Page :
Date : / /

Dimensionality Reduction

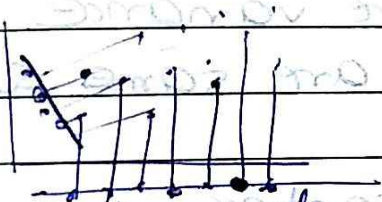
1) Required for data compression
2) Data compression use for speedup operation.

What is? Dimensionality Reduction
1) Reduce data from 2D to 1D

2) Data Visualization
3) Reduce the No. of coloums.

Principal Component Analysis (PCA)

ex)



Housing Data
size
Dimensaility reduction

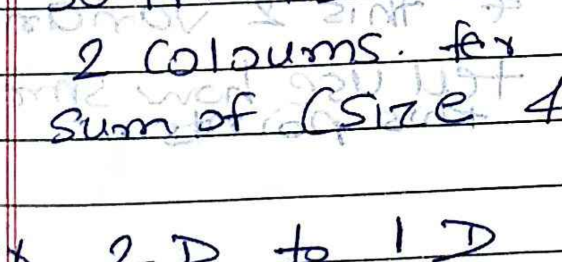
Number of rooms ⟶ Size Features.
No. of bathrooms.

So select correct projection.

Schools around ⟶ Location Features
Crime Rate.

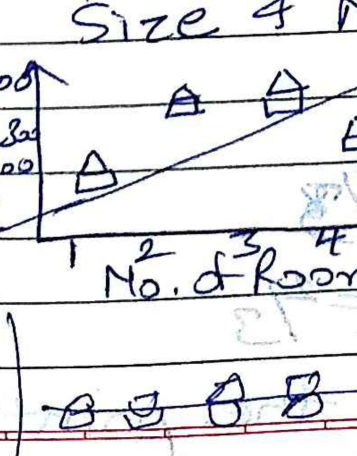
So in this reduce five colums data into 2 coloums. for ex:- ~~redu~~ taking weighted sum of (Size 4 ~~scho~~ area)

For ex 2D to 1D House Two features.
Size 4 No. of rooms.



Here we are not try to fit data we try to make good projection line

table of two coloum (dimension) become one. D.



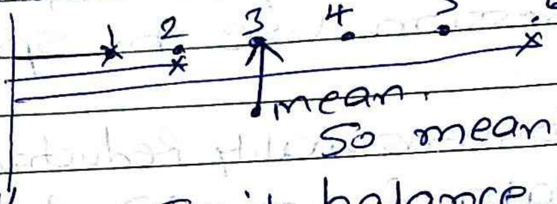
2D

Page :
Date : / /

2D size & No. of rooms → 1D size features.

Some statistics Basic Graphically

1) mean



So mean is $\frac{1+2+6}{3} = 3$

So it balance the weight.

2) Variance

$\frac{1^2+0^2+1^2}{3} = \frac{2}{3}$

1) Point at top are close less variance

$\frac{5^2+0^2+5^2}{3} = \frac{50}{3}$

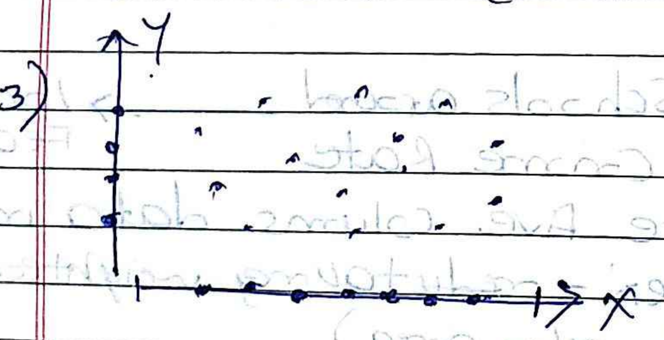
2) Point at bottom wide more variance.

3) Here for both mean are same so need to find variance.

4) take distance from center.

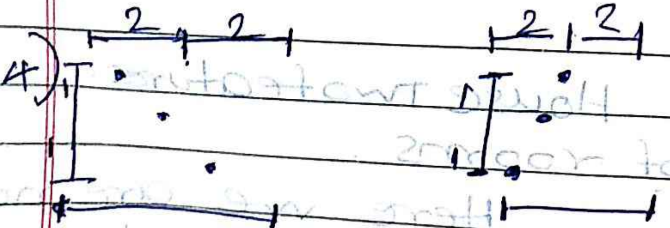
5) In case 1 ~~mea~~ variance is more than case 2

3)



find X-variance
Y-variance
& this 2 variance tell use how spreadout this points.

