Data Matrix, $S = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}$ ↓ No. of samples ;

$N = 3 \leftarrow$ no. of samples
$M = 2 \leftarrow$ no. of attributes (each attribute being a RV)

$C = \text{Covariance of } S = \text{Cov}(S)$

$= \dfrac{S^t S}{N-1} = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \equiv \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} = \begin{pmatrix} c_1 & c_2 \end{pmatrix}$

where $\sigma_{xy} = E\big((x-\mu_x)(y-\mu_y)\big) = E(xy) - \mu_x \mu_y$

Here the data matrix S is already "centered" i.e. the column vectors of S are mean subtracted (you may verify that this in this case by checking that each column had mean "0").

We then solve the eigen system for matrix C by solving

$$C\vec{u} = \lambda \vec{u}$$

λ is found by solving $|C - \lambda I| = 0 \Rightarrow \begin{vmatrix} 1-\lambda & -\frac{1}{2} \\ -\frac{1}{2} & 1-\lambda \end{vmatrix} = 0$

$\Rightarrow (1-\lambda)^2 - \frac{1}{4} = 0$

$\Rightarrow 1 - 2\lambda + \lambda^2 - \frac{1}{4} = 0$

$\Rightarrow 4 - 8\lambda + 4\lambda^2 - 1 = 0$

$\Rightarrow 4\lambda^2 - 8\lambda + 3 = 0$

$\Rightarrow \lambda = 3/2, 1/2$
$\qquad \uparrow \qquad \uparrow$
$\qquad \lambda_1 \quad \lambda_2$

the evs $\vec{u}$ corresponding to λ_1 are is called $\vec{u}^{(1)}$ & λ_2 is called $\vec{u}^{(2)}$.

for this example,

$\vec{u}^{(1)} = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\vec{u}^{(2)} = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$\quad = \begin{pmatrix} v_1 \\ u_2 \end{pmatrix} \qquad\qquad = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$

$$U = \begin{pmatrix} \vec{u}^{(1)} & \vec{u}^{(2)} \end{pmatrix} \Rightarrow U U^{\dagger} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I \quad \underline{pg(2)}$$

$$\Rightarrow U \text{ is unitary!}$$

Now, since $C \vec{u}^{(1)} = \lambda_1 \vec{u}^{(1)}$

$$\Rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \frac{3}{2} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \vec{0}$$

$$\Rightarrow \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \frac{3}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \vec{0}.$$

$$\underset{\underset{\text{manipulation}}{\text{arithmetic}}}{\text{By}} \Rightarrow 1 \underset{\underset{C_1}{\uparrow}}{\begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}} - 1 \underset{\underset{C_2}{\uparrow}}{\begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}} - \frac{3}{2} \underset{\underset{\lambda_1}{\uparrow}}{\begin{pmatrix} 1 \\ -1 \end{pmatrix}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \underset{\underset{\frac{1}{\sqrt{2}}\vec{u}^{(1)}}{\text{from}}}{\Big\}}$$

Similarly,

$$C \vec{u}^{(2)} = \lambda_2 \vec{u}^{(2)} \text{ gives us}$$

$$1 \underset{\underset{C_1}{\uparrow}}{\begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}} + 1 \underset{\underset{C_2}{\uparrow}}{\begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}} - \frac{1}{2} \underset{\underset{\lambda_2}{\uparrow}}{\begin{pmatrix} 1 \\ +1 \end{pmatrix}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \underset{\underset{\frac{1}{\sqrt{2}}\vec{u}^{(2)}}{\text{from}}}{\Big\}}$$

Note $C_1 - C_2 = \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix} - \begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} 3/2 \\ -3/2 \end{pmatrix} = \frac{3}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$$= \left(C \vec{u}^{(1)} \right) \sqrt{2}$$

and $C_2 + C_2 = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \left(C \vec{u}^{(2)} \right) \sqrt{2}$

