

# Motivation:

↳ We can visualize a maximum of 3 dimensions.

How to visualize dimensions  $> 3$ ?

Solution 01:

↳ We either plot scatter for every combination of columns.

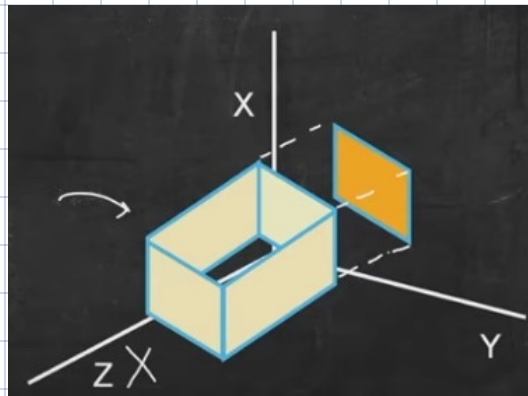
Solution 02:

↳ Reduce dimensions by projecting data to a lower dimension with minimum loss of information. One way is principle component analysis (PCA).

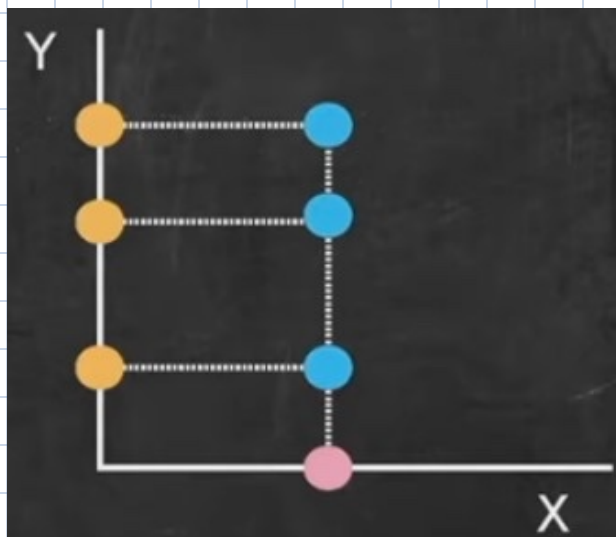
Projection to Reduce Dimension:

↳ Consider you have a cube in 3D space.

↳ We project to either  $X$ ,  $Y$  or  $Z$  & discard the others. Let's discard  $Z$ .



↳ Let's see another example,

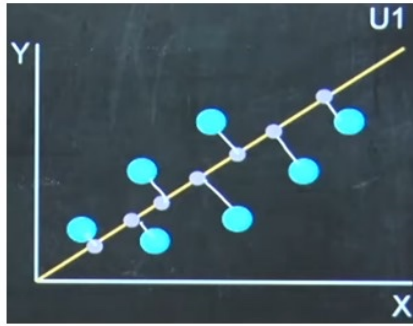
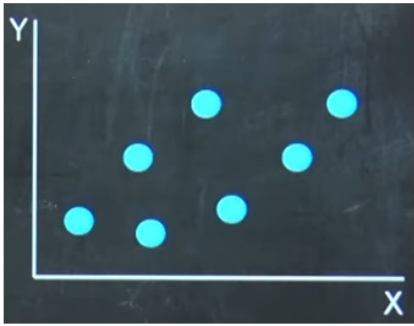


↳ If we discard  $Y$ , most information (Variance) is lost.

↳ If we discard  $X$ , most information (Variance) is maintained.

Project on new axis:

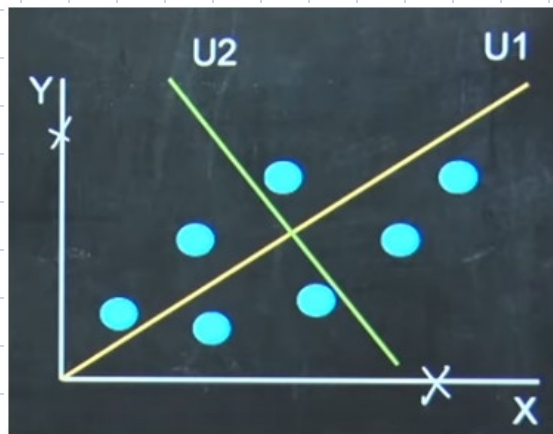
↳ Instead of limiting ourselves to existing, why don't we create a new axis.



