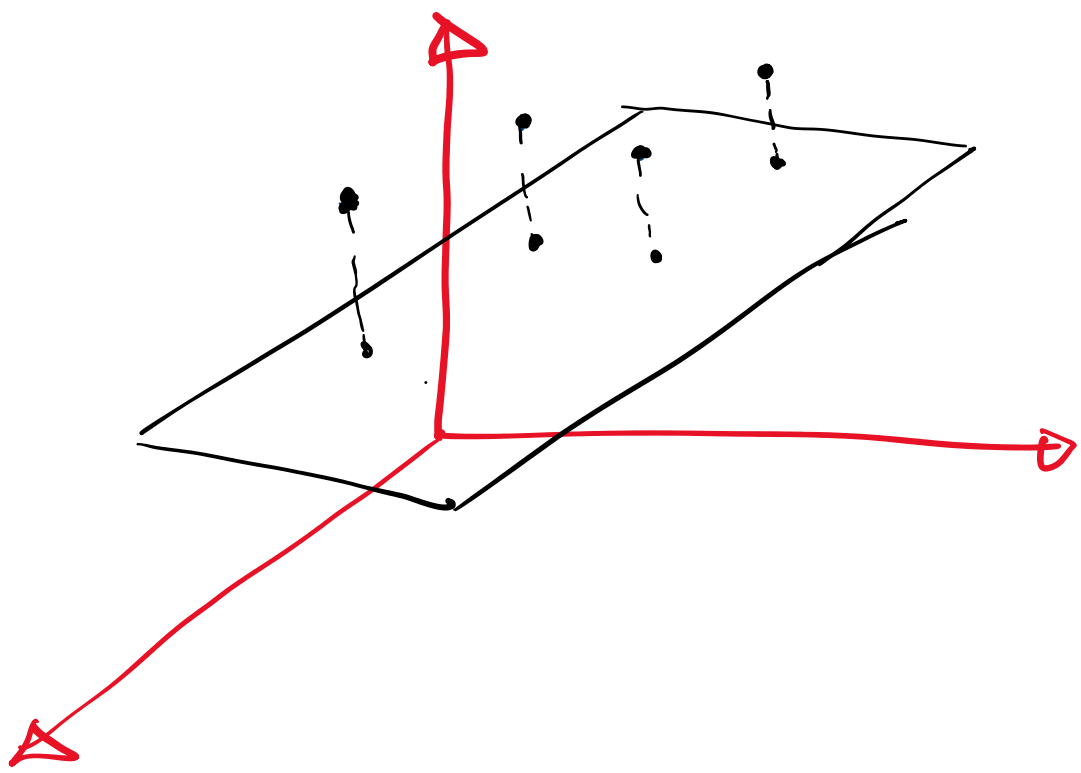
Principal Component Analysis (PCA)

: $x_n \in \mathbb{R}^D$; D : # of variables

x_n, $n = 1, \ldots, N$: # of observations

* Goal: To project the data onto a space having dimension $M < D$ while maximizing variance of the projected data



Let $M = 1$: one-dimension space

u_1 : D-dimensional vector

also, $u_1^T u_1 = 1$: Unit vector

Then, the Projection, for x_n a scaler value $= u_1^T x_n$

* Mean vector : $\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$

The variance of the projected data

$$\frac{1}{N} \sum_{n=1}^{N} (u_1^T x_n - u_1^T \bar{x})^2$$

Page 2

10 August 2023 21:53

$$\frac{1}{N} \sum_{i=1}^{N} (u_1^T x_n - u_1^T \bar{x})^2 = u_1^T S u_1$$

S : $D \times D$ - dimensional covariance

$$: \frac{1}{N} \sum_{i=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T$$

Then, find u_1 that maximize the variance of the projected data

$$\boxed{\underset{u_1}{\text{maximize}} \quad u_1^T S u_1 \quad \text{s.t.} \quad u_1^T u_1 = 1} \quad - (01)$$

$\therefore$ The constraint optimization function 01 has to be solved.

