Principal Component Analysis

Find set of orthogonal directions in data space along which data are maximally variable

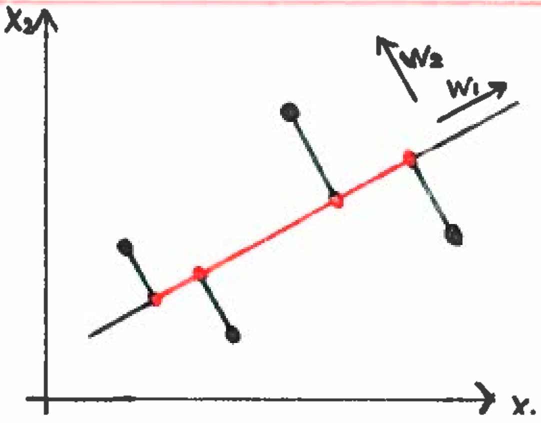
- PC directions $\underline{w}_1, \dots, \underline{w}_k \in \mathbb{R}^d$
- PC (scores) $\underline{z}_1, \dots, \underline{z}_n \in \mathbb{R}^k$

$$\max_{\underline{w}_1} \ \underline{w}_1^T \Sigma \underline{w}_1 \quad \text{s.t. } \|\underline{w}_1\| = 1$$

$$\max_{\underline{w}_m} \ \underline{w}_m^T \Sigma \underline{w}_m$$
$$\text{s.t. } \|\underline{w}_m\| = 1 \text{ and } \underline{w}_m^T \underline{w}_i = 0 \quad i = 1, \dots, m-1$$

Equivalent simultaneous variance maximisatio

$$\max_{\underline{w}_1, \dots, \underline{w}_k} \ \sum_{i=1}^{k} \underline{w}_i^T \Sigma \underline{w}_i$$
$$\text{s.t. } \|\underline{w}_i\| = 1, \ i = 1, \dots, k \text{ and } \underline{w}_i^T \underline{w}_j = 0, \ i \neq j$$



Minimisation of the approximation error

Equivalent PCA view

- Orthonormal $\underline{w}_1, \dots, \underline{w}_k \in \mathbb{R}^d$ span k-dim subspace

  Projection matrix : $P = \sum_{i=1}^{k} \underline{w}_i \underline{w}_i^T = W_k W_k^T$

  where $W_k = (\underline{w}_1 \dots \underline{w}_k)$

  $\leadsto$ decompose $\underline{x}$ into $\hat{\underline{x}} = P\underline{x} = \sum_{i=1}^{k} \underline{w}_i \underline{w}_i^T \underline{x}$

  and orth. residual $\underline{x} - P\underline{x}$

  $\hat{\underline{x}} \in \mathbb{R}^d$ is approximation of $\underline{x}$ in k-dim subspace

Optimisation problem for subspace with smallest expected approximation error :

$$\boxed{\begin{aligned} &\underset{\underline{w}_1, \dots, \underline{w}_k}{\text{minimise}} \ \mathbb{E}\left[\left\| \underline{x} - \sum_{i=1}^{k} \underline{w}_i \underline{w}_i^T \underline{x} \right\|^2\right] \\ &\text{s.t. } \|\underline{w}_i\| = 1, \quad i = 1, \dots, k \text{ and } \underline{w}_i^T \underline{w}_j = 0, \ i \neq j \end{aligned}}$$

is equivalent to simultaneous variance maximisation

$\leadsto$ Optimal subspace spanned by

$\underline{u}_1, \dots, \underline{u}_k$

Optimal $P = U_k U_k^T$

Optimal approximation: $\hat{\underline{x}} = U_k U_k^T \underline{x}$

$$= \sum_{i=1}^{k} \underline{u}_i \underline{u}_i^T \underline{x}$$

$$= \sum_{i=1}^{k} \underline{u}_i z_i$$

$\rightsquigarrow$ z_i is i-th coordinate of $\hat{\underline{x}}$ in space spanned by $\underline{u}_1, \dots, \underline{u}_k$

