$D \in \mathbb{R}^{n \times d}$

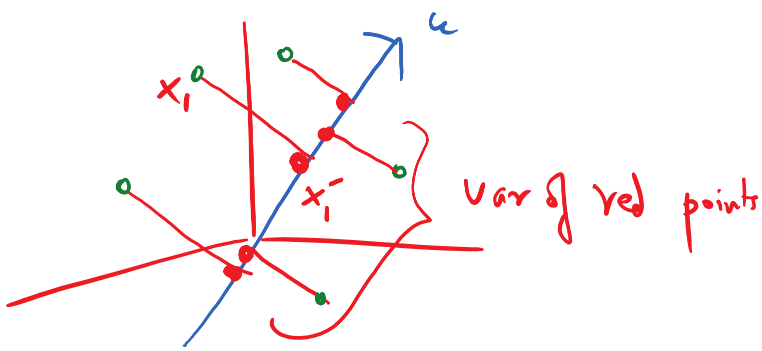
d can also be large

a direction that capture some info about D

objective

max. variance

$\vec{u}$?



$\vec{x} \in \mathbb{R}^d$

$\vec{x}_1'$ projection onto $\vec{u}$

$$\boxed{\vec{x}' = \left(\frac{x^T u}{u^T u}\right)\vec{u}}$$

orthogonal projection

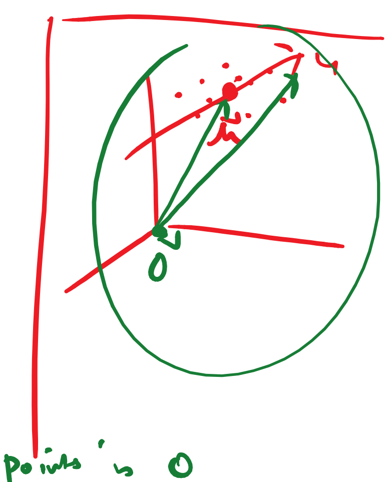
$\vec{u}$ has been normalized

$u^T u = 1 \qquad \|\vec{u}\| = 1$

assume that D has been centered, i.e. $\hat{\vec{\mu}} = \vec{0}$

$$\vec{x}_i' = \overbrace{(x_i^T u)}^{a_i}\,\vec{u}$$

scalar coefficient/ coordinate value

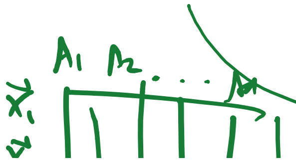


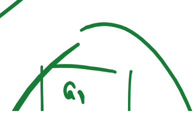
$D \rightarrow$ center $\rightarrow$ project onto $\vec{u}$

$\downarrow$

$\vec{x}_i \in \mathbb{R}^d \qquad a_i \in \mathbb{R}$

mean of projected points is 0





x_1 x_2 $\ldots$ x_L $\Rightarrow$ $a_1, a_2, \ldots, a_n$

$$\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - \underline{\underline{0}})^2$$

$$\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n} a_i^2$$

$$a_i = \underbrace{x_i^T u}$$

$$\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^T u)^2$$

$$(x_i^T u)\cdot(x_i^T u)$$

$$(u^T x_i)(x_i^T u)$$

$$u^T(\underbrace{x_i x_i^T})u$$

$$= \frac{1}{n}\sum_{i=1}^{n}\vec{u}^T(x_i x_i^T)u$$

$$= \vec{u}^T\Big(\underbrace{\frac{1}{n}\sum x_i x_i^T}\Big)u$$

Covariance matrix for centered Data

$$\sigma_u^2 = \vec{u}^T\Sigma\vec{u}$$

Objective

$$\max_{\vec{u}}\{\sigma_u^2\} = \max_{\vec{u}}\{\underbrace{\vec{u}^T\Sigma\vec{u}}\}$$

