Dim Reduction

10/23

- if 2^d in an alg time complexity
  $c << d$ run alg on lower dim data 2^c in run time

- data viz 2 & 3 d are easier to understand

Background

data matrix D

$$\vec{x}_1^T \in \mathbb{R}^d$$

$$D = \begin{pmatrix} \boxed{x_{11} \quad x_{12} \quad \cdots \quad x_{1d}} \\ x_{21} \quad x_{22} \quad \cdots \quad x_{2d} \\ \vdots \qquad \ddots \qquad \vdots \\ x_{n1} \quad x_{n2} \quad \cdots \quad x_{nd} \end{pmatrix}$$

Recall e_i i^{th} std basis vector
$(0,0,\dots 0,1,0\dots,0)$
↖ i^{th} entry is 1
all other entries are 0

$\vec{e}_i \cdot \vec{e}_i = \|\vec{e}_i\| = 1$ ← all vectors are length 1
$i \neq j$ $\vec{e}_i \cdot \vec{e}_j = 0$ ← all vectors are othogonal
$\Rightarrow$ orthonormal

$u_i \cdot u_i = 1$
$i \neq j$ $u_i \cdot u_j = 0$

Given any other set $\vec{u}_1, \dots, \vec{u}_d$ of $\boxed{\text{orthonormal vectors}}$

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

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \dots + x_d\vec{e}_d$$
$$= a_1\vec{u}_1 + a_2\vec{u}_2 + \dots + a_d\vec{u}_d$$

e.g. $\vec{a} = (a_1, a_2, \dots, a_d)$
↖ our point $\vec{x}$ in the basis of $\vec{u}_i$'s

Lets expess as a linear transformation

$$U = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_d \\ | & | & & | \end{pmatrix} \leftarrow \text{as } u_i\text{'s are orthogonal and each } \|u_i\| = 1 \Rightarrow \text{orthonormal matrix}$$

$\Downarrow$

prop $\Rightarrow$ $U^{-1} = U^T$

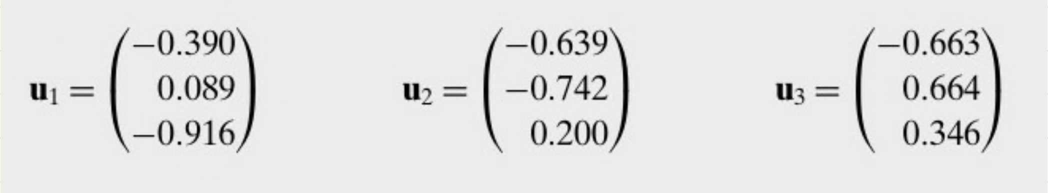
$\vec{x} = U \vec{a}$ — want to find $\vec{a}$

U^T

$$\Leftrightarrow U^{-1}\vec{x} = \overset{I}{\boxed{U^{-1}U}}\vec{a}$$

$$U^T\vec{x} = \vec{a}$$

Eg. convert pt $\begin{pmatrix} -.343 \\ -.754 \\ .241 \end{pmatrix}$ w/ basis (which is orthonormal)



$$\mathbf{u}_1 = \begin{pmatrix} -0.390 \\ 0.089 \\ -0.916 \end{pmatrix} \qquad \mathbf{u}_2 = \begin{pmatrix} -0.639 \\ -0.742 \\ 0.200 \end{pmatrix} \qquad \mathbf{u}_3 = \begin{pmatrix} -0.663 \\ 0.664 \\ 0.346 \end{pmatrix}$$

given point in std basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

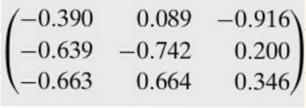
Change the basis

A. Form U

$$U = \begin{pmatrix} -.39 & -.639 & -.663 \\ .089 & -.742 & .664 \\ -.916 & .2 & .346 \end{pmatrix}$$

B take transpose

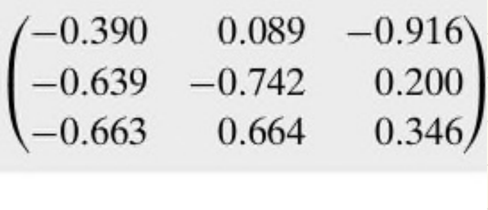
$U^T =$



$$\begin{pmatrix} -0.390 & 0.089 & -0.916 \\ -0.639 & -0.742 & 0.200 \\ -0.663 & 0.664 & 0.346 \end{pmatrix}$$

change of basis

get coods in new basis



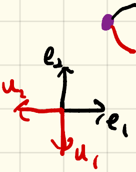
$$\vec{a} = U^T\vec{x} = \begin{pmatrix} -0.390 & 0.089 & -0.916 \\ -0.639 & -0.742 & 0.200 \\ -0.663 & 0.664 & 0.346 \end{pmatrix} \begin{pmatrix} -.343 \\ -.754 \\ .241 \end{pmatrix}$$

$\vec{x}$ in std basis

$$= \begin{pmatrix} -.154 \\ .828 \\ -.190 \end{pmatrix}$$

$\vec{a}$ our pt in basis of U_i's

Interpet geometrically



(1,2) → 1 step in e_1 dir
2 steps in e_2 dir

(-2,-1) → -2 steps in the u_1 dir
-1 step in the u_2 dir

orthonormal transformations
- rotations
- reflections

Consider sorting dim by "importence"

$\vec{u}_1, \vec{u}_2, ---, \vec{u}_d$ in descending order of importence

very important

not very important

reduce dim by dropping unimportandant dim

Consider

$\vec{u}_1, \vec{u}_2, --, \vec{u}_r$ these are the important dims

$r \ll d$

$\Rightarrow \vec{x} = a_1\vec{u}_1 + a_2\vec{u}_2 + --- + a_r\vec{u} + \underbrace{--- + a_d u_d}$

$\vec{x}' = a_1\vec{u}_1 + a_2\vec{u}_2 + \cdots + a_r\vec{u}_r$

(drop unimportant stuff projected onto first r dims)