主成分分析 … 次元削除 データの中で分散が最大となる軸を見つけて測る.

なるべく少ない変数でデータの特徴を記述したい

id	使いやすさ	デザイン	価格
1	4	4	3
2	1	2	1
3	3	3	2
4	5	5	3
⋮		⋮	

一番いいサイトは?

- 合計点
- 主成分得点 イケてる!!

1つの値で見る.
要約.

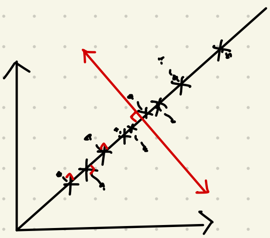
いい具合にやる.
- 量を減らす.
- 元の情報をたもつ

データの中で分散が最大となる軸を見つけて測る.

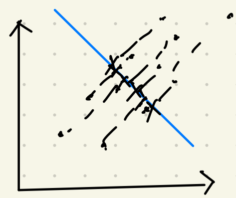
2変数

第1軸

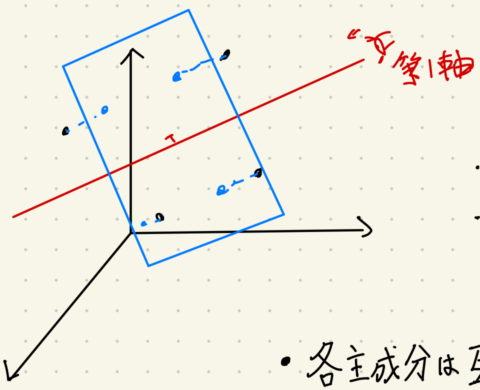
良い軸



よくない軸



第2軸



各主成分は互いに直交
- 主成分は変数の数と同じ数までいける
- 情報量 1>2>3 …

- 各主成分は互いに直交
- 主成分の変数の数と同じ数まで求めることができる
- 情報量 1>2>3 …

線形代数

行列 → $\begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix} = \begin{pmatrix} 0\times1+2\times2 \\ 2\times1+3\times2 \end{pmatrix}$

$(=)$ 行ベクトル　$(||)$ 列ベクトル

対称行列

$$\begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix}$$

単位行列

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I \Rightarrow \underbrace{AA^{-1}}_{\text{逆行列}} = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

固有値・固有ベクトル

$$Ax = \lambda x$$

A：ある行列

λ：定数：固有値

x：ベクトル：固有ベクトル

（固有値と固有ベクトルは）セット

例）

$$\underbrace{\begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix}}_{A}\underbrace{\begin{pmatrix} 1 \\ 2 \end{pmatrix}}_{x} = \begin{pmatrix} 4 \\ 8 \end{pmatrix} = \underbrace{4}_{\lambda}\underbrace{\begin{pmatrix} 1 \\ 2 \end{pmatrix}}_{x}$$

連立方程式の行列表現

$$\begin{cases} 3x + 5y = 13 \\ 4x + 2y = 6 \end{cases} \Rightarrow \begin{pmatrix} 3 & 5 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 13 \\ 6 \end{pmatrix}$$

$Z = a_1 x_1 + a_2 x_2 + \cdots$ ⇒ Zが主成分.

	x_1	x_2
1	P_1	k_1
2	P_2	k_2
平均	0	0
分散	$\frac{P_1^2+P_2^2}{2}$	$\frac{k_1^2+k_2^2}{2}$

共分散 $\frac{P_1k_1+P_2k_2}{2}$

分散共分散行列

$$\begin{pmatrix} 分散_1 & 共分散 \\ 共分散 & 分散_2 \end{pmatrix} = \begin{pmatrix} \frac{P_1^2+P_2^2}{2} & \frac{P_1k_1+P_2k_2}{2} \\ \frac{P_1k_1+P_2k_2}{2} & \frac{k_1^2+k_2^2}{2} \end{pmatrix}$$