Convex function
ie. unique optimal solution

if $\vec{u}$ is an eigenvector for Σ

$$\Sigma\vec{u} = \lambda\vec{u}$$

$$\max_{\vec{u}} \quad \vec{u}^T(\Sigma\vec{u})$$

$$= \vec{u}^T(\lambda\vec{u})$$

$$= \lambda\,\underline{\vec{u}^T\vec{u}}$$

$$= (\lambda)$$

$$\big(\vec{u}^T\Sigma\vec{u}\big)$$

$= \lambda$

to maximize choose largest eigenvalue, i.e. λ_1

solution is $\vec{u}_1$ ← dominant eigenvector!

$$\max_{\vec{u}} \left(\frac{\vec{u}^T \Sigma \vec{u}}{\vec{u}^T \vec{u}} \right)$$

Rayleigh Coefficient

Lagrange multipliers:

$$\underbrace{\max_{\vec{u}} \; \vec{u}^T \Sigma \vec{u}}_{\text{objective}}, \quad \text{s.t. } \underbrace{\vec{u}^T \vec{u} = 1}_{\text{constraint}} \rightarrow (\vec{u}^T \vec{u} - 1) = 0$$

$$\max_{\vec{u}} \underbrace{\left\{ \vec{u}^T \Sigma \vec{u} - \alpha \left(\vec{u}^T \vec{u} - 1 \right) \right\}}_{J(\vec{u})}$$

$\vec{u} \in \mathbb{R}^d$

$$\frac{\partial J(\vec{u})}{\partial \vec{u}} = \underbrace{\Sigma (\cancel{2} \vec{u}) - \alpha (\cancel{2} \vec{u}) = 0}$$

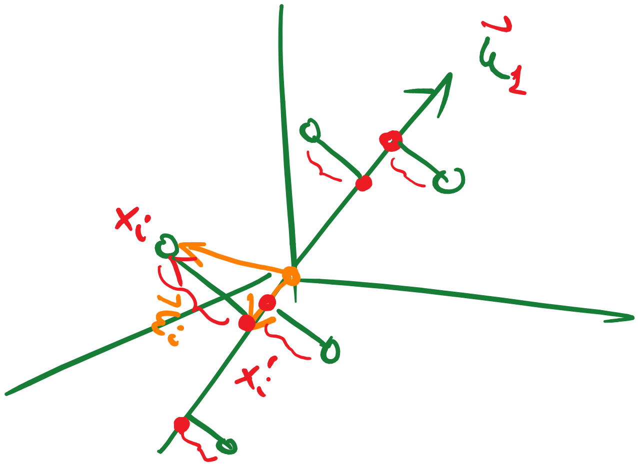
vector derivative w.r.t. $\vec{u}$

$$\Sigma \vec{u} - \alpha \vec{u} = 0$$

or

$$\Sigma \vec{u} = \alpha \vec{u}$$

λ_1 is in fact the projected variance σ_u^2 !



$$\boxed{\sigma_u^2 = \lambda_1}$$

$\vec{u}_1$: 1^{st} principal component (PC)

$$\underbrace{D \in \mathbb{R}^d \longrightarrow \vec{u}_1 \in \mathbb{R}}_{\underbrace{\text{lost info}}_{\text{error?}}}$$

$$\vec{\varepsilon}_i = \vec{x}_i - \vec{x}_i'$$

Mean squared error (MSE)

$$MSE(\vec{u}) = \frac{1}{n} \sum_{i=1}^{n} \|\varepsilon_i\|^2$$

$$\min_{\vec{u}} \{ MSE(\vec{u}) \} \longleftarrow \text{new objective}$$

$u^T u = 1$

$$x_i' = (x_i^T \vec{u}) \vec{u}$$

$$\begin{aligned}
\|\varepsilon_i\|^2 &= \vec{\varepsilon}_i^T \vec{\varepsilon}_i \\
&= (\vec{x}_i - \vec{x}_i')^T (\vec{x}_i - \vec{x}_i') \\
&= \left(\vec{x}_i - \underline{(x_i^T \vec{u})\vec{u}}\right)^T \left(\vec{x}_i - \underline{(x_i^T u)\vec{u}}\right) \\
&= x_i^T x_i - (x_i^T u)^2 \\
&\quad - (x_i^T u)^2 \\
&\quad + (x_i^T u)^2 \underbrace{\cancel{u^T u}}_{1} \\
&= \|x_i\|^2 - \left((x_i^T \ldots)^2 \right.
\end{aligned}$$

