Motivation:

↳ We can visualize a maximum of 3 dimensions.

How to visualize dimensions >3?

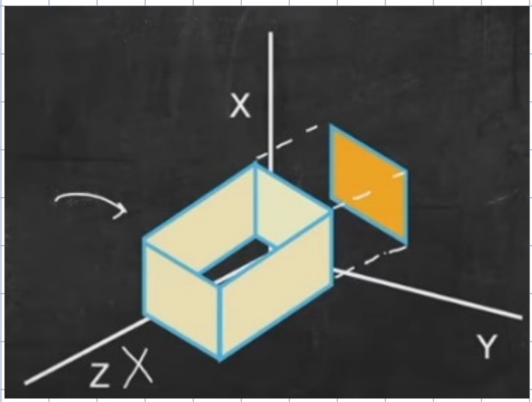
Solution 01:
↳ We either plot scatter for every combination of columns.

Solution 02:
↳ Reduce dimensions by projecting data to a lower dimension with minimum loss of information. One way is principle component analysis (PCA).

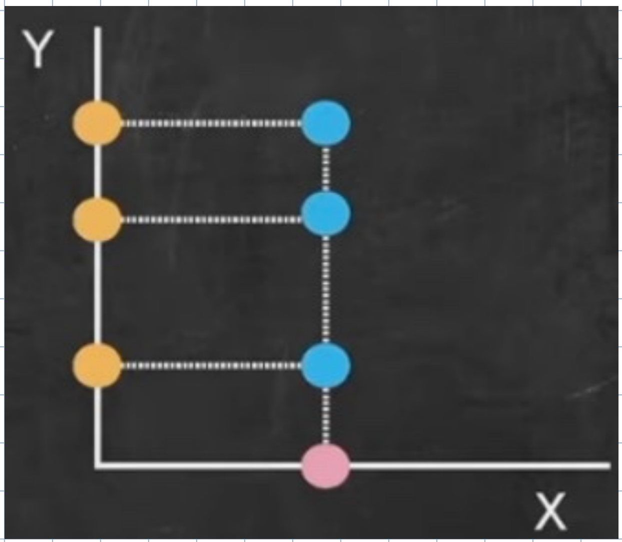
Projection to Reduce Dimension:

↳ Consider you have a cube in 3D space.

↳ We project to either X, Y or Z & discard the others. Lets discard Z.



↳ Lets see an another example,

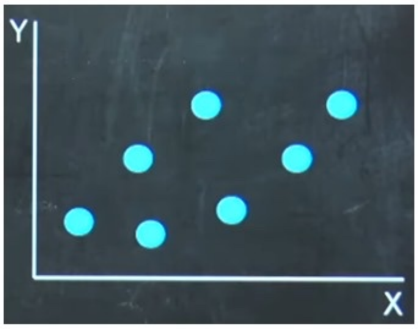


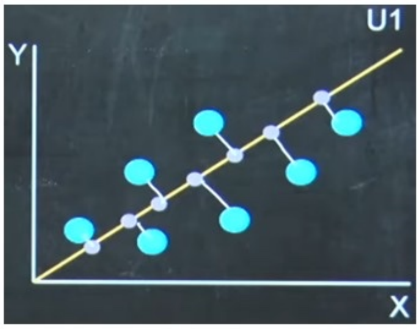
↳ If we discard Y, most information (Variance) is lost.

↳ 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.

