* Dimensionality Reduction :-

→ Need :- We can visualise data in 2D,3D using Scatterplots
" " " " in 4D,5D,6D " pairplots
For higher dimensions pairplot wont work as it creates $^nC_2$ plots

Q) How to visualise data then for high dimensional data like MNIST ??

Soln $nD \rightarrow 2D$, or $3D$

But How ?? ~~using~~ dimensionality reduction using
1. PCA
2. t-SNE

* PCA (Principal Component Analysis) :-

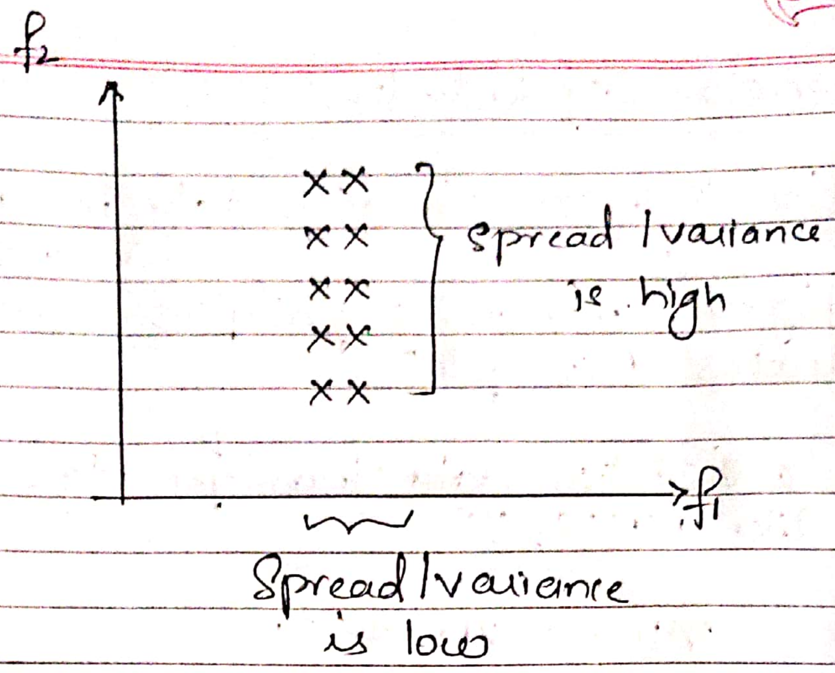
Task :- $D_{dim} \rightarrow D'dim$ $(D' << D)$

For simplicity lets consider a case of $2D \rightarrow 1D$
Later using Linear Algebra we will scale all of this to higher dimension

eg:- Consider data of Indian's having 2 features

$$\text{Indians}(X) = \begin{array}{c c} & \begin{array}{cc} f_1 & f_2 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ n \end{array} & \begin{bmatrix} & \\ & \\ & \\ & \\ & \end{bmatrix} \end{array}$$

where f_1 = blackness of hair
f_2 = height in cms.



Q) If i want to drop 1 feature which feature should I choose to drop.

Answer Offcourse f_1 (Reason! – Because the spread is very low $\therefore$ dropping f_1 will not lead to loss of crucial information)

Is that logically correct ?? Yes...

→ Almost all Indians have black hair; $\therefore$ there will be less variation in this feature. But f_2 which is height can vary alot among different Indians.

Hence we have done $2D \rightarrow 1D$.

$$X = \begin{array}{c c} & \begin{array}{cc} f_1 & f_2 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ n \end{array} & \left[\begin{array}{cc} \; & \; \\ \; & \; \\ \; & \; \\ \; & \; \\ \; & \; \end{array} \right]_{n \times 2} \end{array} \Longrightarrow X' = \begin{array}{c c} & f_2 \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ n \end{array} & \left[\begin{array}{c} \; \\ \; \\ \; \\ \; \\ \; \end{array} \right]_{n \times 1} \end{array}$$

Conclusion:- We can reduce the dimension of data by dropping the feature with minimum spread and preserving the feature with maximal spread.

