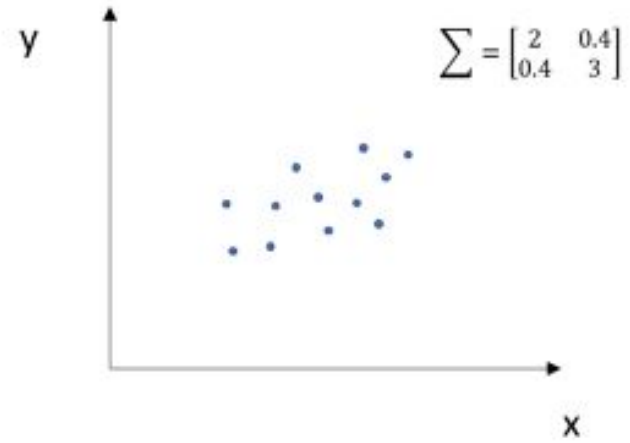
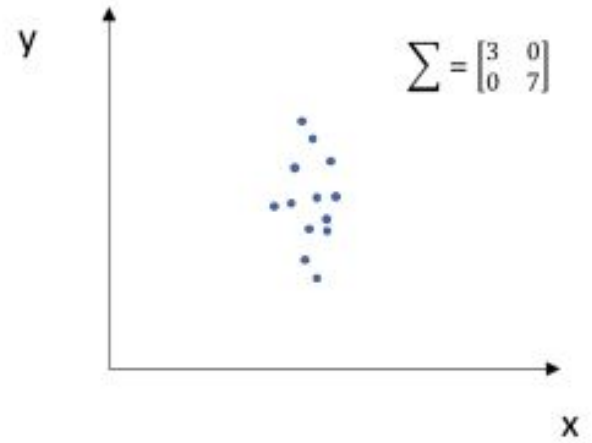
PRINCIPAL COMPONENT ANALYSIS

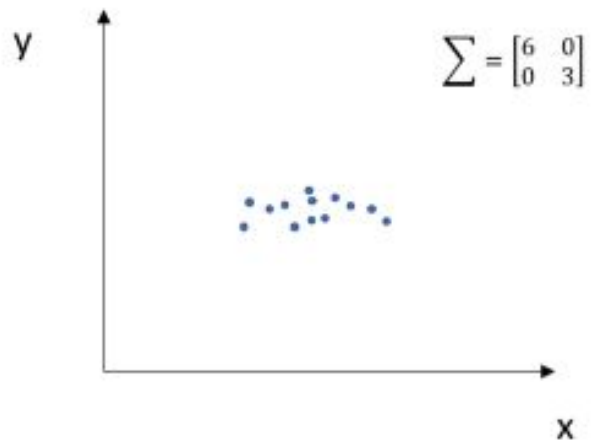
Motivation

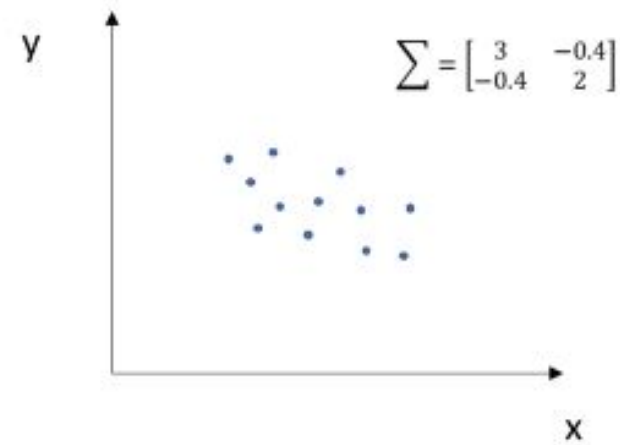
- Large multidimensional data sets X = $\{\vec{x}_1, \ldots \vec{x}_n\} \equiv [\, A_{11}, A_{12}\ldots\ldots, A_{nm}]$
- PCA goal is to reduce the number of dimensions/attributes A_1 .. A_m, so that now there are ***l<m*** dimensions.
- We thus need to eliminate some of them, such that most of the original information is still intact and there is least mean square error
- This reduced set of attributes makes it easier to visualize and analyze data in machine learning
- What is the transformation?