* Lagrangian multiplier technique

(01) can be written with λ, a lagrange variable as follows:

$$\max_{u_1, \lambda} \; u_1^T S u_1 + \lambda (1 - u_1^T u_1)$$

Page 3

10 August 2023 21:59

Result from matrix theory:

$$\frac{\partial x^T B x}{\partial x} = (B + B^T) x$$

Then,

$$\frac{\partial J}{\partial u_1} = (S + S^T) u_1 - 2\lambda_1 u_1 = 0$$

$$\Longleftrightarrow \quad 2 S u_1 - 2\lambda_1 u_1 = 0$$

$$\Longleftrightarrow \quad \boxed{S u_1 = \lambda_1 u_1}$$

eigenvector corresponding to the largest eigenvalue λ_1

* $\rightarrow$ $M = 2$; second unit vector u_2 similar to u_1;

$$\max_{u_2} \quad u_2^T S u_2$$

$$\text{s.t.} \quad u_2^T u_2 = 2$$

$$u_1^T u_2 = 0 \quad \text{(orthogonal)}$$

* Lagrangian multipliers

$$\max_{u_2, \lambda_1, u_2} \quad u_2^T S u_2 + \lambda_2 (1 - u_2^T u_2)$$

$$\max_{u_2, \lambda_2, \mu_2} \quad u_2' S u_2 + \lambda_2 (1 - u_2^T u_2) + \mu_3 u_1^T u_2$$

Page 4

10 August 2023 22:19

$$\frac{\partial J}{\partial u_2} = 0 \iff S u_2 - \lambda_2 u_2 + \mu_3 u_1 = 0$$

$$\iff u_1^T S u_2 - \lambda_2 \underbrace{u_1^T u_2}_{0} + \mu_3 u_1^T u_1 = 0$$

$$\underbrace{\mathrm{cov}(u_1^T \bar{x}_n, u_2^T \bar{x}_n) = 0}_{\text{uncorrelated variables}}$$

$$\Rightarrow u_1^T S u_2 = 0$$

$$\therefore \iff \mu_3 = 0 \text{ as } u_1^T u_1 = 1.$$

$$\therefore \boxed{S u_2 = \lambda_2 u_2}$$

second largest eigenvalue & eigenvector for the same.

$$* \quad S \cong Q^T \Lambda Q$$

$$Q = [u_1 \; u_2 \; \dots \; u_D]$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \; \lambda_D \end{bmatrix}$$

Principal component Analysis

$\{x_n\}_{n=1}^{N}$: 1 compute $\bar{x}$ & S from the data

2. Compute eigenvectors and eigenvalues of S

3. Find the M eigenvectors of S corresponding to the M largest eigenvalues.

* M: Eigenvectors are principal components.
(PC)

1st PC: corresponding to the largest Eigenvalue

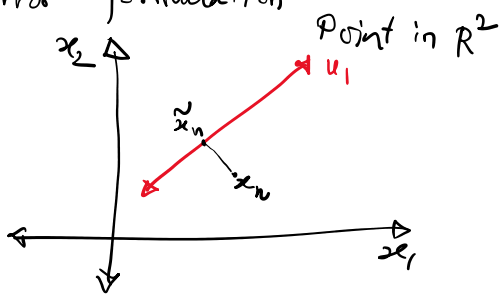
2nd PC: " to the second largest Eigenvalue

⋮

M^{th} PC: " " " M^{th} largest Eigenvalue.

Minimum-error formulation

$\{x_n\}_1^N$:



Problem: Find a new set of orthogonal basis in M dimensional subspace ($< D$), such that projection error is minimized

* Consider a set of new basis vectors $\{u_1, u_2, \ldots, u_D\}$ (complete basis)
  Orthonormal basis

$$u_i^T u_j = \delta_{ij} \rightarrow \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

* $x_n = \alpha_{n1} u_1 + \alpha_{n2} u_2 + \ldots + \alpha_{nD} u_D$

  co-efficients for different points

* Compute $x_n^T u_j = \alpha_{n1} \cancelto{0}{u_1^T u_j} + \alpha_{n2} \cancelto{0}{u_2^T u_j} + \alpha_{nj} \underbrace{u_j^T u_j}_{=1} + \ldots + \alpha_{nD} \cancelto{0}{u_D^T u_j}$

  $= \alpha_{nj}$

* For the jth α, $\alpha_{nj} = x_n^T u_j = u_j^T x_n$

* $x_n = \sum_{i=1}^{D} \alpha_{ni} u_i = \sum_{i=1}^{D} (x_n^T u_i) u_i$

* Our Interest, finding $M (< D)$ variables

$$\tilde{x}_n = \sum_{i=1}^{M} \underbrace{z_{ni}}_{\text{change with the point}} u_i + \underbrace{\sum_{i=M+1}^{D} b_i u_i}_{\text{constant for a point}}$$

$\therefore$ Then, minimize error between x_n and $\tilde{x}_n$

$\therefore$ Then, minimize error between x_n and $\tilde{x}_n$

$$\min_{z_{nj}, b_j, u_j} J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \tilde{x}_n \|_2^2$$

$$= \frac{1}{N} \sum_{n=1}^{N} \| x_n - \sum_{i=1}^{M} z_{ni} u_i - \sum_{i=M+1}^{D} b_i u_i \|_2^2$$

