Curse of Dimensionality

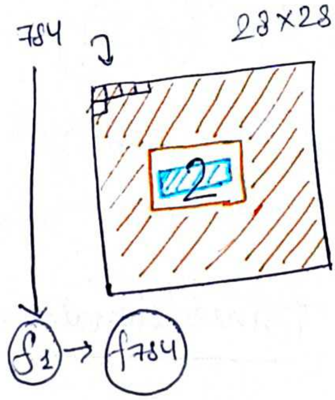
features

↓

dimensionals

f(x) < f(x)

4   5



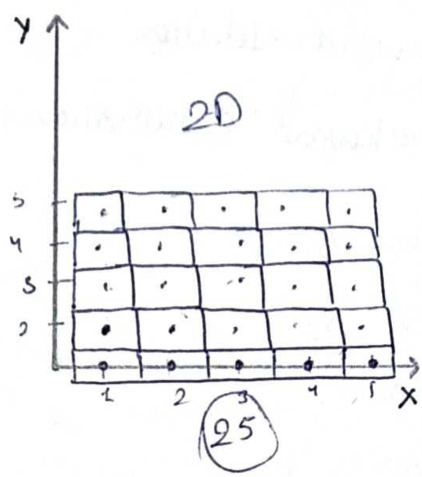
Data

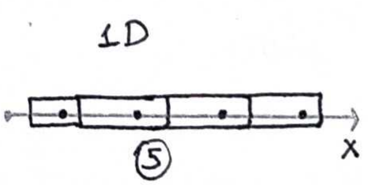
↓

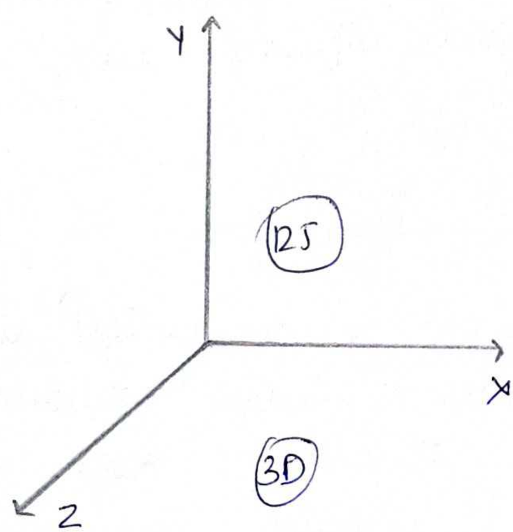
1) Images

2) Text

higher dimension → Sparsity → Increasing Distance







1) Performance Decrease

2) Computation

Dimensionality Reduction

Feature Selection

1. Forward selection
2. Backward elimination

Feature extraction

1. PCA
2. LDA
3. tsne

Feature Extraction

(PCA)

Principal Component Analysis :- 1.) PCA can be abbreviated as principal component Analysis

2) PCA comes under the unsupervised Machine learning category.

3) Reducing the number of variables in a data collection while retaining as much information as feasible is the main goal of PCA. PCA can be mainly used for Dimensionality Reduction and also for important feature selection.

4.) Corelated features to Independent feature.

Technically, PCA provides a complete explanation of the composition of variance and covariance using multiple linear combination of the core variables. Row scattering may be analyzed using PCA, which also identifies the distribution related properties.

Why do we need PCA in Machine Learning?

When a computer is trained on a big, well-organized dataset, machine learning often excels. One of the techniques used to handle the curse of dimensionality in machine learning is principal component analysis (PCA). Typically having a sufficient amount of data enables us to create a more accurate prediction model since we have more data to use to train the computer. But working with a huge data collection has its own drawbacks. The curse of dimensionality is the ultimate trap.

The title of an unreleased Harry Potter novel does not refer to what happens when your data has too many characteristics and perhaps not enough data points; rather, it refers to the curse of dimensionality. One can use dimensionality reduction to escape the dimensionality curse. Having 50 variables may be cut down to 40, 20, or even 10. The strongest effects of dimensionality reduction are found here.

Overfitting issue will arise while working with high-dimensional data, and dimensional data, and dimensionality reduction will be used to address them. Increasing interpretability and minimizing information loss. aids in locating important characteristics. Aids in the discovery of a linear combination of varied sequences.

When to use PCA?

1) Whenever we need to know our features are independent of each other.