Covariance Illustration

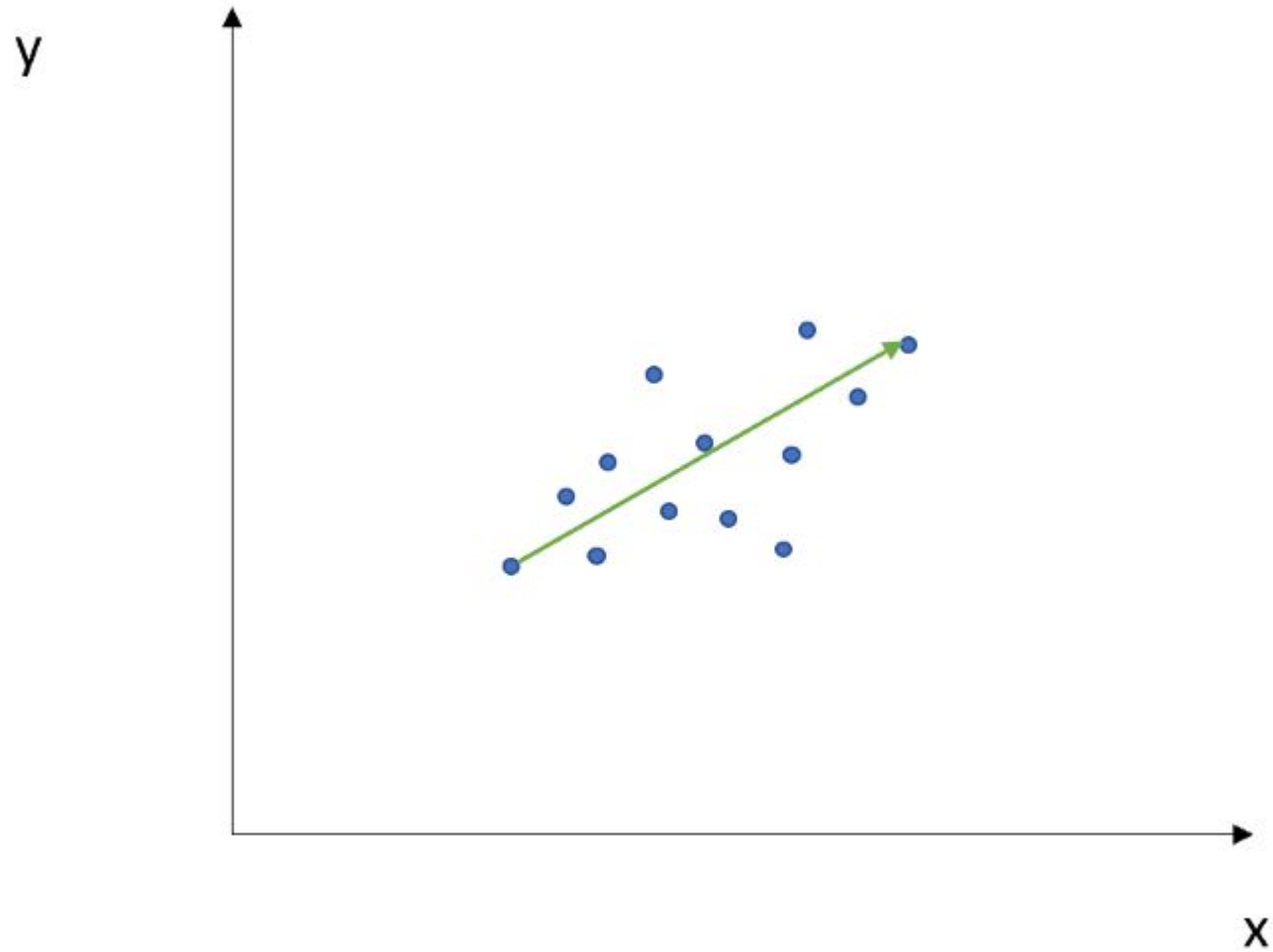




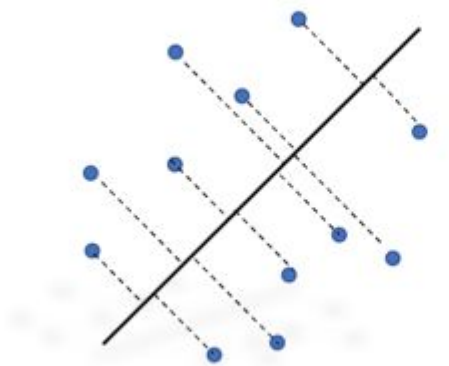




Eigen Vector and Value



PCA



$$\Sigma V = LV$$

$$\begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(y,x) & Var(y) \end{bmatrix} \begin{bmatrix} \boldsymbol{v_1} \\ \boldsymbol{v_2} \end{bmatrix}' = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{v_1} \\ \boldsymbol{v_2} \end{bmatrix}$$

Standardization

- PCA examines the variances in each attribute.
- To ensure all attributes are examined fairly, they must be standardized. A variable with variance between 1 and 1000 will otherwise dominate over another ranging between 0 and 1!

- Each value of each variable is standardized as:
  - $new\ value = \frac{old\ value - mean}{standard\ deviation}$
- Now all attributes are on same scale

Covariance matrix

- If two attributes are correlated, it indicates that there is lot of redundant information.
- For example, cost of house in rupees and in dollars. Number of rooms and total area of house. Number of votes and number of people who visited voting booth.
- Compute *m* X *m* square symmetric covariance matrix between the attributes considering all data points. **C[m][m]:**

$Var(x_1)$	$Covar(\overrightarrow{x_1,x_2})$	$Cov(\overrightarrow{x_1,x_3})$
	$Var(x_2)$	$Cov(x_2,\overrightarrow{x_3})$
		$Var(x_3)$

- If covariance is +ve, attributes synchronize positively. Else they vary in opposite directions.

Covariance Matrix

- Note: Covariance matrix represents the full set of self/joint variances. In that sense, variances are pairwise coalesced.
- Covariance Matrix Covariance Matrix can serve as Transformation matrix to any set of vectors
- In general, any vector can be transformed when a transformation matrix is applied to it.

$$
\begin{aligned}
T\boldsymbol{v}_1 &= \boldsymbol{b}_1 \\
T\boldsymbol{v}_2 &= \boldsymbol{b}_2 \\
&\ldots\ldots \\
T\boldsymbol{v}_n &= \boldsymbol{b}_n