Expected approximation error:

$$\mathbb{E}[\|\underline{x} - U_k U_k^T \underline{x}\|^2]$$

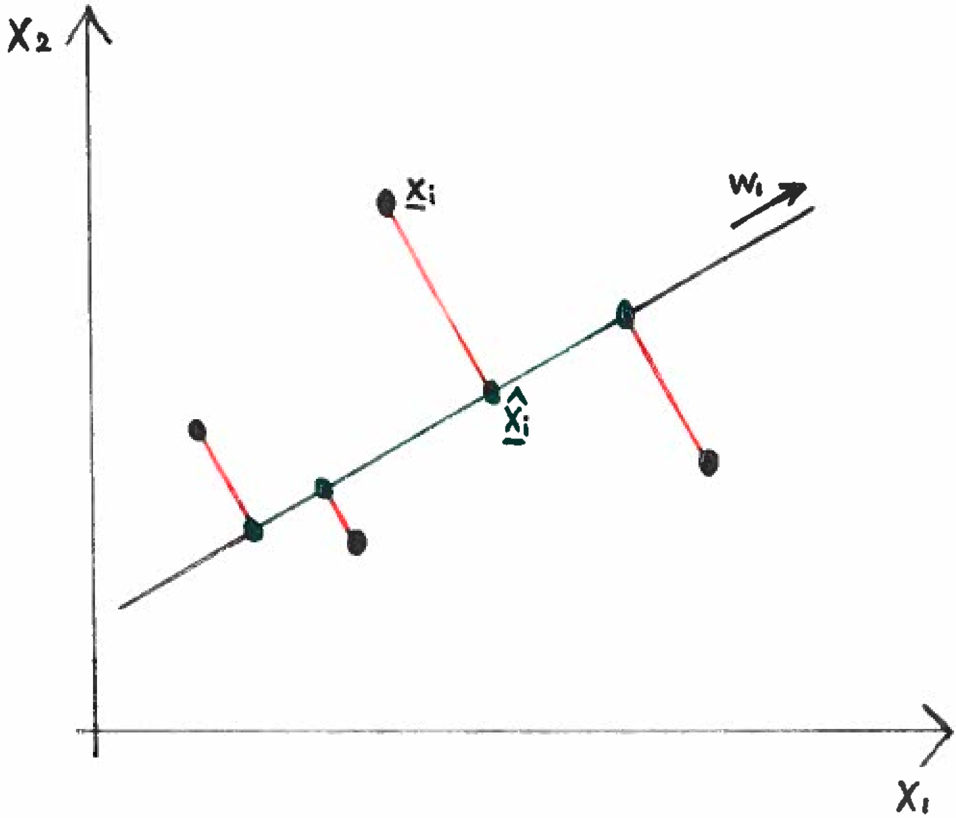
$$= \dots = \mathbb{E}[\underline{x}^T \underline{x}] - \mathbb{E}[\underline{x}^T U_k U_k^T \underline{x}]$$

$$= \mathbb{E}[\underline{x}^T \underline{x}] - \mathbb{E}[z_{1:k}^T \, z_{1:k}]$$

$$= \sum_{i=1}^{d} \mathrm{Var}[x_i] - \sum_{i=1}^{k} \mathrm{Var}[z_i]$$

$$= \sum_{i=1}^{d} \lambda_i - \sum_{i=1}^{k} \lambda_i = \sum_{i=k+1}^{d} \lambda_i$$

$\rightsquigarrow$ Minimising this error maximises variance explained



Variance maximisation

$\hat{=}$ Approximation error minimisation

Optimal low rank matrix approximation

Equivalent PCA view

Given: Centred $d \times n$ data matrix $X = (\underline{x}_1, \dots, \underline{x}_n)$
(Generally no EV decomposition available)

Singular value decomposition (SVD):

$$X = USV^T = \sum_{i=1}^{r} \underline{u}_i \, s_i \, \underline{v}_i^T \quad \text{where}$$

$d \times d$ $\quad U = (\underline{u}_1, \dots \underline{u}_d)$ orth. ("left singular vectors")

