PCA principal component analysis

a linear dimensionality reduction technique

$\nearrow$ centered but not scaled

the data is linearly transformed onto a new coordinate system such that the directions (principal components) capturing the largest variation in the data can be easily identified.

使结果尽可能地分散(方差变大)

centered data $\longrightarrow$ transformation $\longrightarrow$ new vector

X $n \times p$ W $l \times p$ T

$$x_{(i)} \;—\; \begin{array}{lcl} w_{(1)} = (w_1, \cdots, w_p)_{(1)} & — & x_{(i)} w_{(1)} \\ w_{(2)} = (w_1, \cdots, w_p)_{(2)} & — & x_{(i)} w_{(2)} \\ \vdots & & \vdots \\ w_{(l)} = (w_1, \cdots, w_p)_{(l)} & — & x_{(i)} w_{(l)} \end{array} \quad \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_l \end{pmatrix}_{(i)} = t_{(i)}$$

to maximize variance, $w_{(1)}$ has to satisfy

$$w_{(1)} = \underset{\|w\|=1}{\arg\max} \left\{ \sum_i (t_1)^2_{(i)} \right\}$$

$$= \underset{\|w\|=1}{\arg\max} \left\{ \sum_i (x_{(i)} w)^2 \right\}$$

$\downarrow$ matrix

$$= \underset{\|w\|=1}{\arg\max} \left\{ \|Xw\|^2 \right\}$$

$$= \underset{\|w\|=1}{\arg\max} \left\{ w^T X^T X w \right\}$$

$$\Rightarrow w_{(1)} = \arg\max \left\{ \frac{w^T X^T X w}{w^T w} \right\}$$

$\Downarrow$

recognised as a Rayleigh quotient

with $w_{(1)}$ found, we get

$t_{1(i)} = x_{(i)} w_{(1)}$ first PC of $x_{(i)}$ in the transformed co-ordinator

$\{x_{(i)} \cdot w_{(1)}\} w_{(1)}$ corresponding vector in the original variables.

further components (K-th)

subtracting the first K-1 principal components from X

$$\hat{X}_K = X - \sum_{s=1}^{K-1} X w_{(s)} w_{(s)}^T$$

maximum variance from the new data matrix

$$w_{(K)} = \underset{\|w\|=1}{\arg\max} \left\{ \| \hat{X}_K w \|^2 \right\} = \arg\max \left\{ \frac{w^T \hat{X}_K^T \hat{X}_K w}{w^T w} \right\}$$