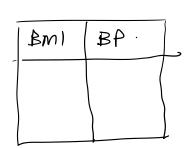
> Dimensionality Reduction > (1) PCA [Panapal Component Analysis] Consider Say beath of leason depends weight thought DBP SBP Health? 4 features weight ~ Syldic - Hught DBP Diastolic B.P. BP To represent this data > We need 4 Dimension For more features -> We had more Dimension. * Visualizing data in More than 3 dimension is difficult. Compulation on if features il also Complex Possible Solution: Reget weight DBP SBP Health-BP [Blood Preserve] [Boly Mars Index] BM! Height & weights Tare 1) look for Strong Correlated features: Heare DBP and SBP Also

Ele. His Correlation of Rught & Inevalt >BMI

> We can combine the Effective Correlation of Hught & Weight ⇒BMI

" DBP and SBP ⇒ BP.

Now Instead of 4 columns we have 2 columns.



= Thus is 2-15 data. Easy to Compute

Easy to Visualize.

We have Reduced the Dinemenon of data from 4-b to 2-D.

$$\begin{array}{ccc}
(SBP) & (SBP) \\
(SBP) & + & \times \\
(BP) & + & \times \\
\end{array}$$

X1 & AZ are weights of teatures X1 and X2 are features

 $\frac{P}{BP} = 0.8 DBP + 0.6 SBP$

wo are giving more weighte to DBP-as compared to SBP.

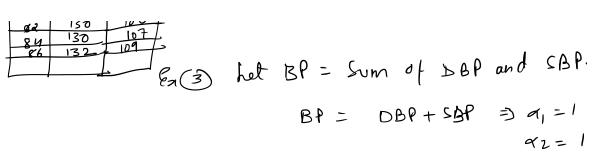
Ex 2 Let BP = mean of DBP & SBP.

$$BP = DBP + CBP = 0.5DAP + 0.5SBP$$

$$d_1 = 0.5$$

$$q_2 = 0.5$$

1 R DD - C - 11 LD and SBP.



For	Sum

SBP	Sum
(26=	204
120	208
127	208
130	212
130_	214-1
132	218-1
	4
	(26) 128 127 130 130

Mow PCA > Promupal Component Analysis

= 9t is a method to find the linear Combination that

accounts for as much variability as possible

in Combined Variable > (Maximum clasions)

To orderstand > y = \alpha_1 \times_1 \times_1 + \alpha_2 \times_2

(we would value of \alpha_1 and \alpha_2 such that the variance in

the value of y) Should be Maximum

gioz move the variance > Move is the information]

Note: > for y = \alpha_1 \times_1 + \alpha_2 \times_2

(combined)

Note: > for y = \alpha_1 \times_1 + \alpha_2 \times_2

(combined)

7 It there use n features, we will need 'n' weighte

gnoug Ex > we had a features! we need or, and orz

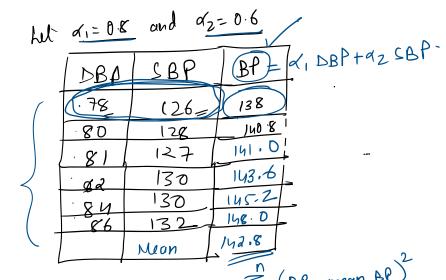
onehante $\alpha_1^1 + \alpha_2^2 = 1$

Let 01=08

 $(6.8)^2 + (6.6)^2 = 0.64 + 0.36 = \frac{1.00}{}$

$$(6.8)^2 + (6.6)^2 = 0.64 + 0.36 = \frac{1.00}{2}$$





Mean
$$\sqrt{142.8}$$

 $\sqrt{2}$ Variance = $\frac{1}{100} = \frac{1}{100} = \frac{1}{$

Let for above 6 input Sample.

> We can take

Many value of

of, and of 2

Such that of 2+12=1 sand (alculate vamana far

Ţ	L 1	42	\0x(A)	
	0.8	0.6	12.74	\mathcal{M}
ŀ	8.6	0.8	11.8	
	8.98	0.2	10.4	
	0.2	0.98	7.4	
	02			

Each α_1 and α_2 [we see for $\alpha_1 = 0.8$ and $\alpha_2 = 0.6$ of the above 4 Gases! we see for $\alpha_1 = 0.8$ we get hangest Variance = 12.74

Taxika 1 °° BP = 0.8 DBP + 0.65BP

To find d, and d2

we got our a, and a2

we got our a, and az

Touka 2 -> Now > (onsoler again >

	DBA	SBP
	78	(26=
_	80	128
Į	81	127 [
	g2	130
Ī	84	130
	86	132
-		
_	2	4

$$(ov(x,y) = \frac{1}{N-1} \ge (x_1^2 - x_2)(y_1 - y_2) = (ov(y,n))$$

For above Sample data set

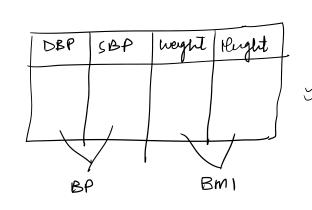
Covariance DBP
$$8.17$$
 5.97 5.97 5.97 5.97 5.97 5.97

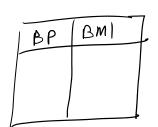
$$BP = -0.8DBP - 0.6SBP$$

$$BP = 0.8DBP + 0.6SBP$$

Applications >

(i) (oner der





Reduced Dimension by

(2)

	DBA	SBP	weight	Hught	Pulse
Pat (78	(26=	82	S-)	80
Pata	80	120	44	S-11	70
{	811	127	69	6.2	90
+	92	130	92	6-3	95
	86	132	85	<u>c.tb</u>	94-
7					/
	•	•			

Here if we have

Many No of

Realure and we

red to compare

the realth

chart parameters

of samples, it is

difficult

But if we combone the above 5 col in a col-

CI	(2	-,
VI	Y2	
V3	Yy	
75	V 6	

Now With less no of features (ompaning the Health chart / parameters of samples is relatively Easy:

* P(A > 91 is Dimensionally Reduction Technique.

Consider! For House Price Prediction

Size	Location	yoc	oc AW	Builder	(olor	gaeden	Swimming	Mela O Bistana	Pn'e

- * Here Price of House depends on above features.
- + Hure Many No of features (Problem of Plenty).
- * Ksues's x Represent Problem > Over filting Problem
- * We need to Reduce Dimension
- * We need to identify Impostant Components.

Dimengionality Reduction >.

* Reduces the dimension of feature Space

to if there are no features/col in dataset and you would to get only 10 features than with dimensionally reduction technique we can achieve this.

* It transforms dataset which is in n dimension space to n' dimension space where n'<n.

Why Dimensionally Reduction >.

- b Normally it is argued that many features gives

 More accurate result
 - T Boil after Some point the model performance decreases (overfitting) with increase in number of features.
 - * This is Cure of Dimensionality
 - * So Dinension Reduction is Concial.
- * PCA enables us to identify the correlation and pattern on the dataset so that it can be transformed into new dataset with lower dimension willow loss of important unformation.