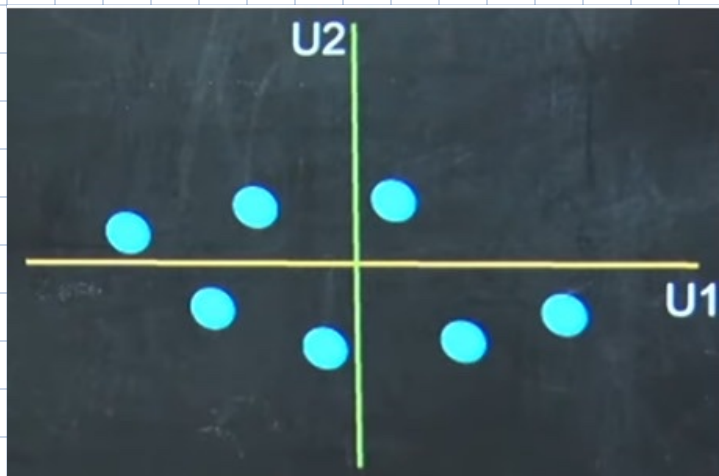
Direction of  $U1$  is such that maximum information i.e. Variance is maintained

↳ Now, if you discard both  $X$  &  $Y$ . You are still able to differentiate b/w data points.

↳ However, some variation in other direction is NOT being captured. So, we create another axis  $U2$  such that  $U2 \perp U1$ .



↳ Then, we discard the original axis.



In a nutshell,

1. We project our higher dimensional data to a new coordinate system.
2. We then choose only those that explain/capture the most information/variance.
3. These axes are orthogonal/perpendicular to each other.

The target is to:

↪ Project data  $X$  on a vector  $u$  such that:

↪ The variance of the projected data is maximum

↪  $u$  is a unit vector.

Example,

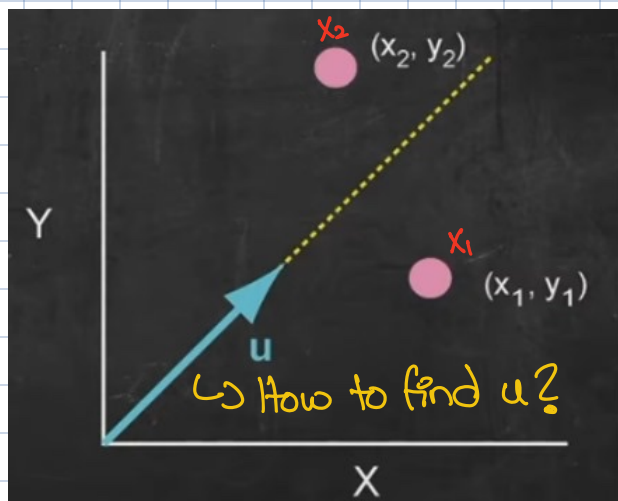
↪ The data:

$$X = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$

$u$  is a unit vector:

$$\|u\| = 1$$

↪ Project  $X$  on  $u$ .

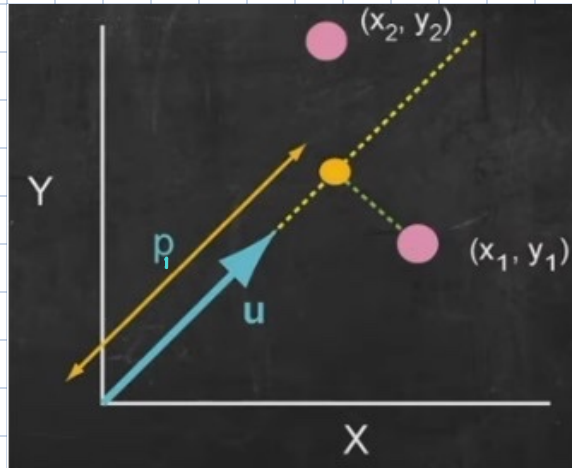
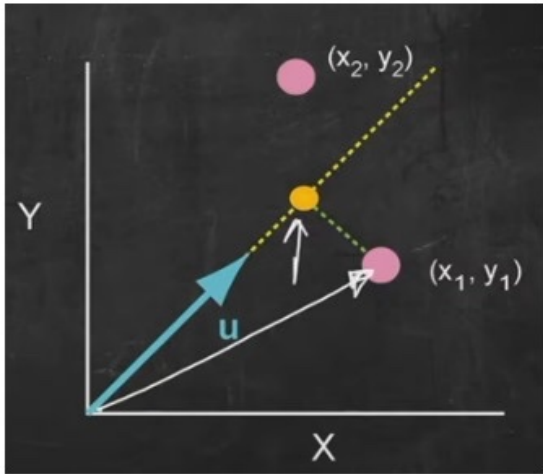


