & u_2 \end{bmatrix} = \begin{bmatrix} v_1 & u_1 \\ v_2 & u_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} v_1 & u_1 \\ v_2 & u_2 \end{bmatrix}}_{\text{基底変換}} \underbrace{\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}}_{\text{倍率}} \underbrace{\begin{bmatrix} v_1 & u_1 \\ v_2 & u_2 \end{bmatrix}^{-1}}_{\text{逆基底変換}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

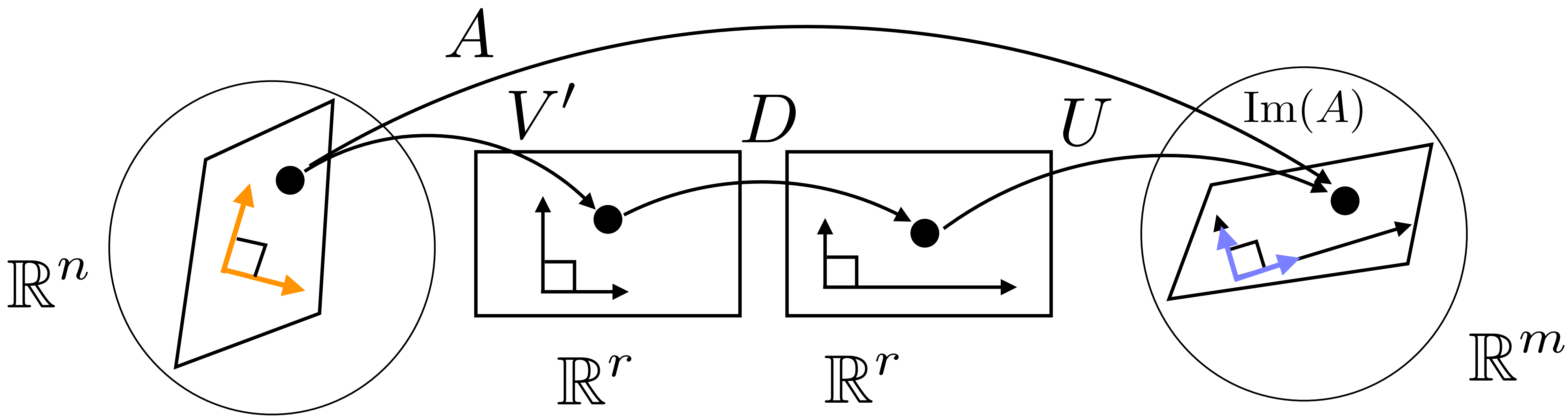
基底変換: $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \to \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

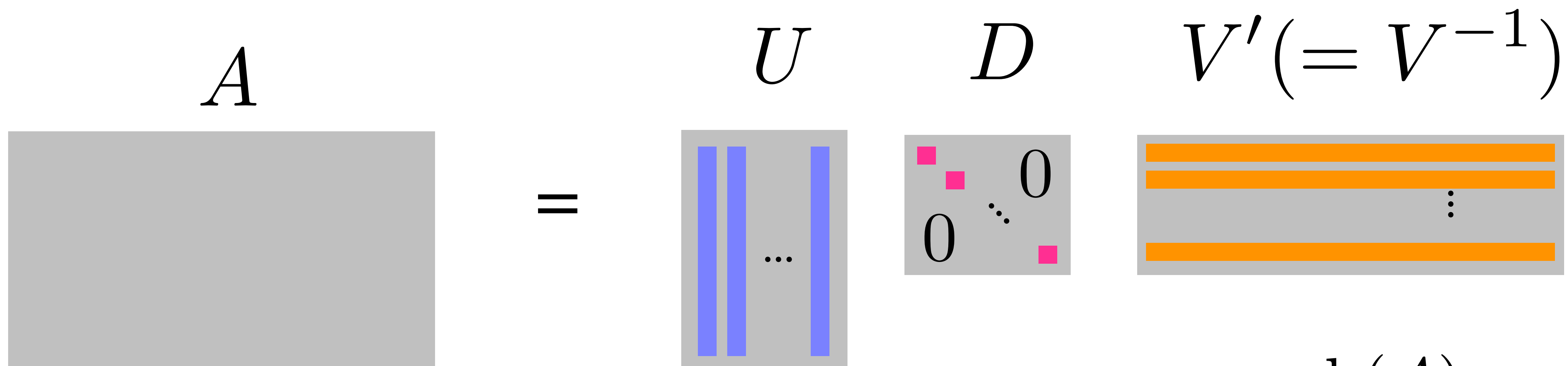
逆基底変換: $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \to \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$