2) Whenever we need fewer features from higher features

Geometric Intuition

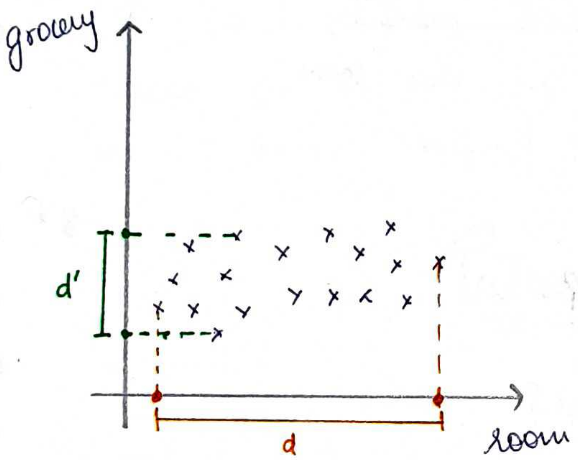
No. of rooms	No. of grocery shop	Price
3	2	60
4	0	130
5	6	170
2	10	90

PCA

[Feature extraction]

↓

we have to know feature selection

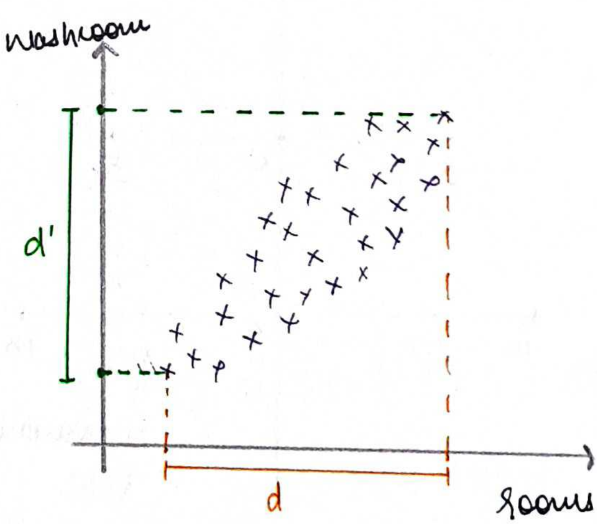


$d > d'$

* which feature have high variance are choosed for model.

$\underbrace{x \rightarrow \text{room}}_{\text{variance} \uparrow}$

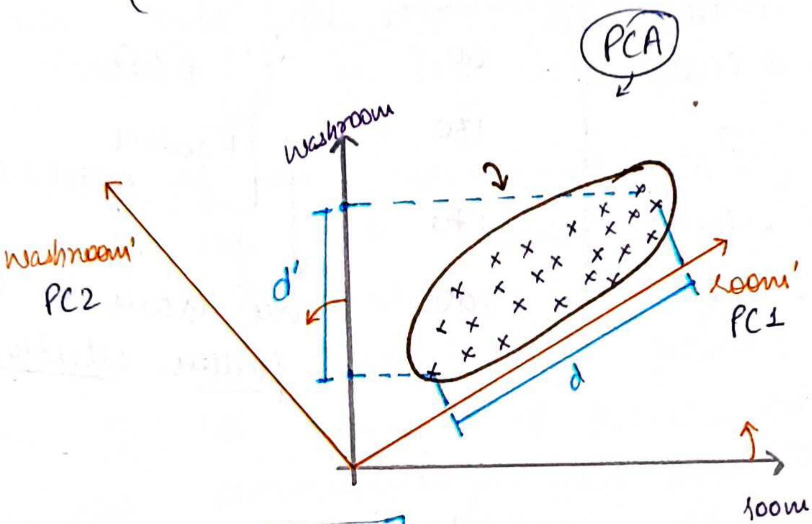
No. of rooms	No. of washroom	Price



$\boxed{d \simeq d'}$

convert → Size | Price (1D) } PCA does

Room | Washroom | Price
(2D)

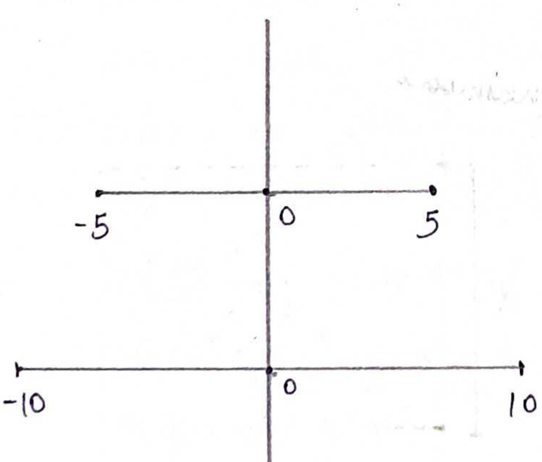


$\boxed{d > d'}$

#1

Why variance is important?