4)



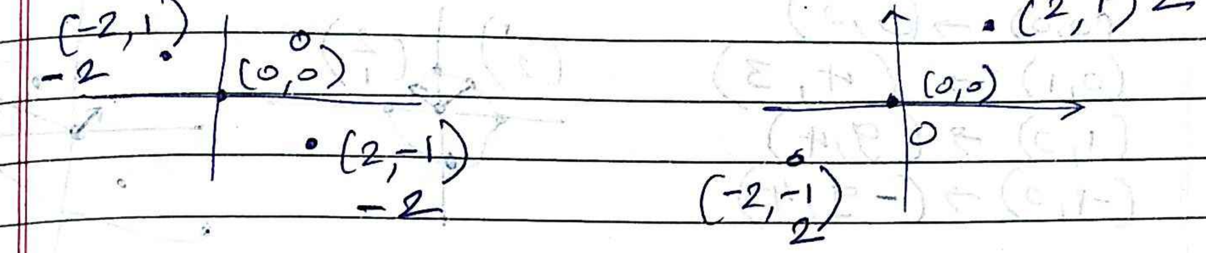
X-variance $= \frac{2^2+0^2+2^2}{3} = 8/3$

Y-variance $= \frac{1^2+0+1^2}{3} = 2/3$

Both are same but the points are on different projection.

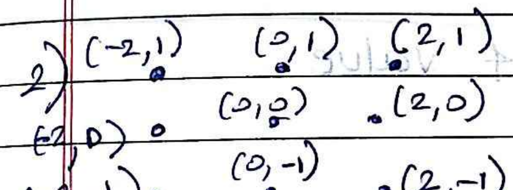
5) Covariance

1) Product of coordinate



$\therefore$ covariance $= \frac{(-2)+0(-2)}{3} = \frac{-4}{3}$ -ve coorelated

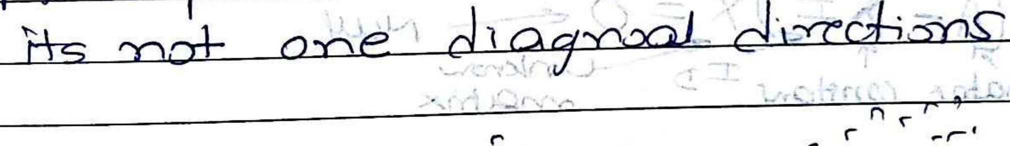
covariance $= \frac{2+0\,2}{3} = \frac{4}{3}$ +ve correlated.

2)


covariance $= \frac{-2+0+2+0+0+0+2+0+-2}{9} = 0$

* Direction of this points are middle

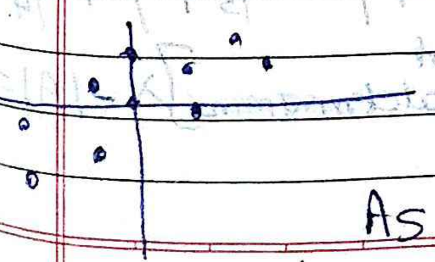
* This points are not +ve or -ve correlated. its not one diagnoal directions.

3)


negative covariance — covariance zero (very small) — Poistive covariance

6) Covariance Matrix

$$\Sigma = \begin{pmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{pmatrix}$$



for ex $\Sigma = \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$

As data is more on x directio : 9>3

+ve correlate : 4

Linear Transformation.

$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$

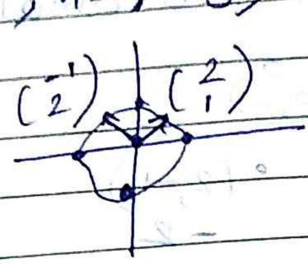
$(x, y) \rightarrow (9x + 4y, 4x + 3y)$

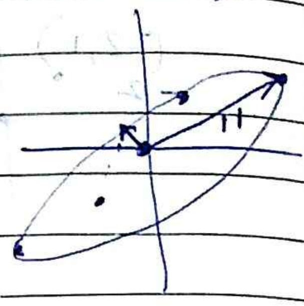
$(0,0) \rightarrow (0,0)$

$(0,1) \rightarrow (4,3)$

$(1,0) \rightarrow (9,4)$

$(-1,0) \rightarrow (-9,4)$





Eigenvectors (direction)