特異値分解の直感的意味

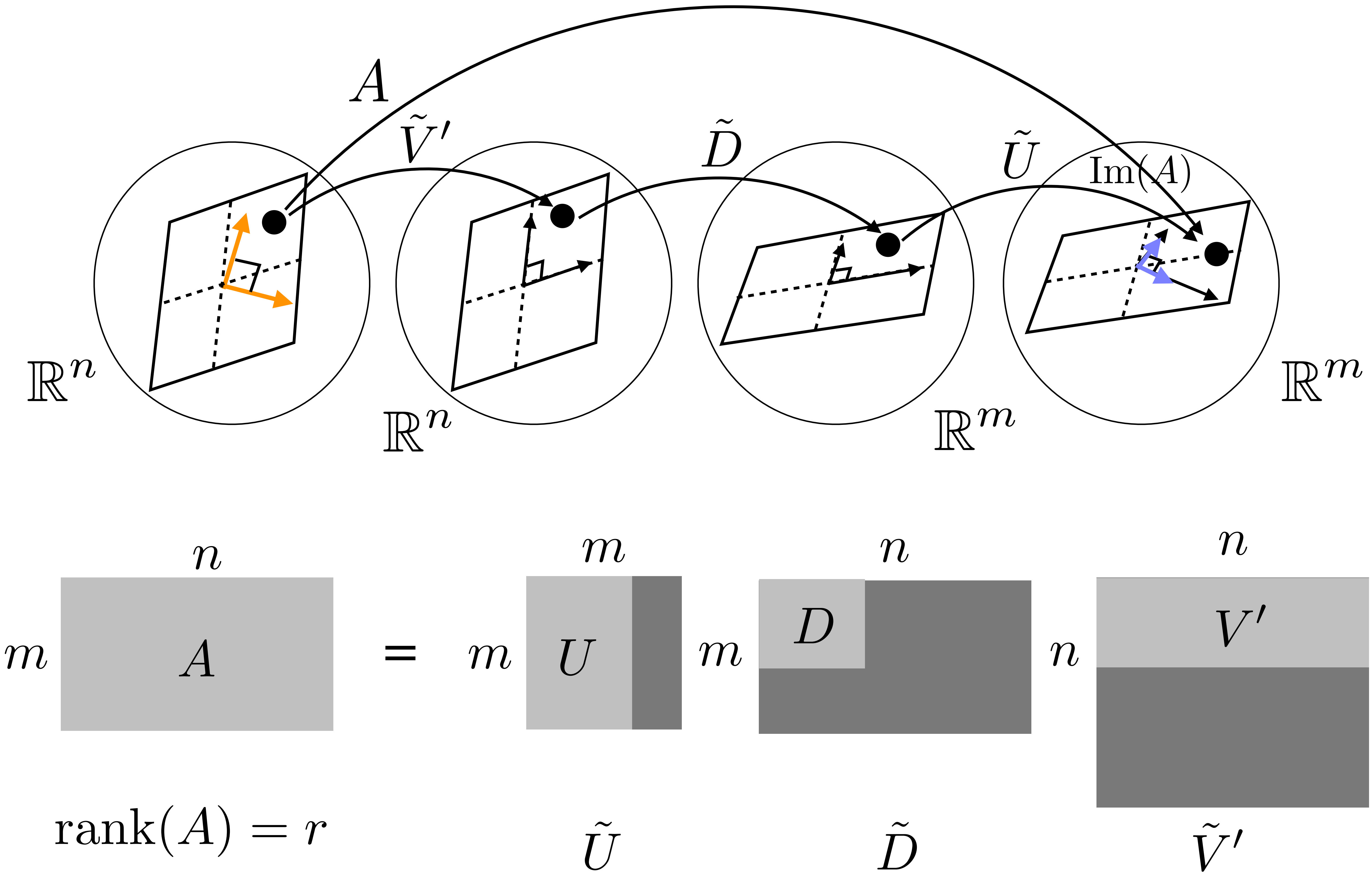
高次に埋め込まれてはいるが、結局 r 次元空間での固有値展開





$$\mathrm{rank}(A) = r$$

拡張特異値分解の直感的意味



特異値の求め方 (1/2)

特異値分解 $A = UDV'$ $U'U = V'V = I$

以下の2つの行列を考える。

$$\begin{aligned} A'A &= (UDV')'(UDV') = (V')'D'U'UDV' \\ &= VD'DV' \end{aligned}$$

$$\begin{aligned} AA' &= (UDV')(UDV')' = UDV'(V')'D'U' \\ &= UDD'U' \end{aligned}$$

特異値の求め方 (2/2)

特異値分解　$A = UDV'$　　$U'U = V'V = I$

固有値分解を $A'A = Q_1 \Lambda_1 Q_1'$ とすると

$A'A = VD'DV'$ より $V = Q_1, D = \Lambda_1^{1/2}$

固有値分解を $A'A = Q_2 \Lambda_2 Q_2'$ とすると