Pictorially / Graphically / Geometrically

$$C_1 - C_2 = (C\vec{u}^{(1)})\sqrt{2}$$
$$= \lambda_1 \vec{u}^{(1)} \sqrt{2}$$

$$C_1 + C_2 = (C\vec{u}^{(2)})\sqrt{2} = (\lambda_2 \vec{u}^{(2)})\sqrt{2}$$
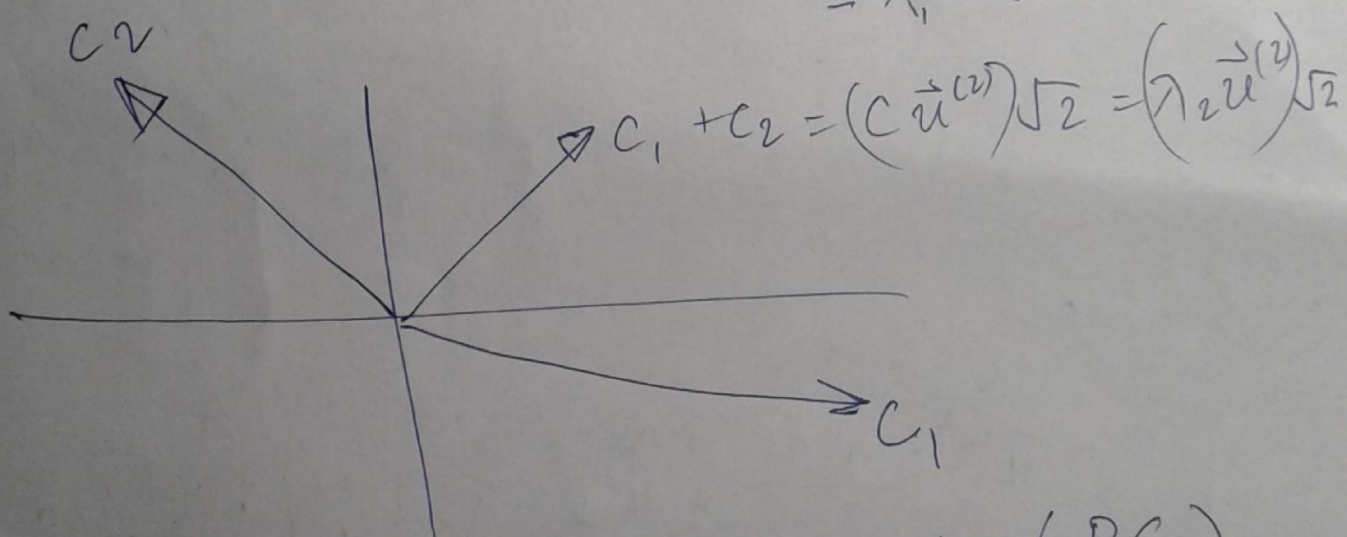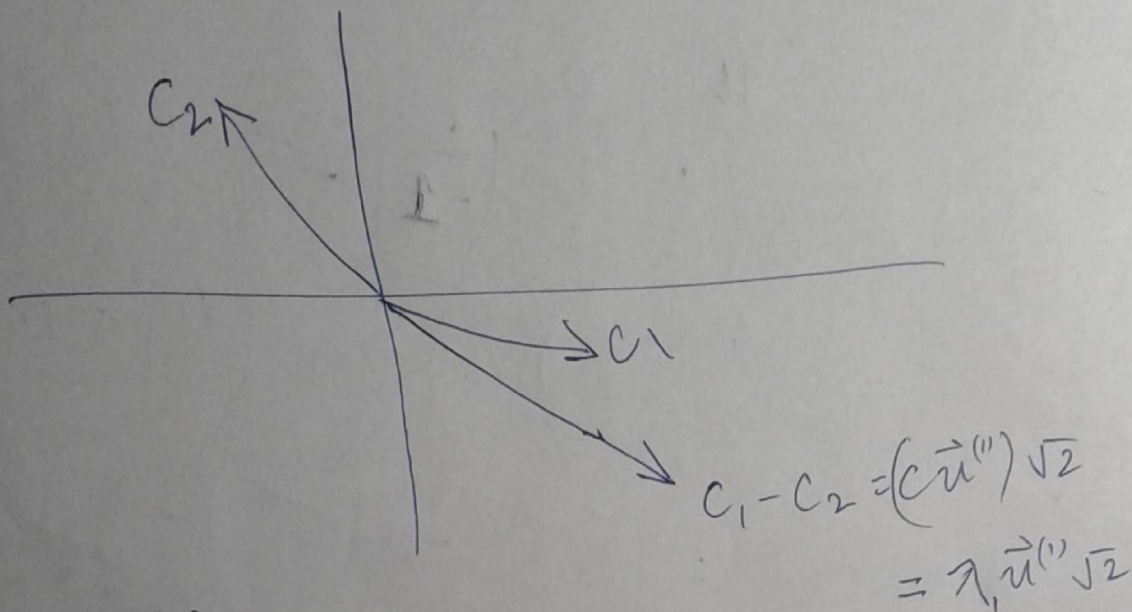
Computing principal Components. (PCs).