$\begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix}$

Eigenvalues (magnitude)

11 1

---

* How to find Eigenvectors & value.

$$A = \begin{bmatrix} 8 & -8 & -2 \\ 4 & -3 & -2 \\ 3 & -4 & 1 \end{bmatrix}$$

1) $(\underset{\text{given matrix}}{A} - \underset{\text{constant}}{\lambda} \underset{\text{ID}}{I}) \underset{\text{unknow matrix}}{X} = \underset{\text{Null}}{0}$

2) $|A - \lambda I| =$

$$\left(\begin{bmatrix} 8 & -8 & -2 \\ 4 & -3 & -2 \\ 3 & -4 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 8-\lambda & -8 & -2 \\ 4 & -3-\lambda & -2 \\ 3 & -4 & 1-\lambda \end{bmatrix} = 0$$

$\begin{vmatrix} -3 & -2 \\ -4 & 1 \end{vmatrix} = \begin{vmatrix} 8 & -2 \\ 3 & 1 \end{vmatrix} = 14 \begin{vmatrix} 8 & -8 \\ 4 & -3 \end{vmatrix}$

3) $1\lambda^3 - [\text{sum of Diagnoal Element}]\lambda^2 + [\text{sum of Diagnoal element minor}]\lambda - |A| = 0$

$\lambda^3 - 6\lambda^2 + 11\lambda - 6 = 0$

$\lambda = 1, 2, 3$ - Eign values.

Page :
Date : / /

* Eign Vectors,

If $\lambda = 1$ $\begin{bmatrix} 7 & -8 & -2 \\ 4 & -4 & -2 \\ 3 & -4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

By Crammer's Rule,

$7x_1 - 8x_2 - 2x_3$ — (1)

$4x_1 - 4x_2 - 2x_3$ — (2)

---

* PCA

1) Take data set center it

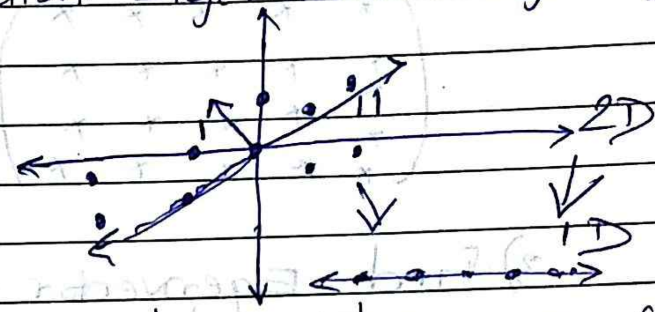
2) Find covariance matrix $\Sigma = \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$

3) Find Eigenvectors (direction) Eigenvalues (magnitude)

direction $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ (11) magnitude

direction $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ (1) magnitude

Eigenvectors (direction) Eigenvalue



1) This two lines are perpendicular because of symmetric matrix,
Also Eigenvectors are orthogonal - perpendicular

2) Eigenvalues are ~~a~~ real as matrix is symmetric

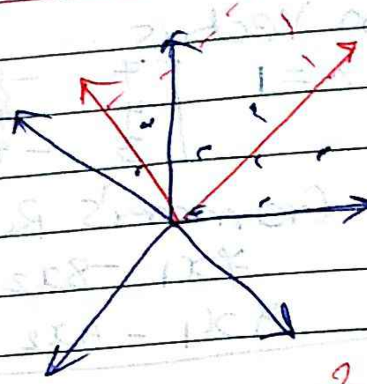
3) So we want to sumarize our data set in this 2 vector & which one more important
one with the largest eigenvalue
Here 11 is good, & only creare abut line

4) we extent that line & take projection on that line, ie the picture of data set.

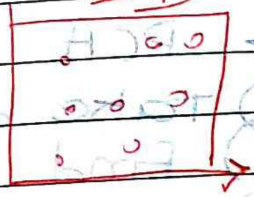
5) We picked projecting over the eigenvector with the highest eign value that means we project over the axis that carries the most amount of information

Largest Table.

X_1	X_2	X_3	X_4	X_5
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.



5 D to 2 D



1) So this 5 D datata.

2) Covariance Matrix

$$\begin{pmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{pmatrix}$$

3) Find Eigenvector & Eigenvalues.