$AA' = UDD'U'$ より $U = Q_2, D = \Lambda_2^{1/2}$

※上の２つを逆に見れば特異値分解の成立が示せる

特異値分解は任意の行列で可能！

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{6}} \end{pmatrix}$$

固有値分解ができない行列でもOK!

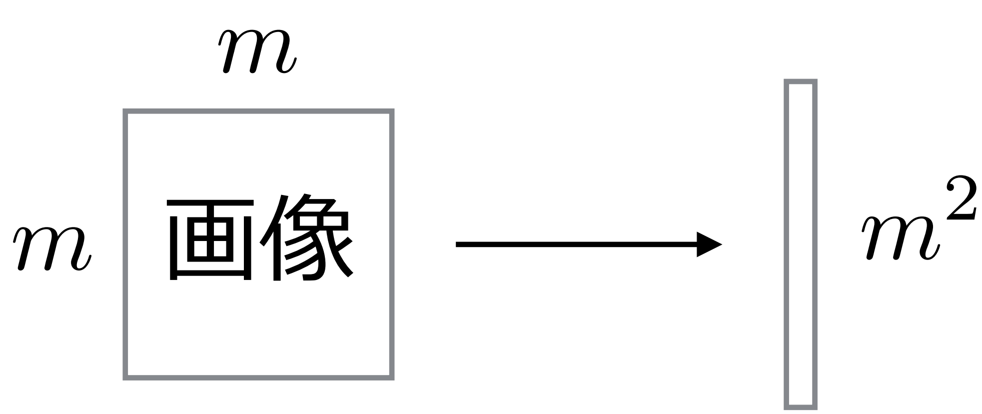
$$\begin{pmatrix} 6 & 12 & 19 \\ -9 & -20 & -33 \\ 4 & 9 & 15 \end{pmatrix} = \begin{pmatrix} 0.471 & -0.875 & 0.11 \\ -0.803 & -0.375 & 0.463 \\ 0.364 & 0.307 & 0.88 \end{pmatrix} \begin{pmatrix} 49.3 & 0. & 0. \\ 0. & 0.775 & 0. \\ 0. & 0. & 0.0262 \end{pmatrix} \begin{pmatrix} 0.233 & 0.507 & 0.83 \\ -0.842 & -0.321 & 0.433 \\ 0.486 & -0.8 & 0.352 \end{pmatrix}$$