PCs are computed by multiplying the components of each EV by the attribute vectors & summing the results.

i.e. $P_1 = u_1 X + u_2 Y$

$P_2 = v_1 X + v_2 Y$

$$\therefore \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

and $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$P_1 = \frac{1}{\sqrt{2}} (X - Y) = \frac{1}{\sqrt{2}} \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}$$

$$P_2 = \frac{1}{\sqrt{2}} (X + Y) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

$$P = \begin{pmatrix} P_1 & P_2 \\ 1 & 1 \end{pmatrix} \quad \text{can also be}$$

computed by $\quad P = SU$

$$= \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ -1 & -1 \end{pmatrix}$$
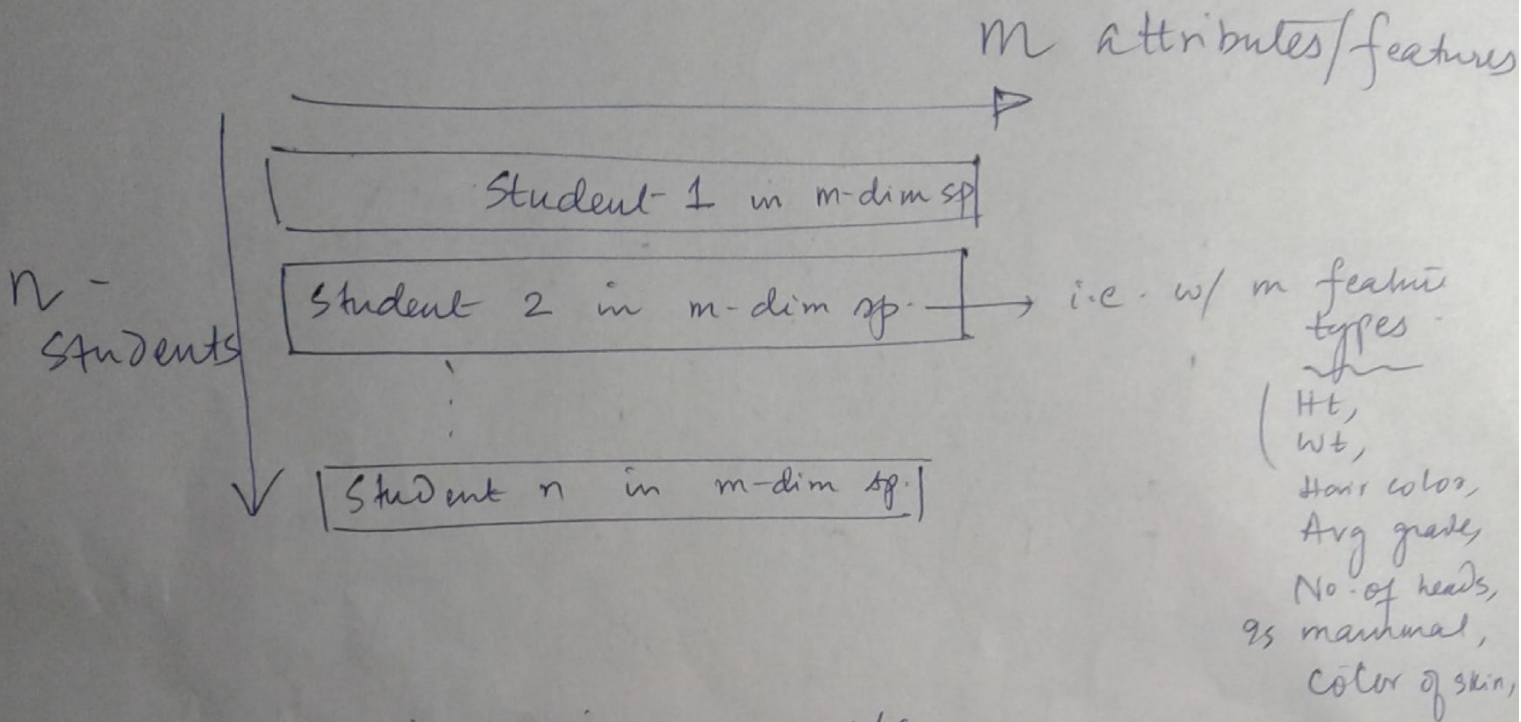
$$P^T P = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

