# PCA

- Motivation

  - Usually, datasets have lot more dimensions than required.
  - Ex: Movement of spring in 1-d captured by three cameras at different angles.
  - Therefore reduce dimensionality of data to get rid of redundant information.

    Ex: Remove two extra cameras. Only one camera needed to capture 1-d motion.

- Working

  - Suppose the spring oscillating in x-direction If the cameras were recording three separate dimensions (x, y and z), we could easily remove 2 of them and keep the one that is capturing the 1-d movement

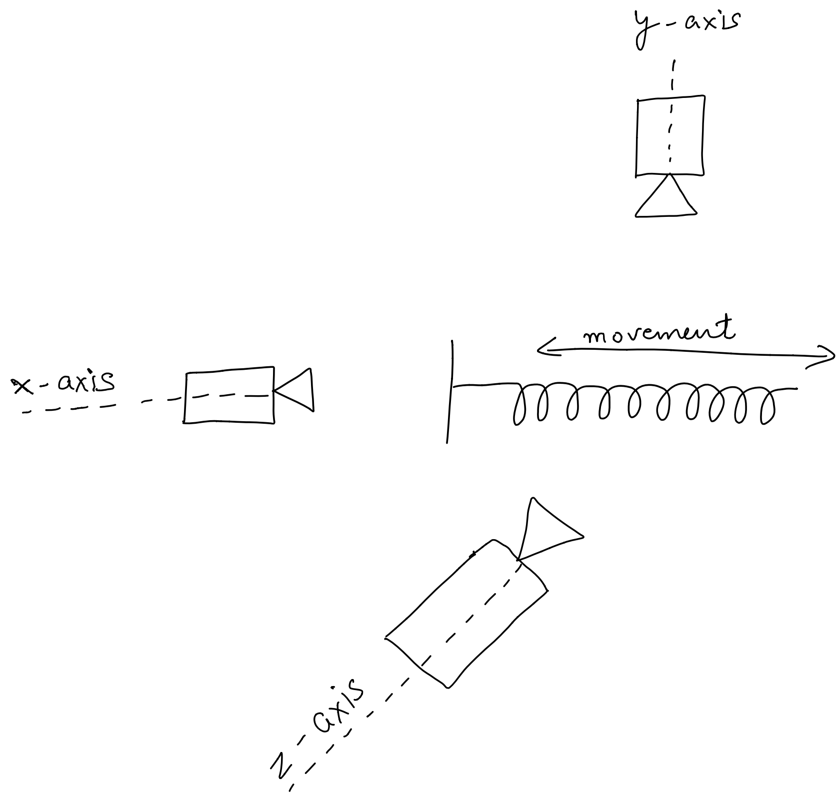


Figure:- Camera x should be deleted. One of camera y or z should be deleted to capture the movement of the spring uniquely.

- From the figure, it is clear that two dimensions can be removed to capture non-redundant information. But, only if the dimensions are orthonormal to each other. In real world datasets, it is unlikely that we get data where attributes are orthonormal.

- <u>Change of basis</u>:

  - We need to find orthonormal basis such that an entire dimension (column) could be removed from the dataset without loss of information.

$$\underbrace{P}_{\text{Principal component matrix}} \overset{\text{original data}}{X} = \underbrace{Y}_{\text{new data with orthonormal columns.}}$$

  - Need to find $P$ that transforms $X$ to $Y$ where $Y$ has orthonormal columns.

- <u>Maximising Variance</u>

  - Another nice thing would be to capture the maximum information (= maximum variance) in the first dimension of $Y$, second highest information in second dimension, and so on.

- In other words, we want $Y$ where the covariance matrix is diagonal (= covariance between any two dimensions is 0)

$$PX = Y$$

such that $C_Y$ is diagonal.

$$C_Y = \frac{1}{n} YY^T$$

$$= \frac{1}{n} (PX)(PX)^T$$

$$= \frac{1}{n} P X X^T P^T$$

$$= P\left(\frac{1}{n} X X^T\right) P^T$$

$$C_Y = P C_X P^T$$

Diagonalizing $C_X$:

$$C_Y = P(EDE^T)P^T$$

$$= (PE) D E^T P^T$$

$$C_Y = (PE) D (PE)^T$$

By selecting $P = E^T \Rightarrow PE = E^T E = I$

$$\Rightarrow \boxed{C_Y = D}$$

$P$ is the (transpose of) eigenvector matrix of $C_X$.

Moreover, if $D$ is arranged such that $d_1 > d_2 > \ldots$ then $var_1 > var_2 > \ldots$ where $var_i$ is variance of $i^{th}$ dimension of $Y$.

---

## Summary

---

- Need $P$ in $PX = Y$.
  - $P$ is principal component matrix.
  - $Y$ is new reduced data.
- Need $var_1 > var_2 > \ldots$ in columns of $Y$.
  - $\Rightarrow$ covariance between any pair of features of $Y = 0$. $\left( cov_{i,j} = 0,\ i \neq j \right)$
  - $\Rightarrow$ Covariance matrix of $Y$ $\left(= \frac{1}{m} YY^T\right)$ is diagonal matrix.
  - $\Rightarrow$ $P$ is eigenvector matrix of covariance matrix of $X$.

<u>Note</u>:- Formula used for covariance matrix $\left(C_x = \frac{1}{n} X X^T\right)$ works only if the data is mean-centered.

## Matrix sizes

$f$ = number of features

$s$ = number of samples

- $$\underset{(f,f)}{P} \; \underset{(f,s)}{X} = \underset{(f,s)}{Y}$$

- $$\underset{(f,f)}{C_x} = \frac{1}{s} \underset{(f,s)}{X} \; \underset{(s,f)}{X^T}$$

- $$\underset{(f,f)}{P} = \underset{(f,f)}{E^{-1}} = \underset{(f,f)}{E^T}$$