固有値 -1,1,1

固有値 1に対する固有ベクトルが2つ作れない

⇨ 対角化できない / 固有値分解できない

特異値分解(SVD)は実用上のメリットが！



画素値のベクトル

m=512の50画像で
分散共分散行列は
512^2 = 687億個の要素

平均画像を引いた画像 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{50}$ に対して

$$\boldsymbol{M} = \sum_{i=1}^{50} \boldsymbol{x}_i \boldsymbol{x}_i'$$ の固有値・固有ベクトルを求めたい

$\boldsymbol{X} = (\boldsymbol{x}_1 \boldsymbol{x}_2 \cdots \boldsymbol{x}_{50})$ として特異値分解！

$\boldsymbol{X}'\boldsymbol{X}$ と $\boldsymbol{X}\boldsymbol{X}'$ の固有値・固有ベクトルが分かる

0固有値を無視して実質ランクr分のみの計算(効率が良い！)

数量化Ⅲ類

表 10.1　児童の得意科目のデータ（〇印が得意科目）

児童 No.	国語	社会	算数	理科	音楽	図工	体育
1	〇			〇		〇	
2			〇		〇	〇	
3	〇						〇
4	〇	〇	〇	〇			
5		〇					〇
6				〇	〇	〇	
7			〇	〇	〇		
8	〇	〇		〇			〇
9			〇			〇	〇
10	〇	〇	〇	〇			

表 10.2　表 10.1 のデータの行と列の並び替え

児童 No.	音楽	図工	算数	理科	国語	社会	体育
2	〇	〇	〇				
6	〇	〇		〇			
7	〇		〇	〇			
1		〇		〇	〇		
9		〇	〇				〇
4			〇	〇	〇	〇	
10			〇	〇	〇	〇	
8				〇	〇	〇	〇
3					〇		〇
5						〇	〇

