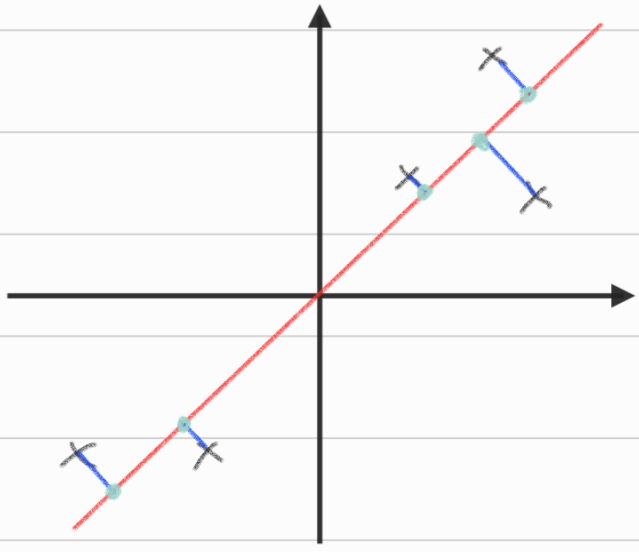
PCA problem formulation



blue line segments are short

PCA tries to find a

lower dimensional surface

onto which to project

the data so sum of squares

of blue line segments are

"minimized" ↳ Projection Error

〈분산최대한유지〉

PCA is not linear regression

PCA algorithm 주성분분석

전, Data preprocessing - feature scaling / mean normalization

PCA: tries to find a lower dimensional subspace onto which to project the data, to minimize projection error

Reduce data from n-dim to k-dim

- Compute "covariance matrix": $\Sigma = \frac{1}{m}\sum_{i=1}^{n}(x^{(i)})(x^{(i)})^T$ $n\times 1$ $1\times n$
- Compute "eigenvectors" of matrix Σ: [U,S,V]=svd(Sigma)
- $U = [u^{(1)}\ u^{(2)}\ \dots\ u^{(n)}] \in \mathbb{R}^{n\times n}$
- Z 구하기 (k)

* Singular value decomposition
* 특이값분해

$x \in \mathbb{R}^n \rightarrow z \in \mathbb{R}^k$

$$z = \begin{bmatrix} | & | & & | \\ u^{(1)} & u^{(2)} & \cdots & u^{(k)} \\ | & | & & | \end{bmatrix}^T x = \begin{bmatrix} — (u^{(1)})^T — \\ \vdots \\ — (u^{(k)})^T — \end{bmatrix} x$$

$n \times k$ (U reduce)

$k \times n$

$n \times 1$

↳ $k \times 1$ matrix

$z \in \mathbb{R}^k$