Q) What if the scales are different for 2 features

eg:- f_1 = height , f_2 = Salary , target = weight

Which feature will you drop ???

→ · This is wrong, height is an important feature in determining weight.

Hence Our approach looks misleading...

Then How to take care of this:-

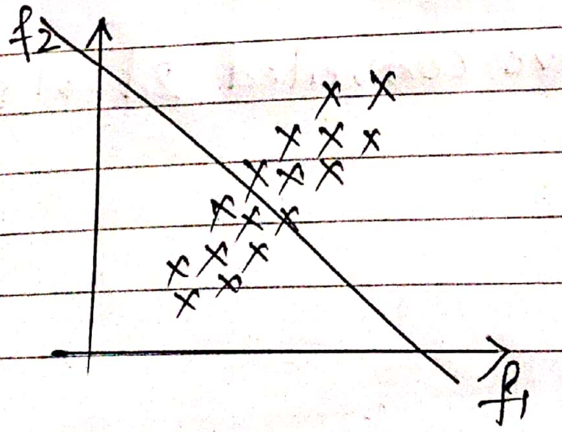
→ Column Standardization:-

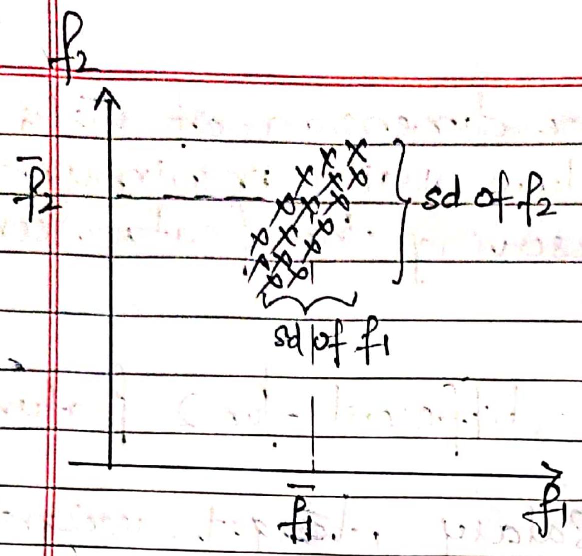
$$a_{i\,std} = \frac{a_i - \bar{a}}{s}$$