V_1	λ_1	↑ Big
V_2	λ_2	
V_3	λ_3	
V_4	λ_4	
V_5	λ_5	Small

4) Pick up highest Eigen value ie PCA compont for ex 2.

* PCA :- Eigen Decomposition of the covariance matrix $(X^T X)_{n \times m} \longrightarrow$ W - Eignvectors, λ - Eignvalues.

Size of $W = (m \times m)$

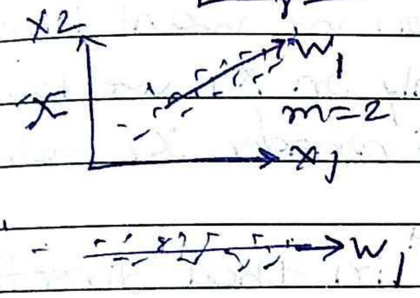
$$\underset{\substack{\text{Score} \\ (n \times m)}}{T} = \underset{(n \times m)}{X} \; \underset{(m \times m)}{W_{\text{Loading}}}$$

* Eeach Coloumn of W is a PCA

* Due to PCA we just transformed data we are not changing size we change projection

* W are ORDERED coloums by value of λ
  ∴ r = No of PCA component in coloumn.
  - W_r = first r coloumn of W.

$$\begin{bmatrix} \underbrace{w_1 \quad w_2}_{r} & \cdots & w_m \end{bmatrix} \qquad \underset{n \times r}{T_r} = X \underset{m \times r}{W_r}$$



$W = 2 \times 2$ size

$r = 1 \quad T = XW_1 = n \times 1$

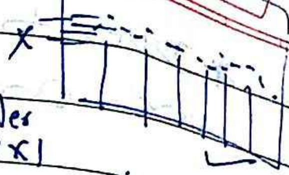
Why dimensionality Reduction.
1) Visualization
2) Reduce Noise
3) Preserve useful information in low memory
4) Less time complexity
5) Less space complexity
6) Analyze
7) We need to understand high dataset.

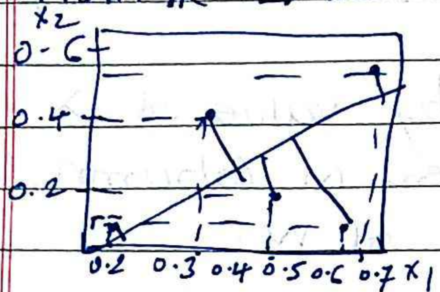
$$X \begin{bmatrix} & \\ & \end{bmatrix} n \text{ rows}$$

$\underline{m}$ measurments ← Higher Dimensional Data set

* Understand data means find patterns in data

For ex:- 1) 2D data convert into 1D
From $\mathbb{R}^2 \rightarrow \mathbb{R}^1$ there is data loss





x_1	x_2	consider only x_1
0.2	0.1	0.2
0.5	0.2	0.5
0.65	0.1	0.65
0.4	0.4	0.4
0.7	0.6	0.7

Here we loose some data, So when dimensionality reduce there is data loss.

* So in dimensionality reduction we want to minimize loss. So when we project data only on x_1 we loose another direction data x_2, If consider x_2 we loose x_1

* So mathematically reduce it in that direction in which the variance is maximized

* Come with a strait line & embed the data in that direction. So new space is one dimensional so direction of the line is not vertical or horizontal this line is linear combination of these two direction

* So PCA is find the line such that if we project data on that line, so projecting data means Project data on maxim variance that gives you the best spread.
    1) Least Loss of information
    2) Best capture the spread

Projecting point on line which gives max^m variance.

Page :
Date : / /

Principal Component Analysis

1) In project method we are interested in finding a mapping from the inputs in original d-dimensional space to a new (K < d) dimensional space, with min^m loss of informat^n

2) The projection of x on the direction of w

$$z = w^T x$$

3) PCA is an unsupervised method in that it does not use the output information. The criterion to be maximized variance.

4) The principal component is w_1 such that the sample, after projection on to w_1 is most spread out.