各列と各行が似るように並び替え

低次元近似して可視化

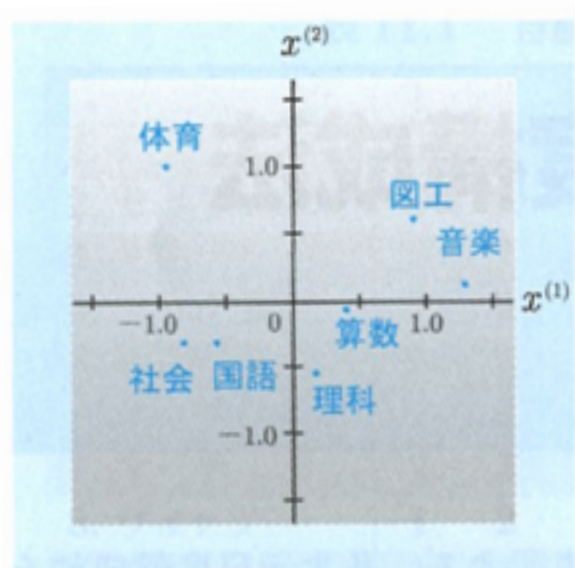


図 10.1　変数スコア散布図

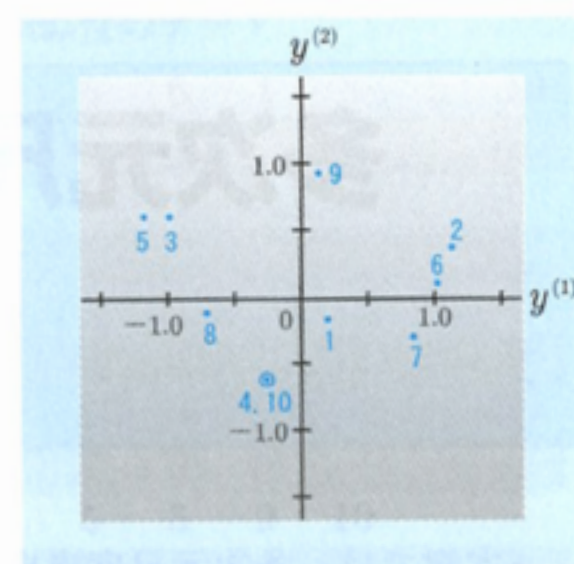


図 10.2　サンプルスコア散布図

数量化Ⅲ類

行、列の各々に数量 x_i, y_j を割り付ける。

表 10.1　児童の得意科目のデータ（○印が得意科目）

児童 No.	国語	社会	算数	理科	音楽	図工	体育	
1	○			○		○		y_1
2			○		○	○		y_2
3	○						○	y_3
4	○	○	○	○				y_4
5		○					○	y_5
6				○	○	○		y_6
7			○	○	○			y_7
8	○	○		○			○	y_8
9			○			○	○	y_9
10	○	○	○	○				y_{10}
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	

例えば、数量 x_i, x_j は対応する列が似ていたら近い値にしたい。 y_i, y_j も似ている行は近い値にしたい。

数量化Ⅲ類

例えば、数量 x_i, x_j は対応する列が似ていたら近い値にしたい。 y_i, y_j も似ている行は近い値にしたい。

表 10.1　児童の得意科目のデータ（○印が得意科目）

児童 No.	国語	社会	算数	理科	音楽	図工	体育	
1	○			○		○		y_1
2			○		○	○		y_2
3	○						○	y_3
4	○	○	○	○				y_4
5		○					○	y_5
6				○	○	○		y_6
7			○	○	○			y_7
8	○	○		○			○	y_8
9			○			○	○	y_9
10	○	○	○	○				y_{10}
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	

○の所だけ

相関係数→max

x_1	y_1
x_1	y_3
x_1	y_4
x_1	y_8
x_1	y_{10}
x_2	y_4
x_2	y_5
x_2	y_8
⋮	
x_6	y_9
x_7	y_3
x_7	y_5
x_7	y_8
x_7	y_9

数量化Ⅲ類で解くべき問題

$$\max_{\boldsymbol{x},\boldsymbol{y}} \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad \text{subject to} \quad \begin{array}{l} \bar{x} = \bar{y} = 0 \\ \sigma_x^2 = \sigma_y^2 = 1 \end{array}$$

相関係数 in [-1,1]

$$\sigma_{xy} = \sum_i \sum_j \delta((i,j) = \bigcirc)\,(x_i - \bar{x})(y_j - \bar{y})$$

表のi行j列が○なら1 else 0

$$\sigma_x^2 = \sum_i \sum_j \delta((i,j) = \bigcirc)\,(x_i - \bar{x})^2$$

$$\sigma_y^2 = \sum_i \sum_j \delta((i,j) = \bigcirc)\,(y_j - \bar{y})^2$$

一意に定めるため
変数に制約を置く
(一般性を失わない)

数量化Ⅲ類で解くべき問題 (行列表記)

$$\max_{\boldsymbol{x},\boldsymbol{y}} \boldsymbol{y}' M \boldsymbol{x} \ \text{ subject to } \ \boldsymbol{x}' D_{\mathrm{col}} \boldsymbol{x} = 1, \boldsymbol{y}' D_{\mathrm{row}} \boldsymbol{y} = 1$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \end{bmatrix} \quad M = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$D_{\mathrm{row}} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

各行の総和

$$D_{\mathrm{col}} = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

各列の総和

$$\max_{\boldsymbol{x},\boldsymbol{y}} \boldsymbol{y}' M \boldsymbol{x} \ \text{subject to} \ \boldsymbol{x}' D_{\mathrm{col}}\, \boldsymbol{x} = 1, \boldsymbol{y}' D_{\mathrm{row}}\, \boldsymbol{y} = 1$$

$\boldsymbol{v} := D_{\mathrm{col}}^{1/2} \boldsymbol{x}$
$\boldsymbol{w} := D_{\mathrm{row}}^{1/2} \boldsymbol{y}$
と置くと
$\boldsymbol{v}'\boldsymbol{v} = \|\boldsymbol{v}\|^2 = 1$
$\boldsymbol{w}'\boldsymbol{w} = \|\boldsymbol{w}\|^2 = 1$
と形がよくなる

