

主成分分析

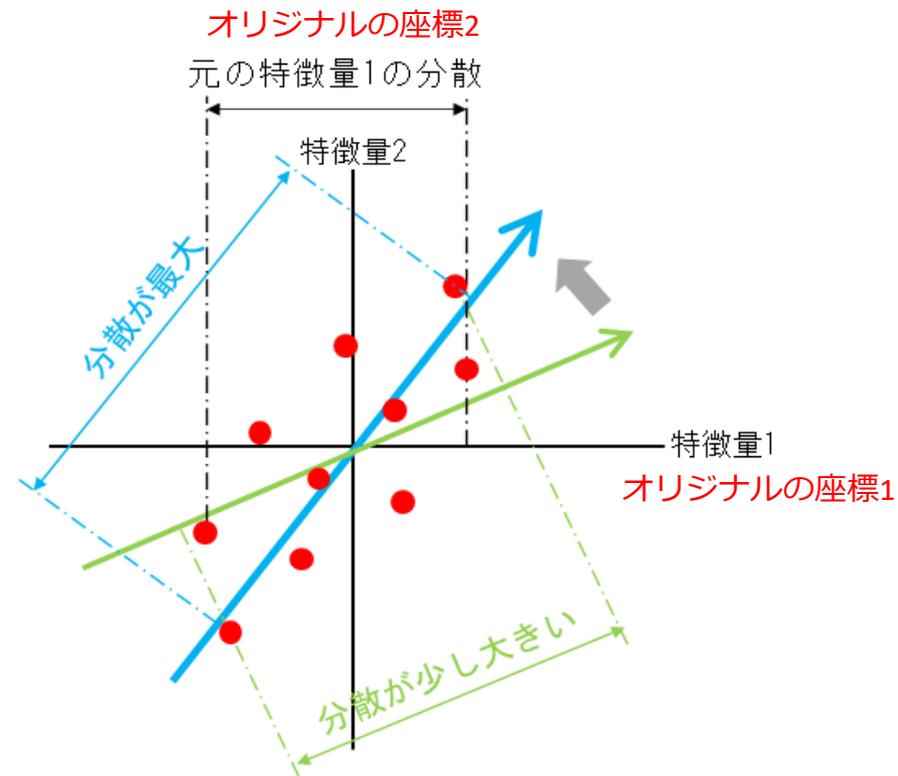
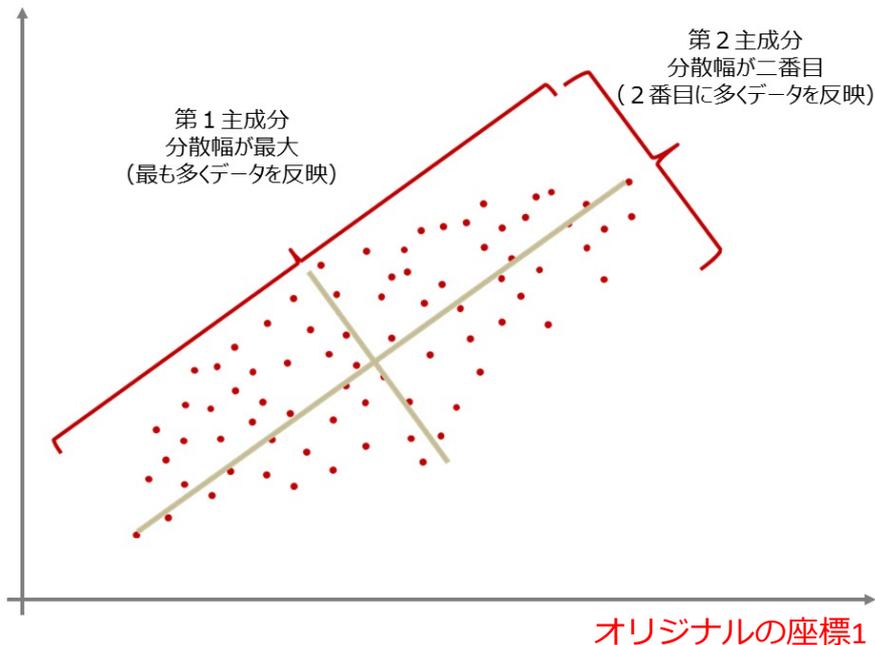
参考