Finding direction of maximum variance

__Derivation__: For unique sol^n & make the direction we require $\|w_1\| = 1$

We seek w_1 such that $Var(z_1)$ is maximized subject to the constraint that $w_1^T w_1 = 1$

Writing this as Lagrange Problem.

$$\max_{w_1} w_1^T \Sigma w_1 - \alpha(w_1^T w_1 - 1)$$

Taking ~~W~~ derivative w.r.t w_1 & setting equal = 0

$$2\Sigma w_1 - 2\alpha w_1 = 0 \quad \therefore \Sigma w_1 = \alpha w_1$$

Page :
Date : / /

which means w_1 is an eigen vector of Σ &
α the corresponding eigen value.

Because we want to maximize $w_1^T \Sigma w_1 = \alpha w_1^T w_1 = \alpha$

* we chose the eigenvector with the largest eigenvalue for the variance to be maximum. i.e. The principal component is the eigenvector of the covariance matrix of the i/p sample with the largest eigenvalue $\lambda_1 = \alpha$.

The second principal component, w_2 should also maximize variance, be of unit length & be orthogonal to w_1. This latter requirement is to that after projection $z_2 = w_2^T x$ is uncorrelated with z_1. for the second principal component.

* Variance of each PC is given by λ_i
variance captured by first L PC $(1 \leq L \leq D)$

$$\frac{\sum_{i=1}^{L} \lambda_i}{\sum_{i=1}^{D} \lambda_i} \times 100$$

* Dimensionality Reduction Using PCA

1) Consider first L eigen value & eign vectors
2) Let W denote the $D \times L$ matrix with first L eigen vectors in the columns (sorted by λ's)
3) PC score matrix $Z = XW$
4) Each input vector $(D \times 1)$ is replaced by a shorter $L \times 1$ vector

Date: / /

* PCA Algorithm (Important) Points are center. means mean is zero

1) center X $\quad X = X - \hat{\mu}$

2) Compute Sample covariance matrix

$$S = \frac{1}{N-1} X^T X$$

3) Find eigen vectors + eigen value for S
4) W consist of first L eigen vectors as columns. $\therefore$ W is $D \times L$
5) Let $Z = \acute{X} W$
6) Each row in Z (or Z_i^T) is the lower dimensional embedding of X_i

* $W^T W = I$ Eeach eigen value is orthogonal to each other

$x_i = W z_i$ $\qquad W^T W = I$ orthogonal

$z_i = W^{-1} x_i \qquad z_i = W^T x_i$

(*) Recovering Original Data

1) Using W (loading matrix) + z_i (latent variable)

$$\hat{x}_i = W z_i$$

Average Reconstruction Error

$$J(W, Z) = \frac{1}{N} \sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2$$

Classical PCA Theorem
Among all possible orthonormal sets of L basis vectors, PCA gives the solution which has the minimum reconstruction error.

* optimal embedding in L dimensional space is given by $z_i = W^T x_i$

Page :
Date : / /

Finding Direction of maximal variance Line

1) Given any direction ($\hat{u}$), the projection of x on $\hat{u}$ is given by $x_i^T \hat{u}$

2) Direction of maximal variance can be obtained by maximizing $\frac{1}{N} \sum_{i=1}^{N} (x_i^T \hat{u})^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{u}^T x_i x_i^T \hat{u}$

$$= \hat{u}^T \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \right) \hat{u}$$

↑ Covariance matrix

3) Find max $\max_{\hat{u}\hat{u}^T=1} \hat{u}^T S \hat{u}$ $\hat{u}\hat{u}^T = 1$ constraint unit vector

where $S = \frac{1}{N} \sum_{i=1}^{N} x_i x_i^T$

S is sample (empirical) covariance matrix of the mean-centered data

4) Finding Directional maximum variance

$\max_{\hat{u}\hat{u}^T=1} \hat{u}^T S \hat{u}$ $\hat{u}\hat{u}^T$ = unit vector

where $S = \frac{1}{N} \sum_{i=1}^{N} x x^T$

$$L(\hat{u}, \lambda) = \hat{u}^T S \hat{u} - \lambda(\hat{u}^T \hat{u} - 1)$$

λ is Lagrange multiplier

$$\frac{dL}{du} = 2S\hat{u} - 2\lambda\hat{u} = 0$$

$$\underset{D\times D\ \text{matrix}}{S}\ \underset{D\times 1\ \text{value}}{\hat{u}} = \underset{\text{scalar}}{\lambda}\hat{u}$$ Need to find $\hat{u}$

$\hat{u}$ = Eigen vector of S λ is Eigen value

* The direction we want is the eigen vector with largest eigen value is 1^{st} PCA.