Lets discuss what is variance?



$$\mu = \frac{-5+0+5}{3} = 0$$

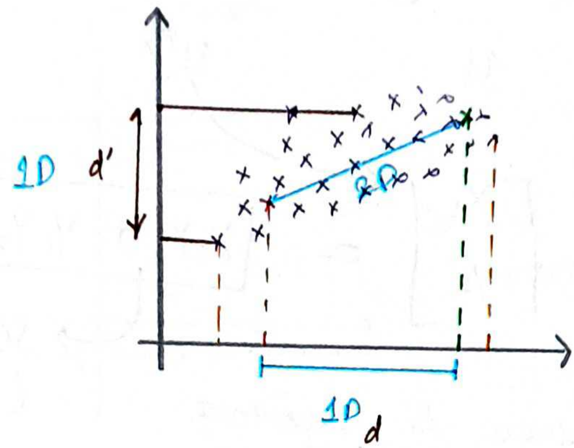
$$\mu = \frac{-10+0+10}{3} = 0$$

when we find mean always 0. and mean find central tedency not find diff^n bet^n data.

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

$$V = \frac{25+0+25}{3} = \frac{50}{3}$$

$$V = \frac{100+0+100}{3} = \frac{200}{3}$$

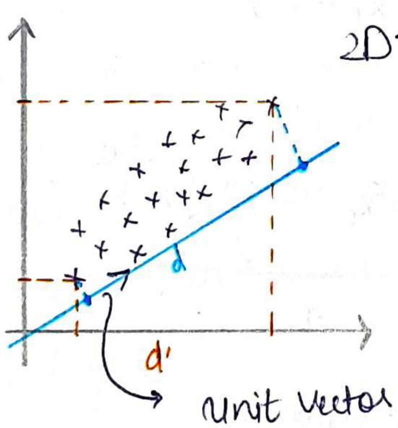


$d > d'$

from 1D(d) we can easily see all points but from d' side we cannot so variance is important.

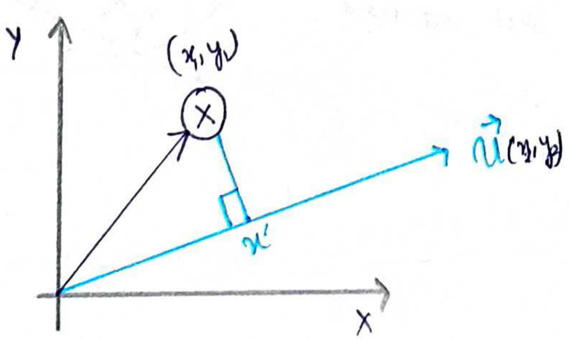
when convert from higher dim to lower dim when convert on x-axis spread or variance ↑↑. but on Y-axis variance ↓↓.

Problem Formulation



2D → 1D

$\boxed{d \simeq d'}$



$$\frac{\vec{u}\cdot\vec{x}}{|u|} = \vec{u}\cdot\vec{x} = u^T x$$

$\therefore |u| = 1$

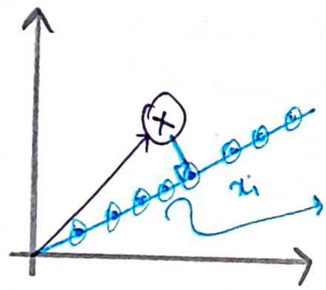
$[x_1\ y_1]$ $[x_2\ y_2]$ $\rightarrow u^T x$

x u

$$[x_1\ y_1] \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \boxed{x_1 y_2 + y_1 y_2}$$

→ Scaler (Number)

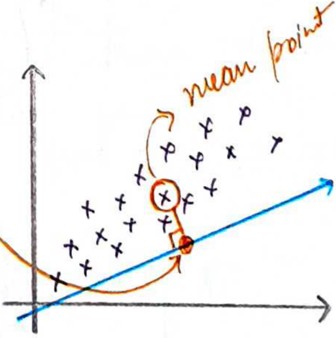
We know that variance is important and we have to find the higher variance vector.



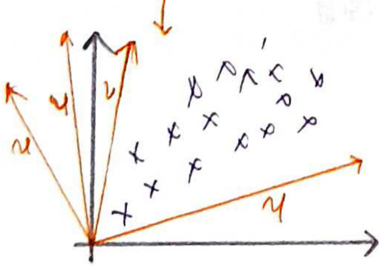
we know that $[u^T x_1]$

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

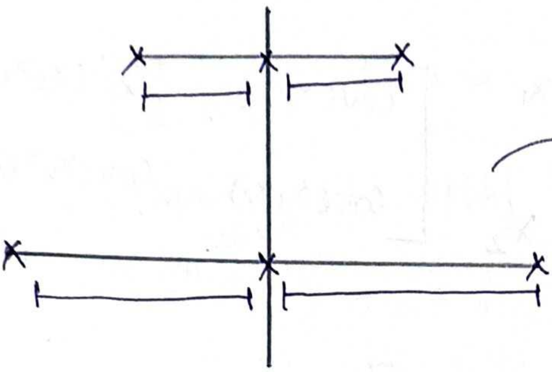
$$\text{variance}(u) = \frac{\sum_{i=1}^{n} (u^T x_i - u^T \bar{x})^2}{n}$$