目的関数は $\boldsymbol{y}' M \boldsymbol{x} = \boldsymbol{w}' D_{\mathrm{row}}^{-1/2} M D_{\mathrm{col}}^{-1/2} \boldsymbol{v}$ となるので

$\tilde{M} := D_{\mathrm{row}}^{-1/2} M D_{\mathrm{col}}^{-1/2}$ と置けば結局以下に変形できる。

$$\max_{\boldsymbol{v},\boldsymbol{w}} \boldsymbol{w}' \tilde{M} \boldsymbol{v} \ \text{subject to} \ \|\boldsymbol{v}\|^2 = 1, \|\boldsymbol{w}\|^2 = 1$$

注意：$D_{\mathrm{row}}, D_{\mathrm{col}}$ は対角行列。対角行列 D について $D' = D$

$$D = \begin{bmatrix} b_{11} & 0 & 0 & \cdots & 0 \\ 0 & b_{22} & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & b_{nn} \end{bmatrix} \quad D^{1/2} = \begin{bmatrix} \sqrt{b_{11}} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{b_{22}} & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{b_{nn}} \end{bmatrix} \quad D^{-1/2} = \begin{bmatrix} 1/\sqrt{b_{11}} & 0 & 0 & \cdots & 0 \\ 0 & 1/\sqrt{b_{22}} & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/\sqrt{b_{nn}} \end{bmatrix}$$

双線型形式の最大化！

$$\max_{\boldsymbol{v},\boldsymbol{w}} \ \boldsymbol{w}'\tilde{M}\boldsymbol{v} \ \text{ subject to } \ \|\boldsymbol{v}\|^2 = 1, \|\boldsymbol{w}\|^2 = 1$$

$\Longleftrightarrow$ Lagrangeの未定乗数法より

$$L(\boldsymbol{v},\boldsymbol{w}) := \boldsymbol{w}'\tilde{M}\boldsymbol{v} - \frac{\lambda}{2}(\boldsymbol{v}'\boldsymbol{v} - 1) - \frac{\mu}{2}(\boldsymbol{w}'\boldsymbol{w} - 1)$$

$$\frac{\partial L(\boldsymbol{v},\boldsymbol{w})}{\partial \boldsymbol{v}} = \tilde{M}'\boldsymbol{w} - \lambda\boldsymbol{v} = \ 0 \quad \Longrightarrow \quad \tilde{M}'\boldsymbol{w} = \lambda\boldsymbol{v}$$

$$\frac{\partial L(\boldsymbol{v},\boldsymbol{w})}{\partial \boldsymbol{w}} = \tilde{M}\boldsymbol{v} - \mu\boldsymbol{w} = \ 0 \quad \Longrightarrow \quad \tilde{M}\boldsymbol{v} = \mu\boldsymbol{w}$$

①目的関数の値は $\boldsymbol{w}'\tilde{M}\boldsymbol{v} = \lambda\boldsymbol{v}'\boldsymbol{v} = \mu\boldsymbol{w}'\boldsymbol{w} = \lambda = \mu$

②このとき $\begin{cases} \tilde{M}\boldsymbol{v} = \lambda\,\boldsymbol{w} \\ \tilde{M}'\boldsymbol{w} = \lambda\,\boldsymbol{v} \end{cases} \Longleftrightarrow \begin{cases} \tilde{M}'\tilde{M}\boldsymbol{v} = \lambda^2\boldsymbol{v} \\ \tilde{M}\tilde{M}'\boldsymbol{w} = \lambda^2\boldsymbol{w} \end{cases}$

①目的関数の値は　$\boldsymbol{w}'\tilde{M}\boldsymbol{v} = \lambda\boldsymbol{v}'\boldsymbol{v} = \mu\boldsymbol{w}'\boldsymbol{w} = \lambda = \mu$