gives us the evs as diagonal terms.

$$PU^T = SUU^T = S = \begin{pmatrix} 1 & 1 \\ X & Y \\ 1 & 1 \end{pmatrix}$$

i.e. the Data Matrix can be recovered from the principal components (Matrix) by simply PU^T.

i.e for this example $\vec{X} = \frac{1}{2}(P_1 + P_2)$ & $\vec{Y} = \frac{1}{2}(P_2 - P_1)$

Why do we perform PCA?

```
                                    m attributes/features
                    ┌──────────────────────────────────►
              ┌─────│────────────────────────────────┐
              │     │  Student 1 in m-dim sp          │
  n -         │     ┌──────────────────────────────┐──┼──► i.e. w/ m features
  students    │     │ Student 2 in m-dim sp ───────┤  │        types
              │     └──────────────────────────────┘  │        ___
              │              ⋮                         │       ( Ht,
              ▼     ┌──────────────────────────────┐   │         Wt,
                    │ Student n  in  m-dim sp. │         Hair color,
                    └──────────────────────────────┘     Avg grade,
                                                          No. of heads,
                                                       as mammal,
                                                          color of skin,
                                                            ⋮
                                                             :)
```

Clearly in this example,
you may guess that
No. of heads and whether or not
she/he is a mammal? is a ~~usef~~
useless feature/attribute b/c it will
be the same for all students so
we may want to find those
attributes which are "less" useful in
identifying a student & eliminate
them in our statistical analysis.

eg of features we may ignore?
 ① Constant parameter (# heads)
 ② " " w/ some noise (or low variance)
                              eg. thickness of hair
 ③ Parameters that are linearly dependent on
                              other parameters

Which parameters do we wish to
keep?

→ parameters that do not depend
on other parameters
(low covariance σ_{XY})

→ High variance (parameters that
change a lot).
↑
high entropy
⇒ more information

Questions

Q1) How do we describe "most important"
features using math?

Ans) Variance.

Q2) How do we represent our data
so that the most important features
can be extracted easily?

Ans) Change of basis -

↦ Essence of PCA machinery!

