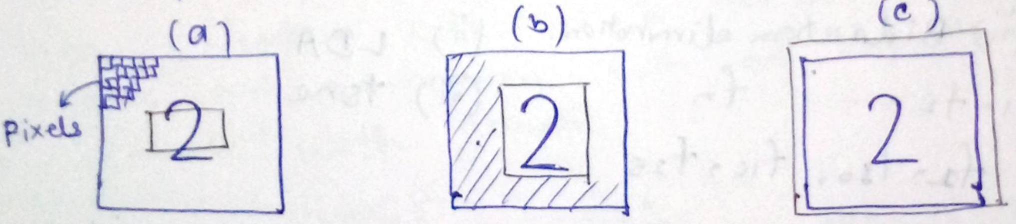
24-01-2024

~~Ps~~ Principle Component Analysis

Curse of dimensionality:

feature - ~~feature~~ columns in dataset

- also called curse of feature.
- always in any dataset there is a only optimal number of feature if we scape or beyond this then there is no effect on the model accuracy and efficiency.
- example in MNIST dataset



less efficiency. | let accuracy = 98%. | Here accuracy is also = 98%

- in the image (b) and (c) there is no effect on the accuracy because in (b) extra shaded region ~~negli~~ co negligible contribution in model efficient that are add in (c) part.
- ~~less~~ in (a) less efficient because focus only in a limited number of pixels.

☐ - model working region ☐ - 28x28 pixel image

- High dimension data may be found in image, text based data set.

- in the high dimension data set there is a high sparsity

- ~~The solution for this~~ is

- disadvantages of high dimension data
  * performance decrease
  * computation increase.

- The solution for this is
  * dimensionality reduction.

feature Selection
(i) forward selection
(ii) backward elimination
(iii) Bidirection elimination.

old: $f_1, f_2 ---- f_n$

new: $f_1, f_{10}, f_{16}, f_{25}$.

feature extraction.
(i) PCA
old: $f_1, f_2 ---- f_n$
new: $F_1, F_2, ---- F_n$
(ii) LDA
(iii) tsne

* PCA is feature extraction technique. it is used to reduced the curse of dimensionality.

* PCA Data के Behavior को Same रखते हुए Data को low dimension में लाता है।

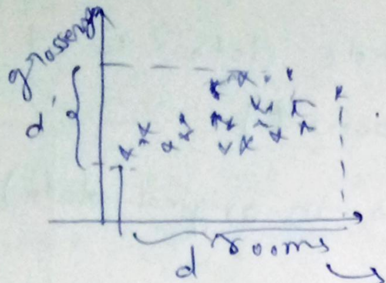
Benefits of using PCA:
- faster execture of Algorithm
- visualization

Geometric Intution:

No. of Rooms	No. of grocessory shop	price
3	2	60
4	0	130
5	6	170
2	10	90

in feature selection:

select the column rooms and price(L)

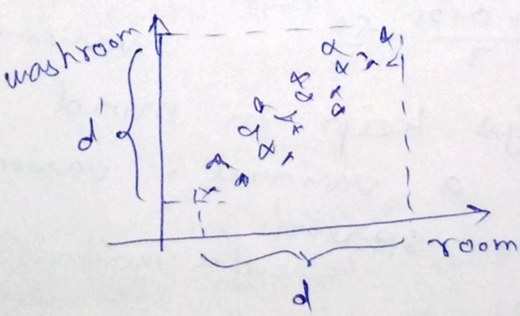


select this because it is more spread in axis (more variate select that)

here $d > d'$

in feature extraction:

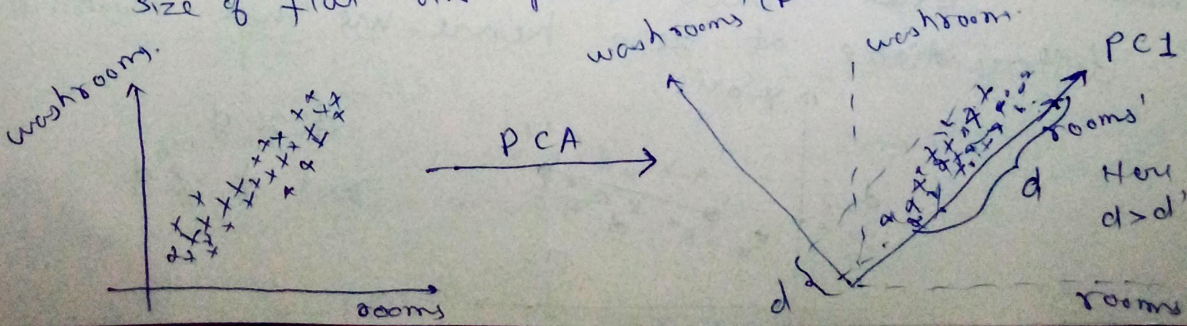
* let Here the columns are Number of rooms, Number of washrooms, price.
* Here feature selection is not use because rooms and washrooms both are important
* example with graph