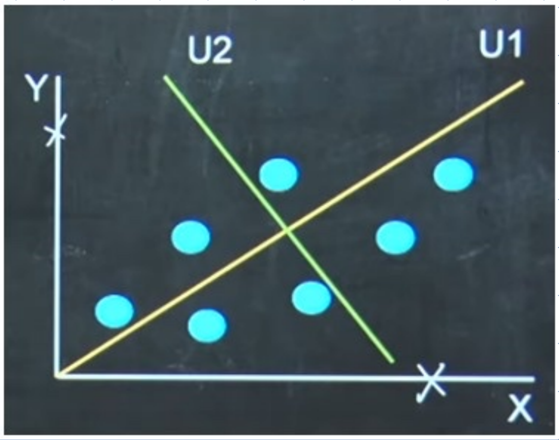




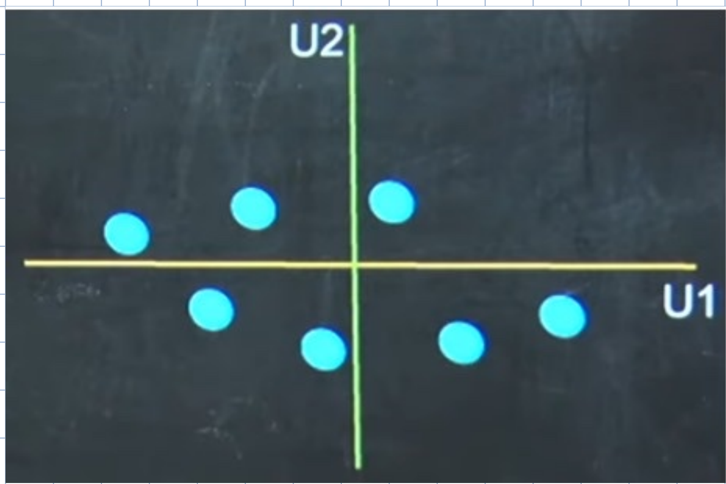
Direction of U1 is such that maximum information i.e Variance is maintained

↳ 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 an 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 the 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,

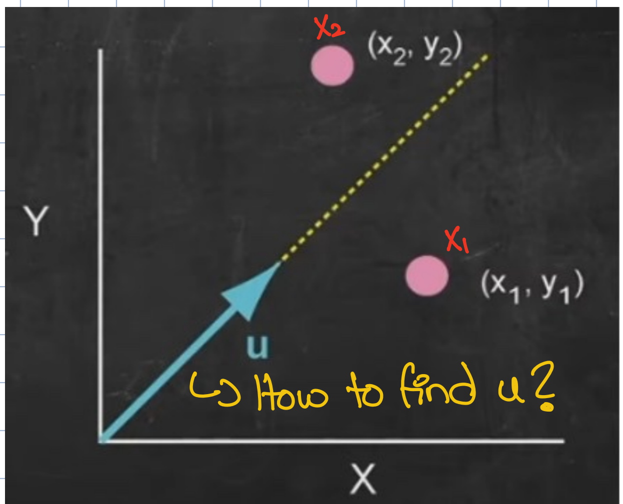
↳ The data:

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

u is a unit vector:

$$\|u\| = 1$$

↳ Project X on u.