Let-

$X' = X - \bar{X}$

Step ① mean subtraction

$$S_0 = \begin{pmatrix} X_1 & X_2 \\ 2.5 & 2.4 \\ 0.5 & 0.7 \\ 2.2 & 2.9 \\ 1.9 & 2.2 \\ 3.1 & 3.0 \\ 2.3 & 2.7 \\ 2.0 & 1.6 \\ 1.0 & 1.1 \\ 1.5 & 1.6 \\ 1.2 & 0.9 \end{pmatrix}$$

$\bar{X_1} = 1.81$

$\longrightarrow$

$\bar{X_2} = 1.91$

$$S = \begin{pmatrix} X_1' & X_2' \\ 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{pmatrix}$$

Step ② Compute Covariance matrix:

$$C = \frac{S^t S}{n-1} = \frac{1}{9}(S^t S) = \begin{pmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}$$
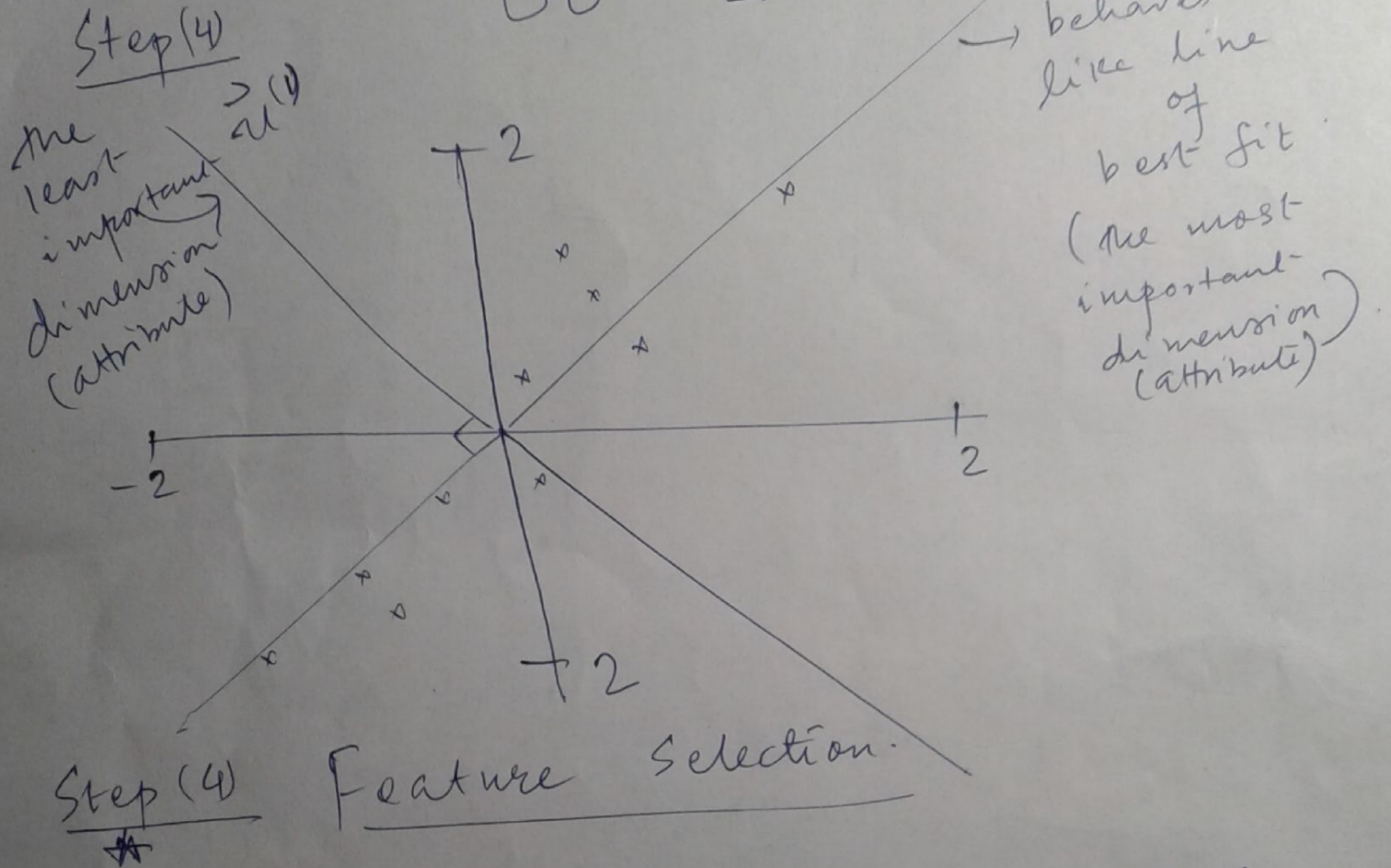
Step ③ find evs & EVs of C