$Z = a_1x_1 + a_2x_2$ … Zの分散が最大となる a_1, a_2 を求める

$a_1^2 + a_2^2 = 1$ … a_1, a_2 が大きくなるほどZの分散が大きくなるので制約

1, Zの平均 … $Z\mu = 0$: $\bar{x}_1 = \bar{x}_2 = 0$ より

2, Zの分散 … $Z\sigma^2 = \dfrac{(a_1P_1+a_2k_1)^2 + (a_1P_2+a_2k_2)^2}{2} = f(a_1, a_2)$

$g(a_1, a_2) = 0$ のもと, $f(a_1, a_2)$ を最大にする a_1, a_2 は.

解くべき問題の設定

$a_1^2 + a_2^2 = 1$ のもとで, $f(a_1, a_2)$ を最大にする a_1, a_2 を求める

ラグランジュの未定乗数法

$g(a_1, a_2) = 0$ のもと, $f(a_1, a_2)$ を最大にする a_1, a_2 は.

$F(a_1, a_2, \lambda) = f(a_1, a_2) - \lambda g(a_1, a_2)$ を a_1, a_2, λ で偏微分

$\dfrac{\partial F}{\partial a_1} = 0$, $\dfrac{\partial F}{\partial a_2} = 0$, $\dfrac{\partial F}{\partial \lambda} = 0$ の連立方程式

$$\boxed{\begin{aligned} &g(a_1, a_2) = 0 \\ &\Rightarrow a_1^2 + a_2^2 - 1 = 0 \\ &a_1^2 + a_2^2 = 1 \end{aligned}}$$

$$F(a_1, a_2, \lambda) = \frac{(a_1P_1+a_2k_1)^2 + (a_1P_2+a_2k_2)^2}{2} - \lambda(a_1^2 + a_2^2 - 1)$$

※合成関数の微分法

$$\frac{\partial F}{\partial a_1} = \frac{2(a_1P_1+a_2k_1)P_1 + 2(a_1P_2+a_2k_2)P_2}{2} - 2\lambda a_1 + 0 = 0$$

$$= \left(\frac{P_1^2+P_2^2}{2}\right)a_1 + \left(\frac{P_1k_1+P_2k_2}{2}\right)a_2 = \lambda a_1$$

$$\frac{\partial F}{\partial a_2} = \left(\frac{P_1k_1+P_2k_2}{2}\right)a_1 + \left(\frac{k_1^2+k_2^2}{2}\right)a_2 = \lambda a_2$$

$$\frac{\partial F}{\partial \lambda} = a_1^2 + a_2^2 = 1$$

$$
\begin{cases}
\left(\dfrac{p_1^2+p_2^2}{2}\right)a_1+\left(\dfrac{p_1k_1+p_2k_2}{2}\right)a_2=\lambda a_1\\[2ex]
\left(\dfrac{p_1k_1+p_2k_2}{2}\right)a_1+\left(\dfrac{k_1^2+k_1^2}{2}\right)a_2=\lambda a_2
\end{cases}
$$

連立方程式を
行列で表現すると

$$
\underbrace{\begin{pmatrix}\dfrac{p_1^2+p_2^2}{2} & \dfrac{p_1k_1+p_2k_2}{2}\\[2ex] \dfrac{p_1k_1+p_2k_2}{2} & \dfrac{k_1^2+k_2^2}{2}\end{pmatrix}}_{A}\underbrace{\begin{pmatrix}a_1\\ a_2\end{pmatrix}}_{x}=\underbrace{\lambda}_{\lambda}\underbrace{\begin{pmatrix}a_1\\ a_2\end{pmatrix}}_{x}
$$

A = 分散共分散行列

=> これを解くと a_1, a_2 を
求めることができる

$$Z_1 = 0.70u + 0.49d - 0.50p$$

$$Z_2 = -0.005u + 0.71d + 0.69p$$

$$Z_3 = -0.71u + 0.48d - 0.50p$$