\end{aligned}
$$

Eigen values and Eigen Vectors

- However, Covariance Vector (like any other square matrix) cannot transform the direction of special vectors, called eigen vectors. It can only stretch it along the same direction. The amount to which they are stretched create Eigen values.

$$Tv_i = \lambda v_i$$

- Solving this gives as many Eigen values λ as the order of the covariance matrix
- Each Eigen Value has its corresponding Eigen Vector (which never changes direction), also called characteristic roots. The Task is to find the Eigen Values and Eigen Vectors

Eigen vectors and values

- Eigen vectors are directions of the axes which contain the covariances.
- Eigen Values are coefficients of Eigen vectors and denote the amount of variance of the direction. Eigen values and vectors always exist in pairs.
- Significant Principal Components have higher coefficients (Eigen Values). Normalize all Eigen Values and discard as per significance.
- Remaining Eigen vectors form Feature Vector

Recast data along the new dimensions

- Let Original dataset be X[n][m].
- Let Feature vector be F[m][l]
- Final Dataset: X’[n,l] = X[n][m] * F[m][l]

Principal components of covariance matrix

- Principal components are new variables constructed from old variances.
- Each Principal components is a unique linear combinations of all old variances. There are *m* principal components for *m* attributes, each a unique combination of them.
- Each component represents some major direction of covariances.
- Each component is independent and uncorrelated with others. They are all orthogonal.

Principal Components

- First Principal component has the maximum information – it explains the maximum percentage of variance in the old data.
- Then next maximum information is in second component, …. And so on.
- Principal components are found out by Eigen Vectors and Eigen Values of covariance matrix.