②このとき $\begin{cases} \tilde{M}\boldsymbol{v} = \lambda\boldsymbol{w} \\ \tilde{M}'\boldsymbol{w} = \lambda\boldsymbol{v} \end{cases} \Longleftrightarrow \begin{cases} \tilde{M}'\tilde{M}\boldsymbol{v} = \lambda^2\boldsymbol{v} \\ \tilde{M}\tilde{M}'\boldsymbol{w} = \lambda^2\boldsymbol{w} \end{cases}$

$\boldsymbol{w}'\tilde{M}\boldsymbol{v} = \lambda$ を最大にするには

$\begin{cases} \tilde{M}'\tilde{M}\boldsymbol{v} = \lambda^2\boldsymbol{v} \\ \tilde{M}\tilde{M}'\boldsymbol{w} = \lambda^2\boldsymbol{w} \end{cases}$ の最大固有値 λ^2 とその固有ベクトル

平方根　2乗

$\Longleftrightarrow$ $\tilde{M}$ の特異値分解で最大特異値 λ を求めるのと同じ！

※ただし先ほどの数量化III類の場合、常に最大特異値1でその特異ベクトルは(1,1,1,...,1)の正規化版になり(確認すること)、意味がないので、二番目に大きい特異値とその固有ベクトルを解に

教科書の例で

表 10.1　児童の得意科目のデータ（○印が得意科目）

児童 No.	国語	社会	算数	理科	音楽	図工	体育
1	○			○		○	
2			○		○	○	
3	○						○
4	○	○	○	○			
5		○					○
6				○	○	○	
7			○	○	○		
8	○	○		○			○
9			○			○	○
10	○	○	○	○			

$$\tilde{M} = \begin{pmatrix} \frac{1}{\sqrt{15}} & 0 & 0 & \frac{1}{3\sqrt{2}} & 0 & \frac{1}{2\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{15}} & 0 & \frac{1}{3} & \frac{1}{2\sqrt{3}} & 0 \\ \frac{1}{\sqrt{10}} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{5}} & \frac{1}{4} & \frac{1}{2\sqrt{5}} & \frac{1}{2\sqrt{6}} & 0 & 0 & 0 \\ 0 & \frac{1}{2\sqrt{2}} & 0 & 0 & 0 & 0 & \frac{1}{2\sqrt{2}} \\ 0 & 0 & 0 & \frac{1}{3\sqrt{2}} & \frac{1}{3} & \frac{1}{2\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{15}} & \frac{1}{3\sqrt{2}} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2\sqrt{5}} & \frac{1}{4} & 0 & \frac{1}{2\sqrt{6}} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{\sqrt{15}} & 0 & 0 & \frac{1}{2\sqrt{3}} & \frac{1}{2\sqrt{3}} \\ \frac{1}{2\sqrt{5}} & \frac{1}{4} & \frac{1}{2\sqrt{5}} & \frac{1}{2\sqrt{6}} & 0 & 0 & 0 \end{pmatrix}$$

λ ($\tilde{M}$ の特異値)

1., 0.749173, 0.528494, 0.443623, 0.354096, 0.30191, 0.171557

λ^2 ($\tilde{M}'\tilde{M}$ と $\tilde{M}\tilde{M}'$ の固有値) $\longrightarrow$ p.162 表10.7

1., 0.56126, 0.279306, 0.196802, 0.125384, 0.0911499, 0.0294318

λ^2 特異値分解か固有値分解で求める $\longrightarrow$ p.162 表10.7

1., 0.56126, 0.279306, 0.196802, 0.125384, 0.0911499, 0.0294318

$\lambda_2^2 = 0.56126$, $\lambda_3^2 = 0.279306$

$\boldsymbol{v}_2$ {-0.317, -0.409, 0.215, 0.0912, 0.556, 0.392, -0.463}

$\boldsymbol{w}_2$ {0.0706, 0.473, -0.352, -0.142, -0.412, 0.427, 0.35, -0.361, 0.047, -0.142}