$$X_1 = \begin{bmatrix} x_1 & y_1 \end{bmatrix}, \quad X_2 = \begin{bmatrix} x_2 & y_2 \end{bmatrix}$$

$$P = \frac{X_1 \cdot u}{\|u\|}$$

$$= X_1 \cdot u$$

↳ This will give position & length from origin of the projected point



lets assume,

$$u = \begin{bmatrix} p \\ q \end{bmatrix}$$

$$P_1 = X_1 \cdot u$$

$$= [x_1 \quad y_1] \begin{bmatrix} p \\ q \end{bmatrix}$$

$$= px_1 + qy_1$$

Similarly, we can do:

$$= X_{n \times d} \cdot u_{d \times 1}$$

$$= P_{n \times 1}$$

Now, we have understood projection. Lets move to maximizing information / Variance aspect.

For this,  
PCA makes use of Covariance matrix,

$$X = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$

mean of column  $x$  is  $\bar{x}$

mean of column  $y$  is  $\bar{y}$

$$X_c = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ x_2 - \bar{x} & y_2 - \bar{y} \end{bmatrix} \leftarrow \text{centered data}$$

$$S = \text{Cov}(X) = \frac{1}{N} X_c^T X_c$$

properties of covariance matrix:

1.  $S$  is square matrix ( $d \times d$ )
2.  $S$  is symmetric  $S = S^T$
3. All eigenvalues of  $S$  are orthogonal
4. All eigenvalues of  $S$  are non-negative
5. If  $n > d$ , then  
 $\hookrightarrow$  All eigenvalues  $\lambda > 0$
6. If  $n < d$ , then  
 $\hookrightarrow$  At least one eigenvalue  $= 0$
7.  $S$  will have  $d$  eigenvalues ( $\lambda$ )

Why we use eigenvector? Why does eigenvector represent the direction of maximum variance?

### Direction of greatest variability

- Select dimension  $\mathbf{e}$  which maximizes the variance
- Points  $\mathbf{x}_i$  "projected" onto vector  $\mathbf{e}$ :
- Variance of projections:  $\frac{1}{n} \sum_{i=1}^n (\sum_{j=1}^d x_{ij} e_j)^2 = \frac{1}{n} \sum_{j=1}^d \left( \sum_{i=1}^n x_{ij} e_j \right)^2$
- Maximize variance  
 - want unit length:  $\|\mathbf{e}\| = 1$   
 - add Lagrange multiplier

$$V = \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=1}^d x_{ij} e_j \right)^2 - \lambda \left( \left( \sum_{j=1}^d e_j^2 \right) - 1 \right)$$

$$\frac{\partial V}{\partial e_a} = \frac{2}{n} \sum_{i=1}^n \left( \sum_{j=1}^d x_{ij} e_j \right) x_{ia} - 2\lambda e_a = 0$$

$$\Sigma \mathbf{e} = \lambda \mathbf{e} \quad \left\{ \begin{array}{l} \sum_{j=1}^d \text{cov}(1,j) e_j = \lambda e_1 \\ \vdots \\ \sum_{j=1}^d \text{cov}(d,j) e_j = \lambda e_d \end{array} \right. \quad \leftarrow \text{hold for } a=1..d \quad 2 \sum_{j=1}^d e_j \left( \frac{1}{n} \sum_{i=1}^n x_{ia} x_{ij} \right) = 2\lambda e_a$$

PCA 11: Eigenvector = direction of maximum variance

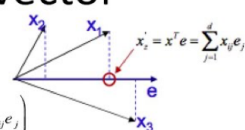
$\hookrightarrow$  When you look for the direction of greatest variance, you will find a

eigen vector.