$n \times n$ $\quad V = (\underline{v}_1, \dots \underline{v}_n)$ orth. ("right " ")

$d \times n$ $\quad S = \left(\begin{array}{ccc|c} s_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & s_r & \\ \hline & 0 & & 0 \end{array}\right)$, $\quad s_1 \geq s_2 \geq \dots \geq s_r > 0$ ("singular values")

r rank of X

Approximate X by matrix M of rank $k < r$

Optimisation problem

$$\boxed{\begin{aligned} &\underset{M}{\text{minimise}} \; \|X - M\|_F \\ &\text{s.t. rank}(M) = k \end{aligned}}$$

where $\|A\|_F = \sum_{i,j} (A)_{ij}^2$ "Frobenius norm"

Solution from linear algebra: $M = \sum_{i=1}^{k} s_i \, \underline{u}_i \, \underline{v}_i^T$

with approximation error $\|X - M\|_F = \sum_{i=k+1}^{r} s_i^2$

Relation to PCA: via sample covariance matrix: $\Sigma = \mathrm{cov}(X)$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i \underline{x}_i^T = \frac{1}{n} X X^T$$

$$= \frac{1}{n} U S V^T (V S^T U^T)$$

$$= U \left(\frac{1}{n} S S^T\right) U^T$$

$\Rightarrow$ u_i (left sing. vect.) of X are the EV of Σ with eigenvalues $\lambda_i = \frac{s_i^2}{n}$

$\Rightarrow$ $\underline{u}_i$ is i-th PC direction with PC scores

$$z_{i:} = \underline{u}_i^T X = \underline{u}_i^T U S V^T = \underline{u}_i^T \sum_{j=1}^{r} \underline{u}_j s_j \underline{v}_j^T$$

$$\overset{\text{orth.}}{=} s_i \underline{v}_i^T$$

Approximate data matrix:

$$M = \sum_{i=1}^{k} \underline{u}_i z_{i:}$$

Approximating the sample covariance matrix:

Equivalent PCA view

Low rank approximation of $\Sigma = \mathrm{cov}(X)$ also yields PC directions

Optimisation problem

$$\boxed{\begin{array}{l} \underset{M}{\text{minimise}}\ \|\Sigma - M\|_F \\ \text{s.t. } \operatorname{rank}(M) = k \text{ and } M^T = M \end{array}}$$

equivalent PCA view

→ Directions in data space with maximal variance also maximally preserve covariance structure

Approximating the Gram matrix
Equivalent PCA view

Gram matrix: $n \times n$ matrix $G = X^T X$

Using SVD:

$$G = (USV^T)^T (USV^T) = V S^T S V^T = V \tilde{\Lambda} V^T$$

where $\tilde{\Lambda} = \begin{pmatrix} s_1^2 & & & 0 \\ & s_2^2 & & \\ & & \ddots & \\ 0 & & & s_n^2 \end{pmatrix}$ } can be 0

$\Rightarrow$ $\underline{v}_i$ (right sing. vect.) of X are EV of G with eigenvalues $\tilde{\lambda}_i = s_i^2$

Optimal rank k approximation:

$$\hat{G} = \sum_{i=1}^{k} \underline{v}_i s_i^2 \underline{v}_i^T = \sum_{i=1}^{k} z_{i:}^T z_{i:} \quad \text{where } z_{i:} = s_i \underline{v}_i^T$$

$\Rightarrow$ minimising average approximation error of data also maximally preserves inner product structure

$k \times n$ matrix with PC scores as rows:

$$\boxed{Z = \sqrt{\tilde{\Lambda}_k}\, V_k^T} \quad \text{where } V_k = (\underline{v}_1 \dots \underline{v}_k),\ \tilde{\Lambda}_k \begin{pmatrix} \tilde{\lambda}_1 & 0 \\ 0 & \tilde{\lambda}_k \end{pmatrix}$$

$\rightsquigarrow$ We can compute PC scores directly from Gram matrix without PC directions