Here $d = d'$

it means variance are equal for both washroom and room

* Here when we merge the grossery and washroom it mean size of flat as a new Column Here hence our new columns are size of flat and price

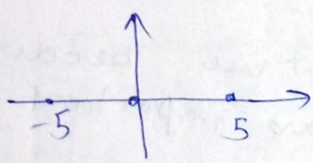


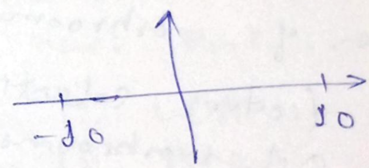
Here $d > d'$

- Here two principal components (PC) is formed 1 and 2. when PC variance is high hence we select it.
- Now ~~tra~~ transform whole data w.r.t PC1 (selected principle component)
- number of PC ≤ n (features in original data)

why variance is important:
- it is a statistical technique जो यह बताता है कि हमारे data का spread कितना है।

$$\text{variance} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$





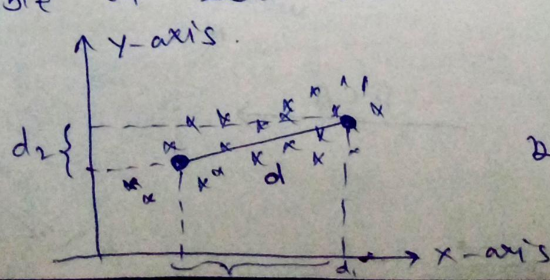
mean = 0 mean = 0

$\text{variance} = \frac{25+0+25}{3} = \frac{50}{3}$ $\text{variance} = \frac{100+0+100}{3} = \frac{200}{3}$

- it is always keep in mind spread is not always a variance. variance is propotional to spread.
  spread $\propto$ variance

Que. why we not use MAD (mean absolute deviation) in PCA

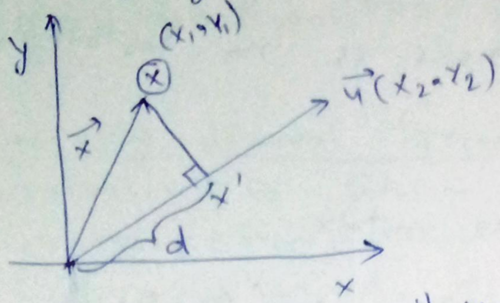
Answer ~~Bust~~ Because we ~~not~~ it is not differentiable at zero hence we use variance



$d_1 \approx d$

But $d_2 \not\approx d$

for given on point only



$$\frac{\vec{u} \cdot \vec{x}}{|u|} = \vec{u} \cdot \vec{x} = u^T x$$

$$= [x_1 \; y_1] \, [x_2 \; y_2]$$

$$= [x_1 \; y_1] \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

यहाँ पर हम उस unit vector को select करेंगे जिसमे variance maximum हो।

$$= x_1 x_2 + y_1 y_2 \rightarrow \text{scalar}$$

Same for other point $[u^T x_1], [u^T x_2] \cdots [u^T x_n]$

$$\text{variance} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Here $[u^T x_m] \longrightarrow$ for mean

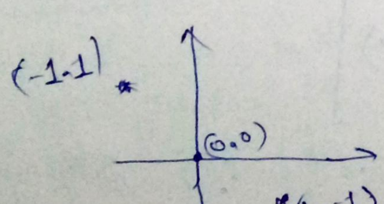
hence

$$\text{variance} = \frac{\sum_{i=1}^{n} (u^T x_i - u^T \bar{x})^2}{n} = \text{mathematical objective function}$$

in the above equation we are find the a value of $\vec{u}$ such that whole equation is maximize

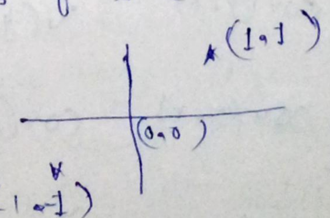
covariance and covariance matrix:

~~var~~ The problem with variance is that the it is not tell us that the what is the relation between the features of the data.



variance $= \frac{2}{3}$

co-variance $= \frac{-1+0-1}{3} = -\frac{2}{3}$

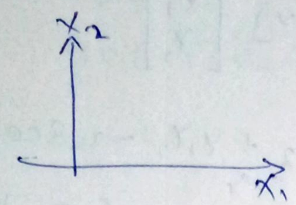


variance $= \frac{2}{3}$

co-variance $= \frac{1+0+1}{3} = \frac{2}{3}$

{ corelation is work in the rang -1 to 1
but in the co-variance there is no restriction in range }

covariance matrix :



it is 2X2 matrix

$$= \begin{bmatrix} cov(X_1, X_1) & cov(X_2, X_1) \\ cov(X_1, X_2) & cov(X_2, X_2) \end{bmatrix}$$

$$or = \begin{bmatrix} var\ X_1 & cov(X_2, X_1) \\ cov(X_1, X_2) & var\ X_2 \end{bmatrix}$$

but $cov(X_1, X_2) = cov(X_2, X_1)$

also we apply for high dimension data X_1, X_2, X_3

Benefits of covariance:
- हर axis पर data का spread कितना है
- किसी भी दो pair of axis पर relation कैसा है। or tell about data orientation.

