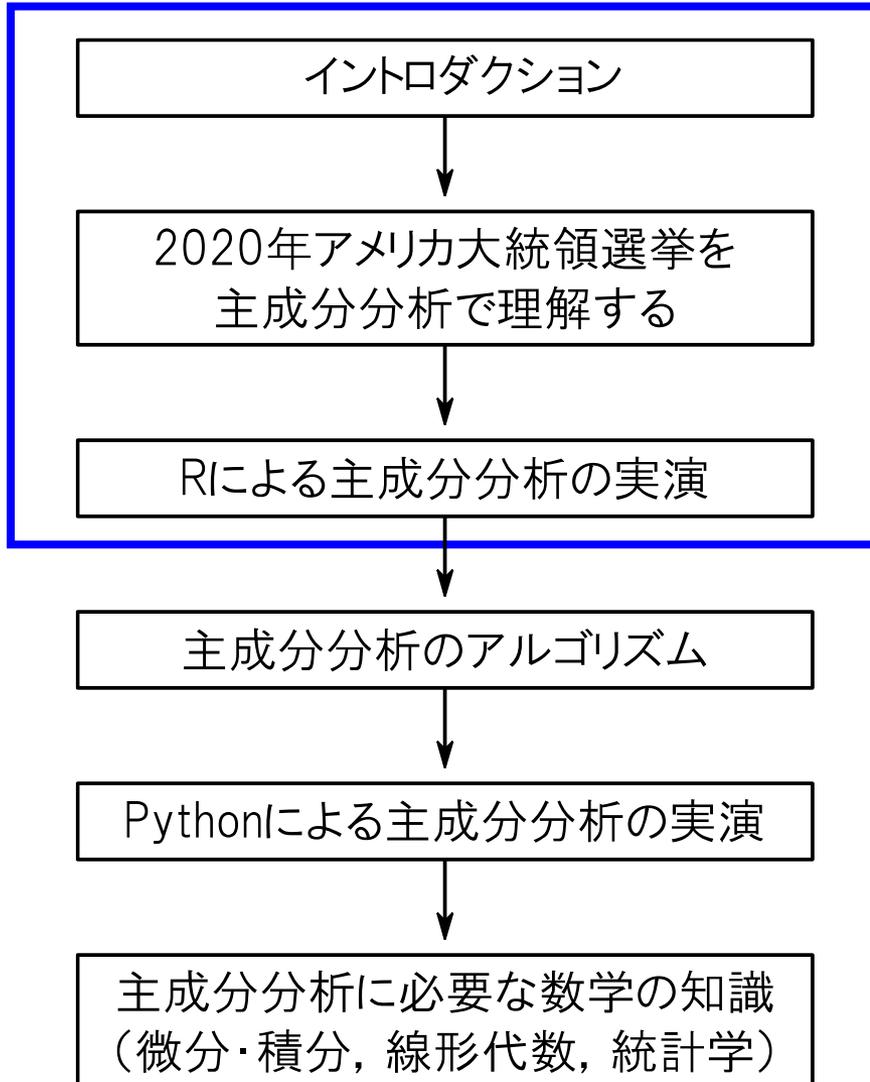


統計数学と主成分分析による ビッグ・データの可視化とパターン認識

Part 1: 入門編

リニア・テック 別府 伸耕

linear tec : Nobuyasu Beppu



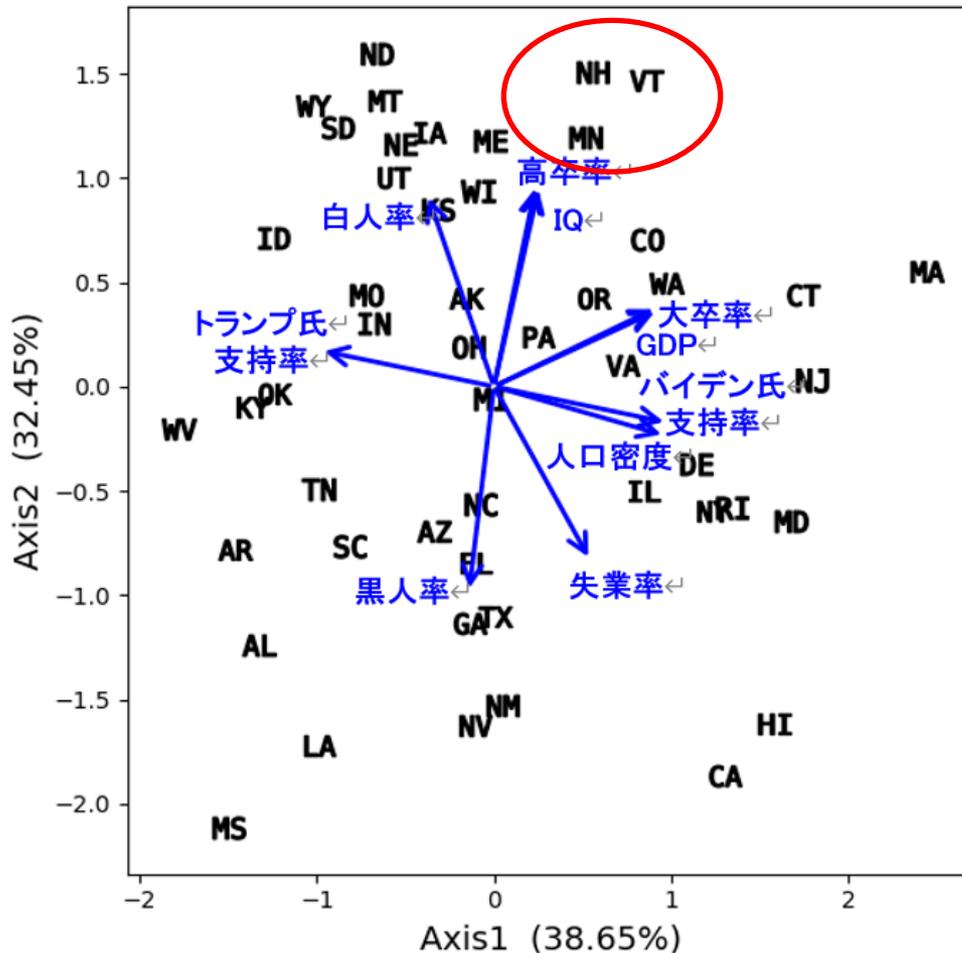
「Part 1 入門編」では
この部分を扱います。

- データ処理の本質は「数学」である.
- データ処理を理解できないのは、数学がわからないから.
- 数学がわかれば、あらゆる理工学への応用につながる.

データ処理で価値を生み出すための正攻法は、数学を学ぶことです。
本セミナーでは主成分分析を題材にして、数学を使いこなすお手伝いをします。

主成分分析の結果の解釈(2)

Sample



● 平均IQおよび高卒率が高い州

- NH (New Hampshire)
- VT (Vermont)
- MN (Minnesota)

エレクトロニクス産業,
重工業産業が多い。

Rで元データを読み込む

Sample

(データ格納先の変数) = read.table("読み込むファイル名", sep="区切り文字")

```
R Console
> data = read.table("data_labeled.txt", sep="\t")
> data
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1		Biden	Trump	IQ	PD	GDP	white	black	high school	bachelor	Unemployment
2	AK	39.7	60.3	99	0.5	63971	60.6	2.9	92.4	29	7.2
3	AL	37	63	95.7	37.4	37261	65.5	26.7	85.3	24.5	6.6
4	AR	35.6	64.4	97.5	22.4	36368	72.3	15.2	85.6	22	7.3
5	AZ	50.2	49.8	97.4	24.7	38590	54.7	4.1	86.5	28.4	6.7
6	CA	65.7	34.3	95.5	97.8	58619	37	5.5	82.5	32.6	11
7	CO	56.9	43.1	101.6	21.5	52795	68.2	3.9	91.1	39.4	6.4
8	CT	60.2	39.8	103.1	284.3	64511	66.7	9.9	90.2	38.4	7.8