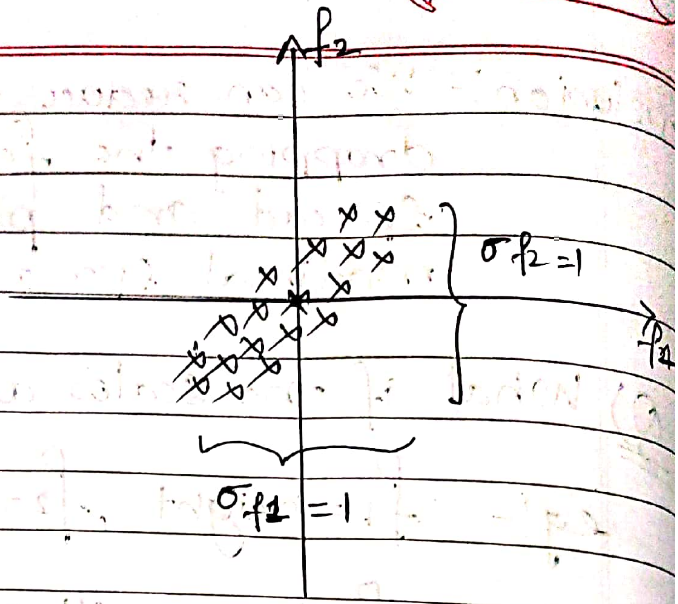
$\bar{a}$ = mean of a_i's

s = std deviation of a_i's

Transforms the vector such that its mean = 0 and std = 1

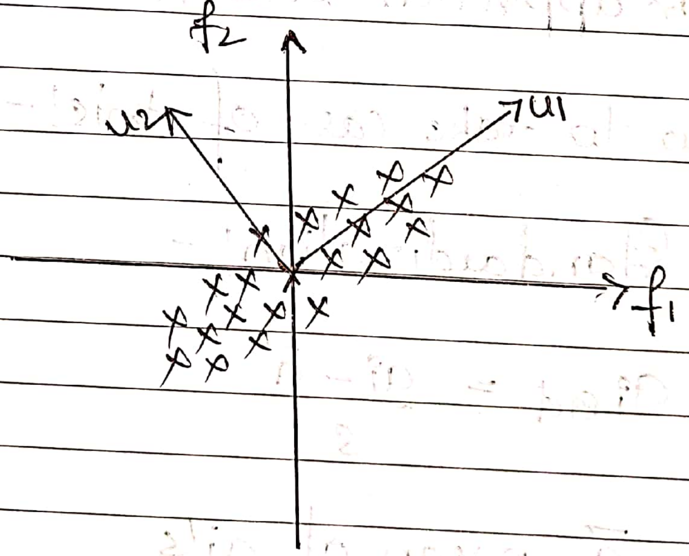






Q) Now which column to drop. Both have same variance.

Solution 1. find a direction with maximal spread u_1
2. find a direction $\perp^r$ to maximal spread u_2



And drop u_2 $\because$ $var(u_2) < var(u_1)$

Done ☺. we have converted $2d \to 1d$

Geometrically all this is fine. Mathematically how to find u_1.

$\lambda u_1 = S u_1$. (This looks similar to the eqn of eigen values)

i.e.:- $\lambda v_1 = S v_1$

where λ = eigen values of $S_{d \times d}$:- $\lambda_1, \lambda_2, \lambda_3 \cdots \lambda_d$ with corresponding eigen vectors $v_1, v_2 \cdots v_d$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_d$

S = covariance matrix of size $d \times d$ (i.e.:- for d features)

$$S = \frac{X^T X}{n}$$

$\therefore$ Finally u_1 is nothing but eigen vector v_1 of $S_{d \times d}$ corresponding to λ_1. (i.e.:- corresponding to maximal eigen value).

Q I got u_1 i.e.:- $u_1 = v_1$; But how to find u_2

Soln Properties of eigen vector :- $v_i \perp^r v_j$ for all $i \neq j$

$\therefore$ $v_2 \perp^r v_1$ (also $u_2 \perp^r u_1$)

$\therefore$ $\boxed{u_2 = v_2}$

Conclusion:- $v_1 \rightarrow$ direction with maximal variance

$v_2 \rightarrow$ direction with 2^{nd} most maximal "

$v_3 \rightarrow$ " " 3^{rd} " "

⋮

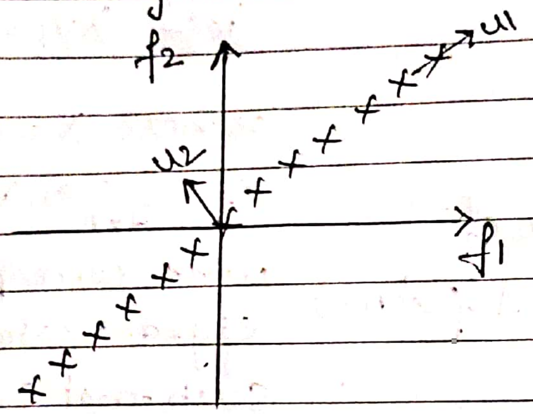
so on...

→ Till Now we have seen that eigen vectors play a crucial role in finding the direction of maximal variance. Is there any use of eigen values (λ) ???