* Determine z_{nj} and b_j by computing

$$\frac{\partial J}{\partial z_{nj}} = 0 \quad \text{and} \quad \frac{\partial J}{\partial b_j} = 0$$

$$\rightarrow \frac{\partial J}{\partial z_{nj}} = 0 \iff \frac{\partial \tilde{x}_n}{\partial z_{nj}}^T (x_n - \tilde{x}_n) = 0$$

$$\iff u_j^T \left(x_n - \sum_{i=1}^{M} z_{ni} u_i - \sum_{i=M+1}^{D} b_i u_i \right) = 0$$

$$\iff u_j^T x_n - z_{nj} = 0$$

$$\iff z_{nj} = u_j^T x_n \text{ or } x_n^T u_j, \; j = 1, \ldots, M$$

$$\rightarrow \frac{\partial J}{\partial b_j} = 0 \iff \frac{\partial \tilde{x}_n}{\partial b_j}^T (x_n - \tilde{x}_n) = 0$$

$$\iff b_j = \bar{x}_n^T u_j, \; j = M+1, \ldots, D$$

Page 2

12 August 2023 17:31

$$x_n - \tilde{x}_n = x_n - \sum_{i=1}^{M} z_{n_i} u_i - \sum_{i=M+1}^{D} b_i u_i$$

$$= \sum_{i=1}^{D} \alpha_{n_i} u_i - \sum_{i=1}^{M} z_{n_i} u_i - \sum_{i=1}^{D} b_i u_i$$

$$= \cancel{\sum_{i=1}^{M} (x_n^T u_i) u_i} + \sum_{i=M+1}^{D} (x_n^T u_i) u_i - \cancel{\sum_{i=1}^{M} (x_n^T u_i) u_i}$$

$$- \sum_{i=M+1}^{D} (\bar{x}_n^T u_i) u_i$$

$$= \sum_{i=M+1}^{D} \{ (x_n^T - \bar{x}_n^T) u_i \} u_i$$

* Let $\|x_n - \tilde{x}_n\|_2^2 = (x_n - \tilde{x}_n)^T (x_n - \tilde{x}_n)$

$$= \sum_{i=M+1}^{D} \{ u_i^T (x_n - \bar{x}_n) \} u_i^T u_i \{ (x_n - \bar{x}_n)^T u_i \}$$

$$= \sum_{i=M+1}^{D} u_i^T (x_n - \bar{x}_n)(x_n - \bar{x}_n)^T u_i$$

* $$J = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=M+1}^{D} u_i^T (x_n - \bar{x}_n)(x_n - \bar{x}_n)^T u_i$$

$$= \sum_{i=M+1}^{D} u_i^T \Big\{ \underbrace{\frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x}_n)(x_n - \bar{x}_n)^T}_{S} \Big\} u_i$$

$$J = \sum_{i=1}^{D} u_i^T S u_i$$

* $$\boxed{\underset{u_i}{\text{minimize}}\ J = \sum_{i=M+1}^{D} u_i^T S u_i} \quad \text{— } J$$

* Let us take $D = 2$ and $M = 1$ for a simplification

* The objective function

$$\underset{u_2}{\text{minimize}}\ J = u_2^T S u_2$$

$$\text{s.t} \quad u_2^T u_2 = 1 \qquad \{ \text{otherwise trivial answer } \|u_2\| = 0 \}$$

* Lagrangian

$$\min_{u_2} \quad u_2^T S u_2 + \lambda (1 - u_2^T u_2)$$

$$\min_{u_2} \quad u_2^T S u_2 + \lambda (1 - u_2^T u_2)$$

* $\frac{\partial J}{\partial u_2} = 0 \iff \boxed{S u_2 = \lambda_2 u_2}$.

* $J = u_2^T S u_2 + \lambda_2 (1 - u_2^T u_2)$
$= u_2^T \lambda_2 u_2 = \lambda_2$

* J takes a minimum value at $\lambda_2 \rightarrow$ by choosing smaller eigenvectors and

Then, to achieve the minimum error projection, in $D = 2$, and $M = 1$, the Principal component subspace corresponding to the largest eigenvalue should be chosen.

* For D-dimensional vector; and $M < D$.
  Then, by taking $\{u_i\}$, $i = 1, \ldots M$ corresponding to the M largest eigenvalues, the error objective fn

$$J = \sum_{i=M+1}^{D} \lambda_i \text{ is minimum.}$$

* PCs $\rightarrow$ the M eigenvectors corresponding to the largest eigenvalues

* orthogonal space: $(D - M)$ eigenvectors corresponding to the smallest Eigenvalues.