9	DE	59.6	40.4	100.4	192.9	63664	62.2	21.5	89.3	31	8.2
10	FL	48.3	51.7	98.4	154.6	39543	53.8	15.4	87.6	28.5	7.6
11	GA	50.1	49.9	98	71.3	44723	52.6	31.1	86.3	29.9	6.4
12	HI	65	35	95.6	85.1	51277	21.8	1.6	91.6	32	15.1
13	IA	45.8	54.2	103.2	21.8	50315	85.9	3.3	91.8	27.7	4.7
14	ID	34.1	65.9	101.4	8.3	35466	82	0.6	90.2	26.8	6.1
15	IL	56.6	43.4	99.9	88.1	54091	61.2	14	88.6	33.4	10.2
16	IN	41.8	58.2	101.7	72.6	45317	79.2	9.2	88.3	25.3	6.2
17	KS	42.3	57.7	102.8	13.8	46982	75.9	5.5	90.5	32.3	5.9
18	KY	36.8	63.2	99.4	43.7	38985	84.6	8	85.2	23.2	5.6
19	LA	40.5	59.5	95.3	41.5	43917	58.5	32.1	84.3	23.4	8.1
20	MA	66.8	33.2	104.3	341.2	65545	71.5	7	90.3	42.1	9.6
21	MD	65.6	34.4	99.7	240.5	55404	50.7	29.4	89.8	39	7.2
22	ME	55.1	44.9	103.4	16.8	38921	93.4	1.2	92.1	30.3	6.1
23	MI	51.4	48.6	100.5	68.2	43372	75	13.6	90.2	28.1	8.5
24	MN	53.6	46.4	103.7	27.3	53704	79.9	6.4	92.8	34.8	6
25	MO	42.1	57.9	101	34.5	43317	79.4	11.4	89.2	28.2	4.9
26	MS	39.7	60.3	94.2	24.5	31881	56.6	37.9	83.4	21.3	7.1
27	MT	41.6	58.4	103.4	2.8	39356	86.3	0.4	93	30.7	5.3
28	NC	49.3	50.7	100.2	83.3	44325	63	21.2	86.9	29.9	7.3
29	ND	32.8	67.2	103.8	4.3	62837	84.4	3	92.3	28.9	4.4