How to get Variance?

### Variance along eigenvector

Variance of projected points ( $\mathbf{x}^T \mathbf{e}$ ):



$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=1}^d x_{ij} e_j - \mu \right)^2 &= \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=1}^d x_{ij} e_j \right)^2 && \mu = \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=1}^d x_{ij} e_j \right) \\ &= \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=1}^d x_{ij} e_j \right) \left( \sum_{a=1}^d x_{ia} e_a \right) && = \sum_{j=1}^d \left( \frac{1}{n} \sum_{i=1}^n x_{ij} \right) e_j \\ &= \sum_{a=1}^d \sum_{j=1}^d \left( \frac{1}{n} \sum_{i=1}^n x_{ia} x_{ij} \right) e_j e_a \\ &= \sum_{a=1}^d \left( \sum_{j=1}^d \text{cov}(a, j) e_j \right) e_a && \text{cov}(a, j) = \frac{1}{n} \sum_{i=1}^n x_{ia} x_{ij} \\ &= \sum_{a=1}^d (\lambda e_a) e_a && \sum_{j=1}^d \text{cov}(a, j) e_j = \lambda e_a \quad \mathbf{e} \text{ is an eigenvector of the covariance matrix} \end{aligned}$$

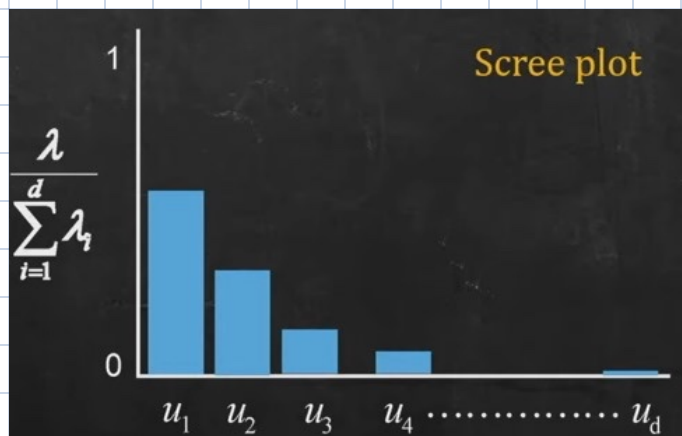
PCA 12: Eigenvalue = variance along eigenvector

↳ After finding the eigen vectors & values. We arrange them in descending order.

$$\begin{array}{cccc} \lambda_1 & \lambda_2 & \dots & \lambda_d \\ \downarrow & \downarrow & & \downarrow \\ u_1 & u_2 & \dots & u_d \end{array}$$

Lets create a probability distribution

$$\frac{\lambda_1}{\sum_{i=1}^d \lambda_i} > \frac{\lambda_2}{\sum_{i=1}^d \lambda_i} > \dots > \frac{\lambda_d}{\sum_{i=1}^d \lambda_i}$$



Select the ones with more variance/probability.



$$\begin{array}{ccc}
 \mathbf{X} & \mathbf{V} & \mathbf{T} \\
 \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{id} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{nd} \end{bmatrix} & \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & \cdots & | \end{bmatrix} & = \begin{bmatrix} x'_{11} & x'_{i1} & \cdots & x'_{1d} \\ \vdots & \vdots & \ddots & \vdots \\ x'_{i1} & x'_{i2} & \cdots & x'_{id} \\ \vdots & \vdots & \ddots & \vdots \\ x'_{n1} & x'_{n2} & \cdots & x'_{nm} \end{bmatrix} \\
 (n \times d) & (d \times m) & (n \times m) \\
 \hline
 \text{Centered data} & m \text{ selected principal components} & \text{Projected data}
 \end{array}$$

PCA algorithm:

1. Given data  $X$
2. Compute mean for every column.
3. Centre  $X$  by subtracting mean
4. Compute Covariance matrix
5. Compute eigen vectors & values for covariance matrix
6. Arrange them in descending order.
7. Select  $m$  of them
8. Find reconstructed Cov matrix  $U \lambda U^T$  called  $P$
9. Project the data  $X_{\text{proj}} = X_c \cdot P$