$$= \|x_i\|^2 - (x_i^T u)^2$$

$$\Sigma$$

$$\|\varepsilon_i\|^2 = \|x_i\|^2 - \vec{u}^T (x_i x_i^T) \vec{u}$$

$$MSE = \frac{1}{n} \sum \|\varepsilon_i\|^2$$

$$\min MSE = \frac{1}{n} \sum \|x_i\|^2 - \vec{u}^T \Sigma \vec{u}$$

D is centered

$$\min MSE = \text{totvar}(D) - \underbrace{\vec{u}^T \Sigma \vec{u}}_{\max}$$

to minimize the squared error you have to maximize the projected variance

$$\vec{u}_1^T \Sigma \vec{u}_1 = \lambda_1$$

$$\boxed{MSE = \text{totvar}(D) - \lambda_1}$$

squared error = sum of remaining total variance

Pick r Principal Components

$$\lambda_1, \lambda_2, \ldots, \lambda_r, \ldots, \lambda_d$$

$$\vec{u}_1 \quad \vec{u}_2 \quad \ldots \quad \vec{u}_r \quad \ldots \quad \vec{u}_d$$

$u_1 \quad u_2 \quad \cdot \quad \cdot \quad \cdot \quad u_r \;|\; \cdot \quad \cdot \quad \cdot \quad \vec{u_d}$

Pick the top r eigenvalue

$$\boxed{MSE = \text{totvar}(D) - \sum_{i=1}^{r} \lambda_r}$$

error in approximation

Variance captured by the new r-dim subspace

$D' \in \mathbb{R}^{n \times r}$

$$\Sigma = U \Lambda U^T$$

Cov matrix

Cov in eigen space

matrix decomposition

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_d \\ | & | & & | \end{bmatrix}$$

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

$\Sigma \vec{u_i} = \lambda_i \vec{u_i}$

$\underbrace{}_{d \text{ equations}}$

$$\Sigma = \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{bmatrix} \begin{bmatrix} - & u_1^T & - \\ - & u_2^T & - \\ & \vdots & \\ - & u_d^T & - \end{bmatrix}$$

$$\boxed{\Sigma = \sum_{i=1}^{d} \lambda_i \underbrace{u_i u_i^T}}$$

Spectral sum

$u_i u_i^T$

$=$

$d \times d$ matrix

rank 1 $\vec{u_i}$

$$= \underbrace{\lambda_1 u_1 u_1^T}_{\text{Most important layer}} + \lambda_2 u_2 u_2^T + \cdots + \lambda_r u_r u_r^T \;|\; + \cdots + \lambda_d u_d u_d^T$$

most important layer

$\underbrace{\qquad\qquad\qquad}_{\text{keep } r \text{ layers}}$

discard remaining $d-r$ layers

Singular Value Decomposition (SVD)

PCA $\rightarrow \Sigma$

Matrix decomposition

$D \in \mathbb{R}^{n \times d}$ $\leftarrow$ original data

$D_q \leftarrow$ rank q approx $\quad q << d$

approximation

$\|A\|_F = \| D - D_q \|_F =$ Frobenius Norm

$\sqrt{\sum_{i=1} \sum a_{ij}^2}$

$D \in \mathbb{R}^{n \times d}$

$\rightarrow$ linear transformation

$T(\vec{x}) : D\vec{x} = \vec{y}$

$T: \mathbb{R}^d \rightarrow \mathbb{R}^n$ $\leftarrow$ function

$\vec{x} \in \mathbb{R}^d \qquad \vec{y} \in \mathbb{R}^n$

$(n \times d) \times (d \times 1)$

$\rightarrow (n \times 1)$

what about the set or space?

$(n \times d) \times (d \times 1)$

$\rightarrow (n \times 1)$

what about the set or space of all possible $\vec{y}$'s ?