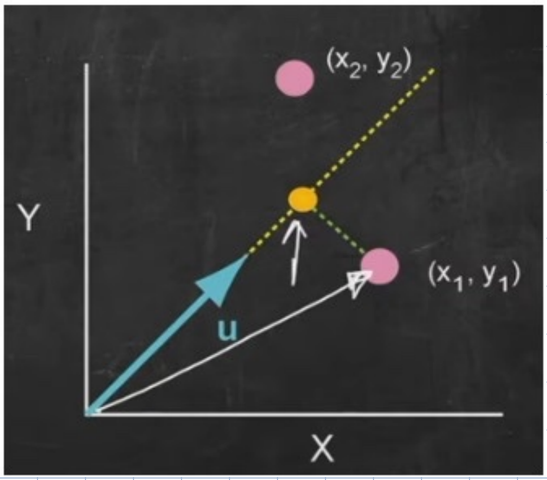


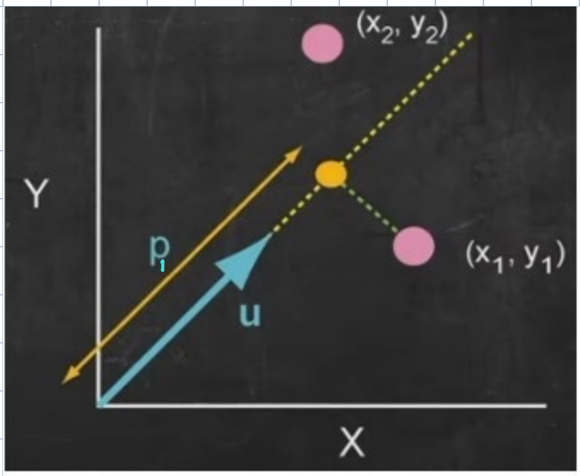
$X_1 = [x_1 \; y_1]$, $X_2 = [x_2 \; y_2]$

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

$$= X_1 \cdot u$$

↳ 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 $\bar{y}$

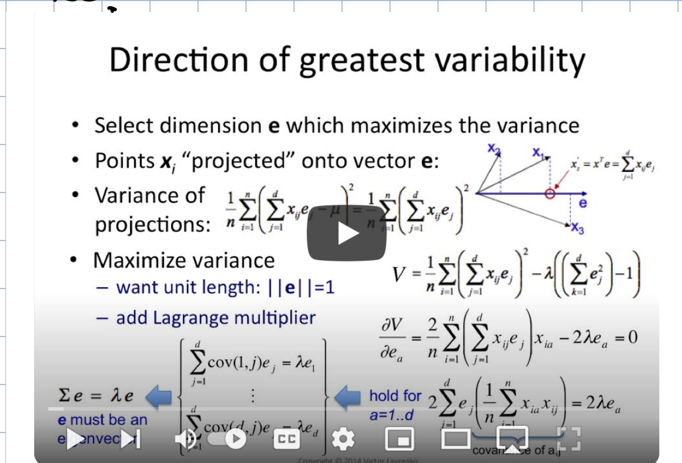
$$X_c = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ x_2 - \bar{x} & y_2 - \bar{y} \end{bmatrix}$$ ← centered data

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

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
   ↳ All eigenvalues $\lambda > 0$
6. If $n < d$, then
   ↳ Atleast one eigenvalue $= 0$
7. S will have d eigenvalues (λ)

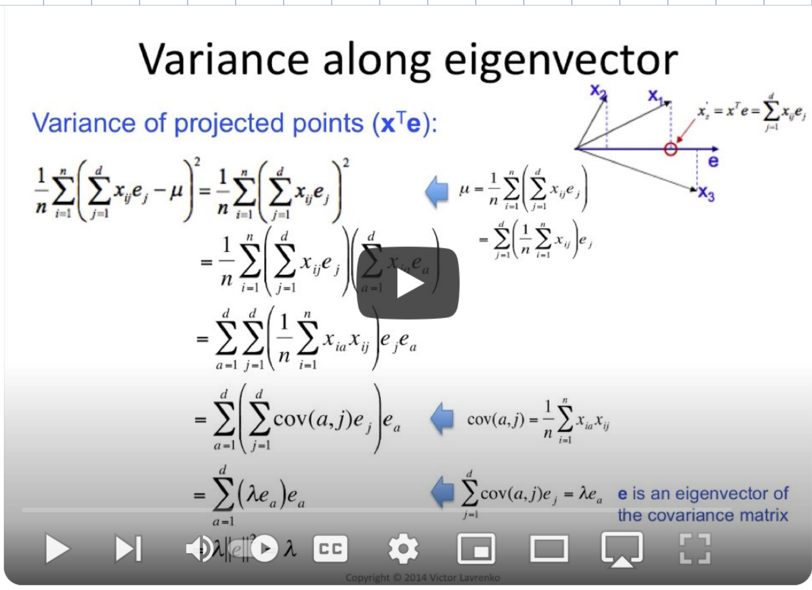
Why we use eigenvector? Why does eigenvector represent the direction of maximum variance?



PCA 11: Eigenvector = direction of maximum variance

↳ When you look for the direction of greatest variance, you will find a eigen vector.