Answer- Yes.. Eigen values λ can help us determine % var explained

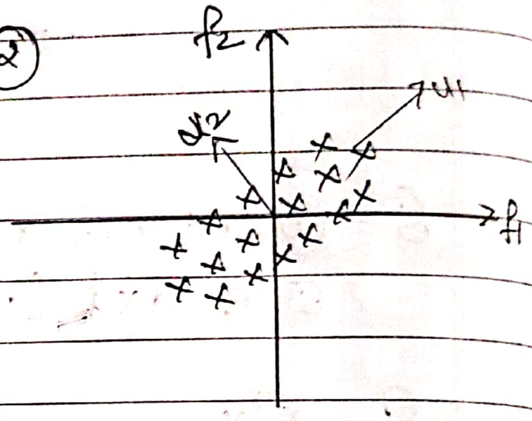
Lets try to understand this with few diagrams.

①



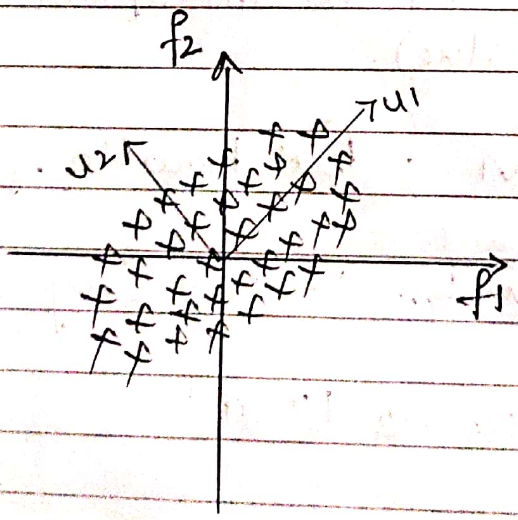
eg:- $\lambda_1 = 3$, $\lambda_2 = 0$ } variance explained in respective directions

②



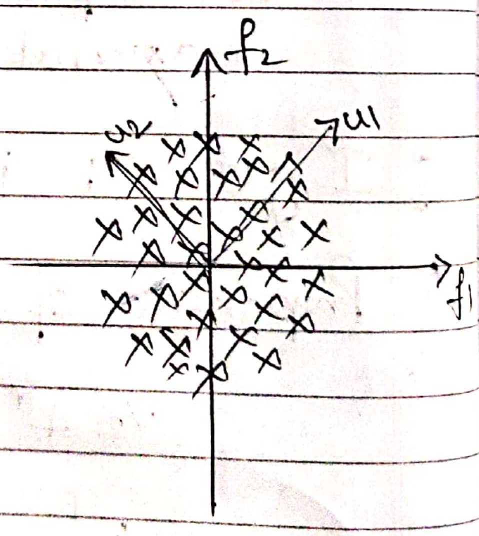
$\lambda_1 = 3$
$\lambda_2 = 1$

③



$\lambda_1 = 3$
$\lambda_2 = 2$

④



$\lambda_1 = 3$
$\lambda_2 = 3$

% variance explained $= \frac{\lambda_i}{\sum_{\forall i} \lambda_i}$

Case1 :- $PVE(\lambda_1) = \frac{3}{3+0} = 1$ 100% of variance is explained by λ_1 itself.

$PVE(\lambda_2) = \frac{0}{3+0} = 0$ 0% of variance is explained by λ_2

Case2 :- $PVE(\lambda_1) = \frac{3}{3+1} = \frac{3}{4} = 0.75$ 75% of variance is explained by λ_1

$PVE(\lambda_2) = \frac{1}{3+1} = \frac{1}{4} = 0.25$ 25% of variance is explained by λ_2

Case3 : $\frac{3}{5}$, $\frac{2}{5}$

Case4 $\frac{1}{2}$, $\frac{1}{2}$

Conclusion :- λ tells us how much information is left over after reducing the dimension

Drawback of PCA :-

Observe Case1 – Case4. PCA works well when data is linear. for case4 where data is non-linear losing any dimension will tend to loose 50% of overall information.

Applications of PCA:-

1. Dimensionality reduction
2. Visualization

Scanned with CamScanner