Column space of D

$$\underbrace{\begin{bmatrix} A_1 & A_2 & \cdots & A_d \end{bmatrix}}_{D}^{\,n} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}}_{\vec{x}} = \vec{y} = \Big(\underbrace{\vec{A}_1 x_1 + \vec{A}_2 x_2 + \cdots + \vec{A}_d x_d}_{\text{linear combination of the cols}} \Big)$$

role of weight vector

row-space of D : all the linear combinations of rows of D

: Col space of D^T

$$\underset{d \times n}{D^T} \cdot \underset{n \times 1}{\vec{y}} \rightarrow \underset{d \times 1}{\vec{x}}$$

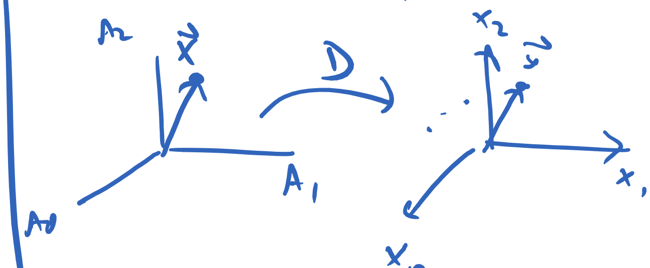
$\mathbb{R}^n \longrightarrow \mathbb{R}^d$

eigen

$$A\vec{x} = \lambda \vec{x}$$

$d \times d$

$D: \mathbb{R}^d \rightarrow \mathbb{R}^n$



$$D\vec{x} = \vec{0}$$

$$\vec{x} \neq 0$$

Null space of D

$$\Sigma = U \Lambda U^T$$

$$\mathbb{R}^d$$

singular value

$$\underbrace{D\vec{x}}_{\text{right}} = \sigma \underbrace{\vec{y}}_{\text{left}} \quad \mathbb{R}^n$$

Singular vectors

$$\underset{n\times d}{D} = \underset{n\times n}{L}\,\underset{n\times d}{\Delta}\,\underset{d\times d}{R}^T$$

Δ is a "diagonal matrix"

d

n

$$= \underbrace{\begin{bmatrix} l_1 & l_2 & \cdots & l_n \\ & L & \end{bmatrix}}_{\substack{\text{left singular}\\ \text{vectors}\\ n\times n}} \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \; \sigma_d \\ \hline & 0 & \end{bmatrix}}_{\substack{\text{singular}\\ \text{values}}} \begin{bmatrix} r_1^T \\ r_2^T \\ \vdots \\ r_d^T \end{bmatrix} R^T$$

orthogonal matrices

$d\times d$

right singular vectors

$$\boxed{D = \sum_{i=1}^{r} \sigma_i\, l_i\, r_i^T}$$

$\sum \lambda_i u_i u_i^T$

eigen

r – layers

r : rank of $D \le \min(n,d)$

D_q : rank q approximation

keep the q largest singular values

l_i, r_i^T

PCA vs SVD

$$\underbrace{D = L\Delta R^T}_{SVD}$$

$l = [\vec{0}\ \vec{0}\ \ \ \vec{0}]$

$$\Sigma = \frac{1}{n} D^T D = (L\Delta R^T)^T (L\Delta R^T)$$

cov

$$\underbrace{d\times n \cdot n\times d}_{d\times d}$$

$$= R\Delta^T \underbrace{L^T L}_{I} \Delta R^T$$

$$L = \left[\vec{l}_1 \; \vec{l}_2 \dots \vec{l}_r\right]$$

$$R = \left[\vec{r}_1 \; \vec{r}_2 \dots \vec{r}_r\right]$$

r : rank of D

Δ = $r \times r$ matrix

$$= R \Delta^T \Delta R^T$$

$$= \frac{1}{n}\left(R \Delta^2 R^T\right)$$

$$= R\left(\frac{\Delta^2}{n}\right) R^T \quad \text{vs} \quad U \Lambda U^T$$

right singular vectors of D are the eigenvectors of Σ

$\sigma_i \leftarrow$ singular value is related to the eigenvalue of Σ

$$\boxed{\lambda_i = \frac{\sigma_i^2}{n}} \geq 0$$

Variance