

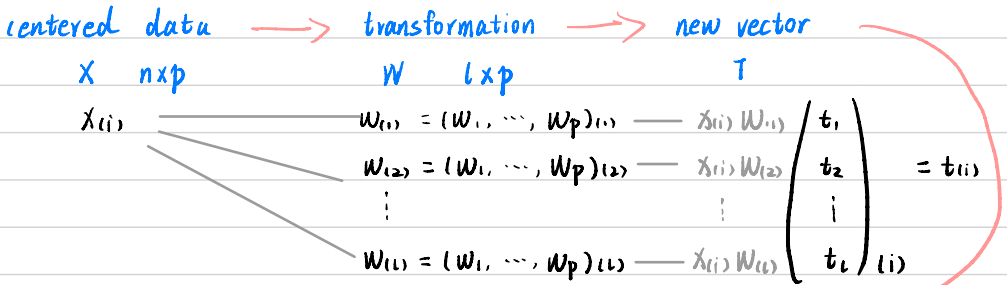
PCA principal component analysis

a linear dimensionality reduction technique

→ 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)



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

$$w_{(1)} = \arg \max_{\|w\|=1} \left\{ \sum_i (t_{(i)})^2 \right\}$$

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

↓ matrix

$$= \arg \max_{\|w\|=1} \left\{ \|X w\|^2 \right\}$$

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

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

↓
recognised as a Rayleigh quotient

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

$t_{(i)} = x_{(i)} w_{(1)}$ 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

$$W_{(K)} = \underset{\|W\|=1}{\operatorname{argmax}} \{ \|\hat{X}_K W\|^2 \} = \operatorname{argmax} \left\{ \frac{W^T \hat{X}_K^T \hat{X}_K W}{W^T W} \right\}$$