$\boldsymbol{v}_3$ {-0.26, -0.232, -0.0594, -0.468, 0.061, 0.418, 0.691}

$\boldsymbol{w}_3$ {-0.107, 0.238, 0.307, -0.426, 0.307, 0.0582, -0.199, -0.0734, 0.577, -0.426}

$$\boldsymbol{x}_2 = D_{\mathrm{col}}^{-1/2}\boldsymbol{v}_2 \qquad \sqrt{(n-1)\lambda_2^2}\cdot\boldsymbol{x}_2$$
$$\boldsymbol{x}_3 = D_{\mathrm{col}}^{-1/2}\boldsymbol{v}_3 \qquad \sqrt{(n-1)\lambda_3^2}\cdot\boldsymbol{x}_3$$

$\longrightarrow$ p.162 表10.8

$$\boldsymbol{y}_2 = D_{\mathrm{row}}^{-1/2}\boldsymbol{w}_2 \qquad \sqrt{(n-1)\lambda_2^2}\cdot\boldsymbol{y}_2$$
$$\boldsymbol{y}_3 = D_{\mathrm{row}}^{-1/2}\boldsymbol{w}_3 \qquad \sqrt{(n-1)\lambda_3^2}\cdot\boldsymbol{y}_3$$

$\longrightarrow$ p.162 表10.9

表 10.1　児童の得意科目のデータ（○印が得意科目）

児童 No.		1 国語	2 社会	3 算数	4 理科	5 音楽	6 図工	7 体育
1	1	○			○		○	
2	2			○		○	○	
3	3	○						○
4	4	○	○	○	○			
5	5		○					○
6	6				○	○	○	
7	7			○	○	○		
8	8	○	○		○			○
9	9			○			○	○
10	10	○	○	○	○			

表 10.2　表 10.1 のデータの行と列の並び替え

児童 No.	音楽	図工	算数	理科	国語	社会	体育
2	○	○	○				
6	○	○		○			
7	○		○	○			
1		○		○	○		
9		○	○				○
4			○	○	○	○	
10			○	○	○	○	
8				○	○	○	○
3					○		○
5						○	○

大きい順に

v_2 {-0.317, -0.409, 0.215, 0.0912, 0.556, 0.392, -0.463}

5, 6, 3, 4, 1, 2, 7

w_2 {0.0706, 0.473, -0.352, -0.142, -0.412, 0.427, 0.35, -0.361, 0.047, -0.142}

2, 6, 7, 1, 9, 10, 4, 3, 8, 5

数量化Ⅲ類の計算 (まとめ)

$$\max_{\boldsymbol{x},\boldsymbol{y}} \boldsymbol{y}' M \boldsymbol{x} \ \text{ subject to } \ \boldsymbol{x}' D_{\text{col}}\, \boldsymbol{x} = 1, \boldsymbol{y}' D_{\text{row}}\, \boldsymbol{y} = 1$$

$$\iff \quad \max_{\boldsymbol{v},\boldsymbol{w}} \ \boldsymbol{w}' \tilde{M} \boldsymbol{v} \ \text{ subject to } \ \|\boldsymbol{v}\|^2 = 1, \|\boldsymbol{w}\|^2 = 1$$

この解は $\tilde{M}$ の2番目に大きい特異値 λ_2 と対応する $\boldsymbol{v}_2, \boldsymbol{w}_2$

$\iff$ $\tilde{M}'\tilde{M}$ の2番目に大きい固有値 λ_2^2 、固有ベクトル $\boldsymbol{v}_2$

$\tilde{M}\tilde{M}'$ の2番目に大きい固有値 λ_2^2 、固有ベクトル $\boldsymbol{w}_2$

このとき

$$\boldsymbol{x}_2 = D_{\text{col}}^{-1/2} \boldsymbol{v}_2, \boldsymbol{y}_2 = D_{\text{row}}^{-1/2} \boldsymbol{w}_2, \lambda_2 = \boldsymbol{y}_2' M \boldsymbol{x}_2 = \boldsymbol{w}_2' \tilde{M} \boldsymbol{v}_2$$

※可視化や低次元近似が目的のときは第3特異値以降も計算