- 〈ゼロからのサイエンス〉 多変量解析がわかった!
- まずはこの一冊から意味がわかる多変量解析

主成分分析のイメージ

- 分散が最大となるような方向（座標、ベクトル）を見つけている
- 分散が大きい空間に写像することでサンプル間の違いが顕著になる

オリジナルの座標2



引用元) <https://www.intage.co.jp/glossary/401/>

引用元) <http://taustation.com/category/information-technology/machine-learning/>

分散を最大化する軸を見つけるとは？

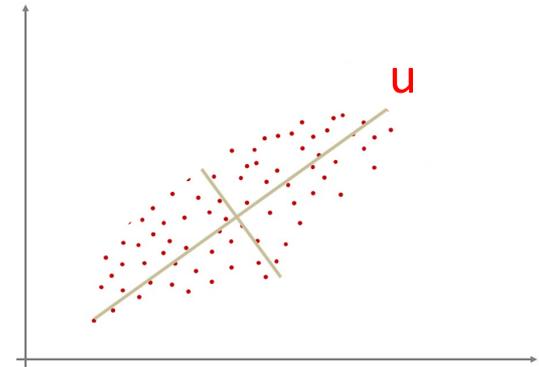
- 4つの変数(w, x, y, z)のデータがある
- 分散を最大にする変数の和を求める $\rightarrow u = k_1w + k_2x + k_3y + k_4z$, $k_1^2 + k_2^2 + k_3^2 + k_4^2 = 1$
- 分散を最大 \rightarrow 分散の二乗、 $S_u^2 = \frac{1}{n} \{(u_1 - \bar{u})^2 + (u_2 - \bar{u})^2 + \dots + (u_{17} - \bar{u})^2\}$
- S_u^2 を最大にする $k_1 \sim k_4$ (主成分負荷量)が求まる $\rightarrow u = 0.382w + 0.277x + 0.784y + 0.402z$

	k_1	k_2	k_3	k_4
番号	身長(w)	座高(x)	体重(y)	腹囲(z)
1	145.9	78.2	39.7	56.2
2	154.4	81.2	42.0	60.5
3	157.8	84.0	45.0	61.0
4	153.1	84.4	51.2	64.7
5	161.3	86.1	46.3	61.2
6	156.7	85.8	57.7	67.4
7	170.0	89.8	56.5	65.5
8	159.2	86.6	47.0	63.9
9	151.9	83.7	54.7	67.1
10	159.6	84.8	42.5	60.1
11	148.4	80.0	48.5	64.3
12	157.5	82.8	45.8	60.6
13	154.3	78.9	47.5	62.0
14	154.0	83.5	49.3	60.8
15	154.3	85.2	63.3	69.1
16	155.5	85.3	45.9	60.7
17	157.9	83.8	55.8	67.0
	(cm)	(cm)	(kg)	(cm)

$$u = k_1w + k_2x + k_3y + k_4z$$



主成分 (y)
u_1
u_2
u_3
u_4
u_5
u_6
u_7
u_8
u_9
u_10
u_11
u_12
u_13
u_14
u_15
u_16
u_17



主成分の寄与率と第2主成分

- 主成分はどれだけ元の情報を表しているか？分散で比較する
- 元の w, x, y, z の分散に対して、主成分 u の分散はどれくらいか？を見る
- (第1主成分の)寄与率 $= \frac{\text{主成分}u\text{の分散}}{\text{元の}w, x, y, z\text{の分散}} = \frac{s_u^2}{s_w^2 + s_x^2 + s_y^2 + s_z^2}$
- 第1主成分の残りについて再度分散が最大になる変数の和が第2主成分
 - 第1主成分の残りとは、
 - $w \rightarrow w - k_1 u$
 - $x \rightarrow x - k_2 u$
 - $y \rightarrow y - k_3 u$
 - $z \rightarrow z - k_4 u$
 - $v = l_1(w - k_1 u) + l_2(x - k_2 u) + l_3(y - k_3 u) + l_4(z - k_4 u)$ 、 $l_1^2 + l_2^2 + l_3^2 + l_4^2 = 1$
 - $v = 0.835w + 0.305x + 0.402y + 0.220z$
 - 第2主成分の寄与率 $= \frac{\text{主成分}v\text{の分散}}{\text{元の}w, x, y, z\text{の分散}} = \frac{s_v^2}{s_w^2 + s_x^2 + s_y^2 + s_z^2}$

主成分得点プロットと変量プロット

- 主成分得点プロットは、サンプルについて、主成分でプロット
- 変量プロットは、変数について、係数でプロット

番号	身長(w)	座高(x)	体重(y)	腹囲(z)
1	145.9	78.2	39.7	56.2
2	154.4	81.2	42.0	60.5
3	157.8	84.0	45.0	61.0
4	153.1	84.4	51.2	64.7
5	161.3	86.1	46.3	61.2
6	156.7	85.8	57.7	67.4
7	170.0	89.8	56.5	65.5
8	159.2	86.6	47.0	63.9
9	151.9	83.7	54.7	67.1
10	159.6	84.8	42.5	60.1
11	148.4	80.0	48.5	64.3
12	157.5	82.8	45.8	60.6
13	154.3	78.9	47.5	62.0
14	154.0	83.5	49.3	60.8
15	154.3	85.2	63.3	69.1
16	155.5	85.3	45.9	60.7
17	157.9	83.8	55.8	67.0

(cm) (cm) (kg) (cm)

変量プロット (係数)

$$u = 0.382w + 0.277x + 0.784y + 0.402z$$
$$v = 0.835w + 0.305x + 0.402y + 0.220z$$



主成分得点プロット

第1主成分 (u)	第2主成分 (v)
u_1	v_1
u_2	
u_3	
u_4	
u_5	
u_6	
u_7	
u_8	
u_9	
u_10	
u_11	
u_12	
u_13	
u_14	
u_15	
u_16	
u_17	