for 3X3 matrix for a x, y, z columns.

$$\begin{array}{cc} x \\ y \\ z \end{array} \begin{bmatrix} V(x) & cov(x,y) & cov(x,z) \\ cov(y,x) & V(y) & cov(y,z) \\ cov(z,x) & cov(z,y) & V(z) \end{bmatrix}$$

$$\begin{array}{ccc} x & y & z \end{array}$$

Eigen decomposition of covariance matrix:

matrices linear transformation:

matrices is linear transformations in which make changes in our coordinate system eg. squeeze, expand, tilt

eigen vectors:

* it is a special vector in which their direction is not change after applying transformation, but magnitude is changes
* if we apply linear transformation in 2-D data then we only get two eigen vector.

$$\boxed{A\vec{V} = \lambda \vec{V}}$$

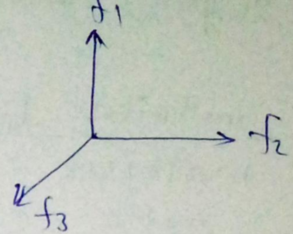
eigen values:

- eigen vector ~~कितना~~ strech या screenk कर रहा है
- example (-1, 1) to (-2, 2) here eigen value = 2

here A = matrix
$\vec{V}$ = eigen vector
λ = eigen value.

* the largest eigenvector of the covariance matrix always points into the direction of the largest variance of the data.
* covariance matrix का eigen vector और eigen value find in which जो सबसे बड़ा eigen value है उस vector के उपर ही variance ~~सबसे~~ सबसे ज्यादा आता है।

Step by Step Approach for PCA



(1.) Mean centring

(2.) find co-variance matrix

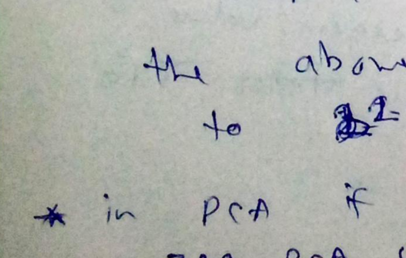
$$\begin{array}{c c} f_1 & \\ f_2 & \\ f_3 & \end{array}\begin{bmatrix} v(f_1) & cov(f_1, f_2) & cov(f_1, f_3) \\ cov(f_1, f_2) & v(f_2) & cov(f_1, f_2) \\ cov(f_3, f_2) & cov(f_3, f_2) & v(f_3) \end{bmatrix}$$

$$\quad f_1 \qquad f_2 \qquad f_3$$

(3.) Find eigen value and eigen vector for the covariance matrix. (here we get 3 set of vector corresponding value $\lambda_1, \lambda_2, \lambda_3$)
let then λ_1 is largest then λ_1 is principle component

(4.) Now transform the points with corresponding principle component (PC)

How to transform the point.



$$u^T \cdot x$$
$$= (3,1) \cdot (1000, 3)$$
$$= (1000, 1)$$

shape (1000, 3)

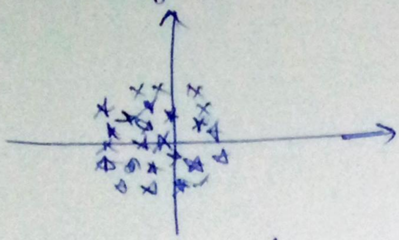
the above transformation example is ~~(3×1)~~ 3-D to 1-D data

* in PCA if data 784 column then we get the 784 PCA with eigen value $\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_{784}$

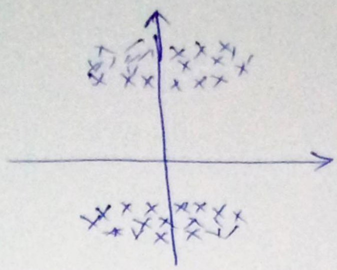
ये यह बता रहा है कि अकेले यह original data मे कितना variance explain कर पा रहा है।

when PCA does not work:

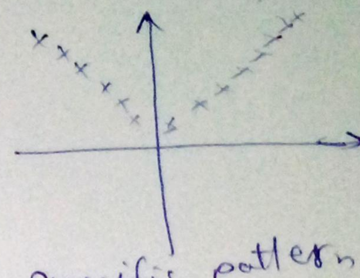
example.



circle ~~sph~~ shape



dumble. shape



Specific pattern

$y = x^2$

Or sin, cosine function