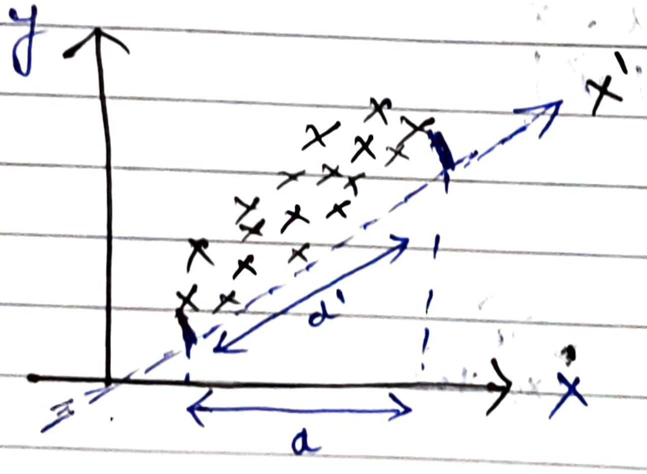
Problem formulation

- You have to find that line in 1D, which will give as good results it would give in 2D

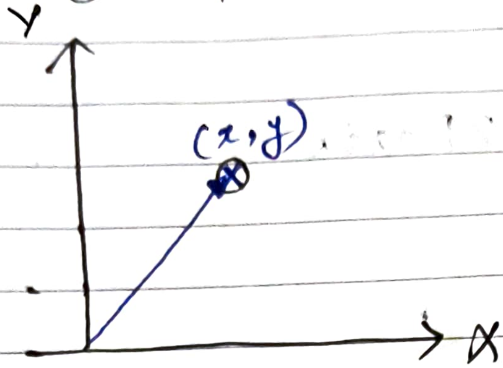
- Maybe, this axis can be that line x'



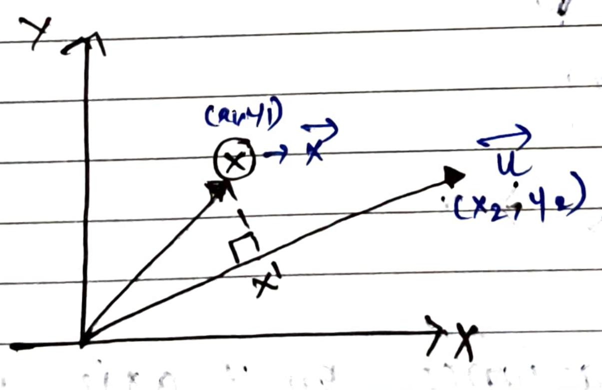
Because the variance on x axis is almost same as on x' axis

Exactly what does PCA solve understanding with single pt.

- Single pt



- ⊗ → This pt is a vector $\vec{x}$
- You have to find a unit vector $\vec{u}$
- Project vector $\vec{x}$ on $\vec{u}$
- Projection = $\dfrac{\vec{u} \cdot \vec{x}}{|\vec{u}|}$

- 

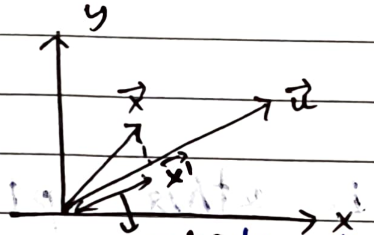
- Since our $\vec{u}$ is unit vector
  ∴ projection of $\vec{x}$ on $\vec{u}$
  is ⇒ $\dfrac{\vec{u} \cdot \vec{x}}{1} = \vec{u} \cdot \vec{x}$
- $\vec{u} \cdot \vec{x} \Rightarrow \vec{u}^T \vec{x}$
- After projecting $\vec{x}$ on $\vec{u}$, you get new vector. you can call it $\vec{x}'$

- $\vec{x}$ is vector with components x_1, y_1
- & $\vec{u}$ " " " " x_2 y_2

⇒ Projection of $\vec{x}$ on $\vec{u}$ = $\boxed{u^T x}$

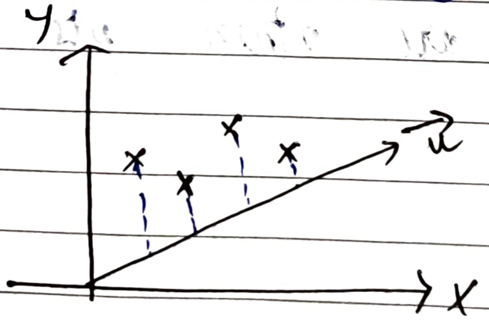
$u^T = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ $\quad x = [x, y]$

- Dot product of $u^T x = x_1 x_2 + y_1 y_2$
which will be scalar / length.



scalar / length of project of $\vec{x}$ on $\vec{u}$

- Similarly, you will project other points on that unit vector ($\vec{u}$)

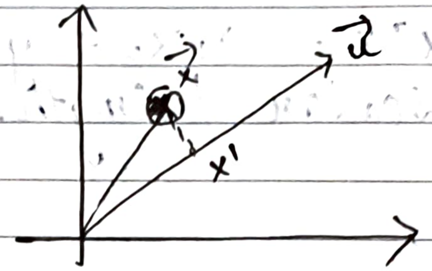


- And from each pt projection you will get scalar which will be length of projection of pt on our unit vector $\vec{u}$.