How to get Variance?



PCA 12: Eigenvalue = variance along eigenvector

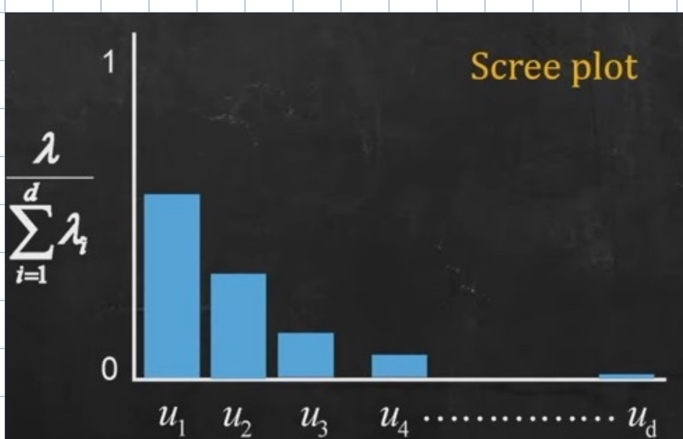
↳ After finding the eigen vectors & values. We arrange them in descending order.

$$\lambda_1 > \lambda_2 > \dots > \lambda_d$$

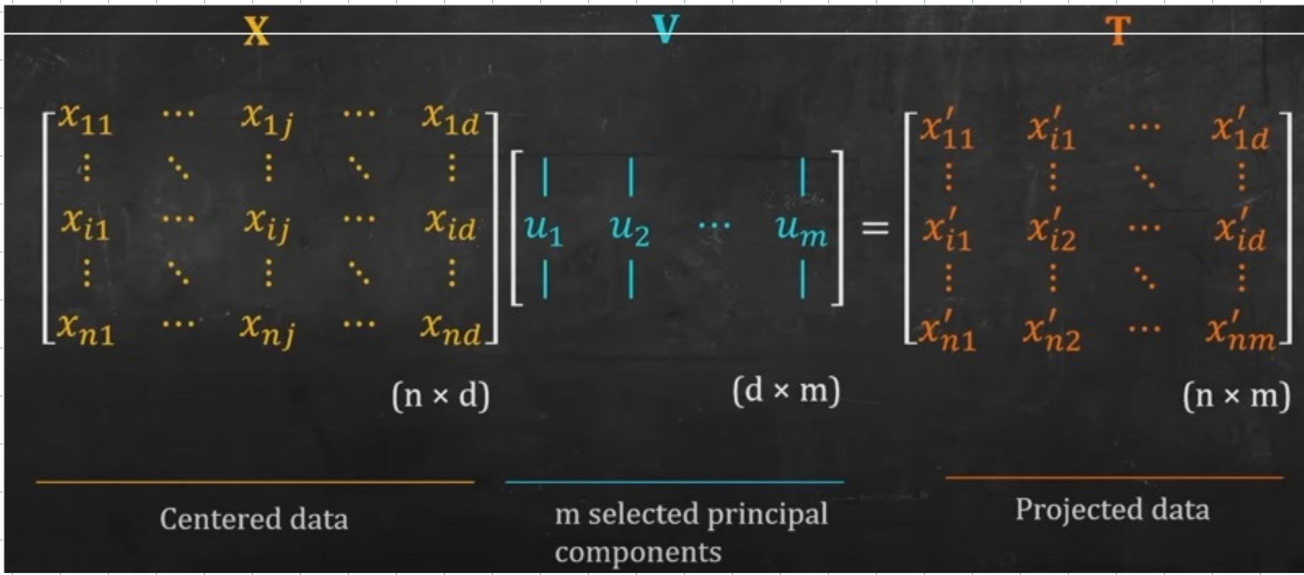
$$\begin{array}{cccc} \downarrow & \downarrow & & \downarrow \\ u_1 & u_2 & \dots & u_d \end{array}$$

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.



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 $X_{proj} = X_c \cdot P$