$\lambda_1 = 0.0491, \quad \lambda_2 = 1.2840.$

$$\vec{u}^{(1)} = \begin{pmatrix} -0.7352 \\ 0.6779 \end{pmatrix}; \quad \vec{u}^{(2)} = \begin{pmatrix} -0.6779 \\ -0.7352 \end{pmatrix}$$

Note $\vec{u}^{(1)} \perp \vec{u}^{(2)}$

i.e. $U = \begin{pmatrix} | & | \\ \frac{1}{|}\vec{u}^{(1)} & \frac{1}{|}\vec{u}^{(2)} \\ | & | \end{pmatrix}$

$$UU^{\dagger} = I$$

Step (4)

the least important dimension (attribute) $\vec{u}^{(1)}$

$\vec{u}^{(2)}$ → behaves like line of best fit (the most important dimension (attribute))

Step (4) Feature Selection

(i) the eigen vector (EV) w/ the highest- eigenvalue is the principal component of the data set.

(ii) Once EVs & evs are found from C; next step is to order them in descending order ~~decreasing~~

$$\lambda_{①} > \lambda_{②} > \lambda_{③} > \cdots$$

In this example

$$\lambda_{(1)} = \lambda_2$$

$$\lambda_{(2)} = \lambda_1$$

Note we can always
recover the data

$$S = P U^t$$

Step (5)   Reduced Data set

$\longrightarrow$ m dimensional original data.

$\rightarrow$ Calculate m EVs & evs.

$\longrightarrow$ choose r largest evs (EVs)

$\rightarrow$ Reduced (final) data has r dim.

Note :-

(i)   If dim are highly correlated;
there will be small no. of EVs
w/ large evs & $r \ll m$

(ii)   If dim. are not correlated;
$r \approx m$ (PCA does not help).

In our example; throwing out the
useless dimension would mean projecting
final data values on the vector $\vec{u}^{(2)}$
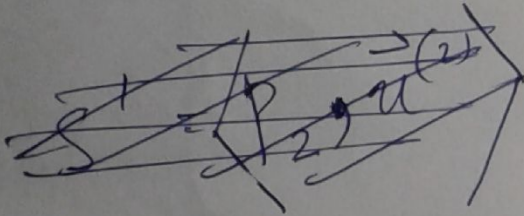
$$P_1 = u_1 X_1 + u_2 X_2$$

$$P_2 = v_1 X_1 + v_2 X_2$$

↳ most imp. data.

$$P = \begin{pmatrix} | & | \\ P_1 & P_2 \\ | & | \end{pmatrix} \longrightarrow \begin{array}{l}\text{principal}\\ \text{component - matrix}\\ \text{(transformed dataset)}\end{array}$$

$$S = P U^{t} \quad \text{will give me original data}$$

But if 9 "Kill"/eliminate some of the useless dim.; the data recovered will not be identical to original data (but almost similar)

$$S_{reduced} = P_2 \left(\vec{u}^{(2)} \right)^{t} \underset{10 \times 1}{\longrightarrow} \begin{array}{l}\text{will give me}\\ \text{data projected}\\ \text{onto } \vec{u}^{(2)} \text{ & will}\\ \text{look like}\end{array}$$

\#