- Now we need to maximize the variance.

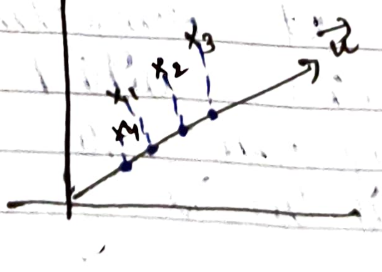
- There could be any unit vectors But we will choose that unit vector for which our variance is maximum.

- <u>Calculating variance using same example.</u>



$\rightarrow x' = u^T x$

- ~~So for~~ There will be other pts too
- for $\vec{x}$; we got $u^T x$
- Let's say this $\vec{x}$ is pt x_1.
- Projection of x_1 on $\vec{u}$ $\Rightarrow$ $u^T x_1$
- Similarly there are other pts x_2, x_3, x_4
- for $x_2 = u^T x_2$
  " $x_3 = u^T x_3$
  " $x_4 = u^T x_4$
  $x_n = u^T x_n$

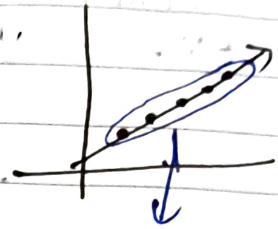


- Now we have to calculate variance

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

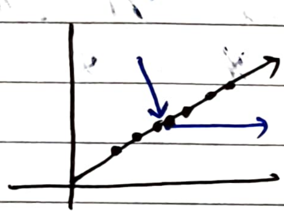
- X_i → each of individual projections → $(u^T X_i)$
  ~~for x_i ⇒~~ X

- for $i = 1$ ⇒ $u^T x_1$
  $i = 2$ $u^T x_2$
  & all these are scalar



we are trying to calculate variance of this projection

- $\bar{X}$ → mean/centre of projection



centre of projection ⇒ $u^T x_{m \to mean}$

⇒ Variance of projection of all points

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

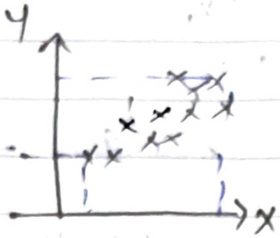
$$\boxed{= \frac{1}{n} \sum_{i=1}^{1} (u^T x_i - u^T x_m)^2}$$

- Now, mathematically we have to find a unit vector for which our variance of projection is maximum.

- Optimization problem → maximizing the variance.

Problem with variance

- Variance is calculated for 1 axis.
- Ex –



- Variance on x axis will different
  " " y axis "

- Variance will not tell the collective relationship between 2 or more axis.

- That's where covariance comes which tells the relation b/w x and y.

- **Covariance**

Ex ↑ We have 2 col x_1 & x_2

- It's covariance will be of [2 X 2]

$$
\begin{array}{c|cc}
 & x_1 & x_2 \\
\hline
x_1 & cov(x_1, x_1) & cov(x_1 x_2) \\
x_2 & cov(x_2, x_1) & cov(x_2, x_1)
\end{array}
$$

Covariance Matrix

$$\begin{array}{c|c} & \\ X_1 & \multirow{2}{*}{$\begin{bmatrix} cov(x_1,x_1) & cov(x_2,x_1) \\ cov(x_1,x_2) & cov(x_2,x_2) \end{bmatrix}$} \\ X_2 & \\ & X_1 \quad\quad X_2 \end{array}$$

Note :-

1. covariance with itself is variance.

$\therefore$

$$= \begin{bmatrix} Var\,x_1 & cov(x_2,x_1) \\ cov(x_1,x_2) & var\,x_2 \end{bmatrix}$$

2. $cov(x_1, x_2) = cov(x_2, x_1)$

$$= \begin{bmatrix} var\,x_1 & cov(x_1\,x_2) \\ cov(x_1\,x_2) & var\,x_2 \end{bmatrix}$$

- Now our matrix is square and symmetrical
- diagnol elements are variance
- Other elements are covariance.

Benefit of covariance matrix

1) tells about variance
2) " " covariance which tells the relationship b/w 2 axis is -ve or +ve

Ex 2 Find covariance matrix of 3 dimensions i.e. x, y, z.

Ans

$$
\begin{array}{c|ccc|}
x & var(x) & cov(x,y) & cov(x,z) \\
y & cov(y,x) & var(y) & cov(y,z) \\
z & cov(z,x) & cov(z,y) & var(z) \\
\hline
 & x & y & z
\end{array}
$$