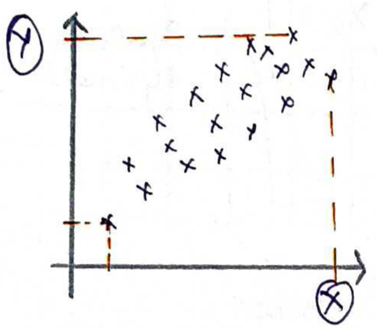
* We have find that unit vector (u) where variance is high. Because there are many unit vector but we have to choose higher one.

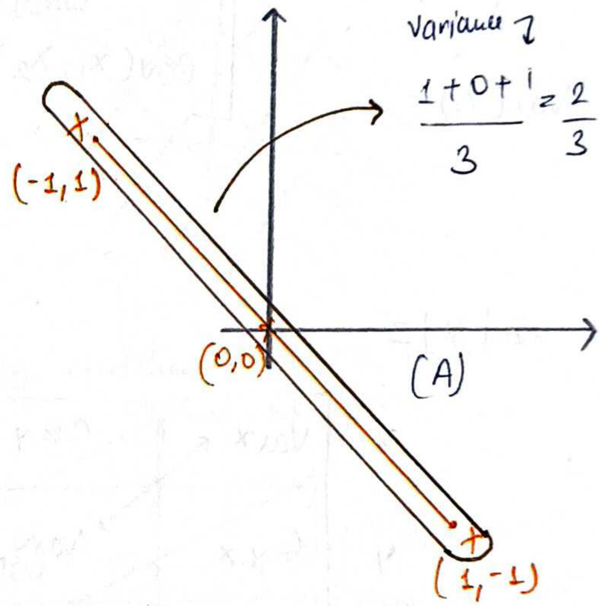


Covariance and Covariance Matrix



Variance





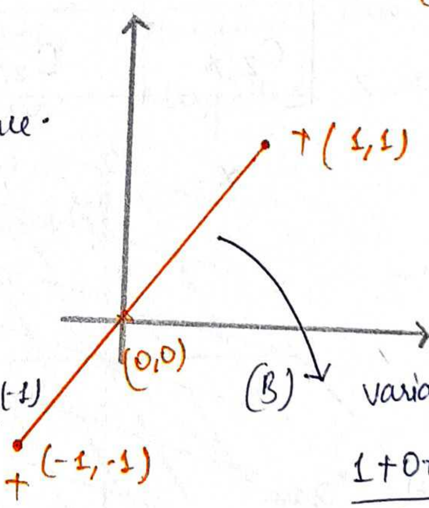
Variance ↓

$\frac{1+0+1}{3} = \frac{2}{3}$

correlation

-1 to 1

* Variance can not find any difference betn the point so, we have to use covariance.
covariance define the relation betn x and y.



variance ↓

$\frac{1+0+1}{3} = \frac{2}{3}$

$$\text{Covariance }(A) = \frac{(-1)\times(1)+0\times0+(1)\times(-1)}{3}$$

$$\text{COV}(A) = \frac{-1+0-1}{3} = \frac{-2}{3}$$

difference

$$\text{Covariance }(B) = \frac{(-1)(-1)+(0)(0)+(1)(1)}{3} = \frac{2}{3}$$

Covariance Metrix

$x_1 \mid x_2$

cov Metrix

$[2 \times 2]$

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

$\underline{cov(x_1, x_1)}$

$\downarrow$

$Var(x_1)$

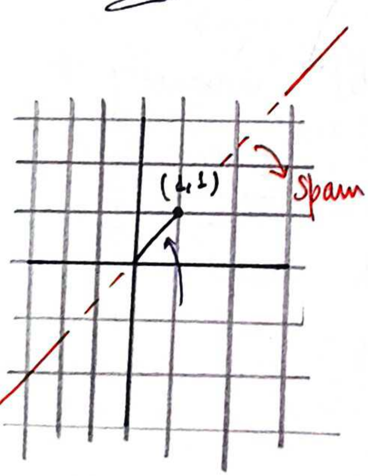
$$\begin{bmatrix} Var\, x_1 & cov(x_2, x_1) \\ cov(x_1, x_2) & var\, x_2 \end{bmatrix}$$

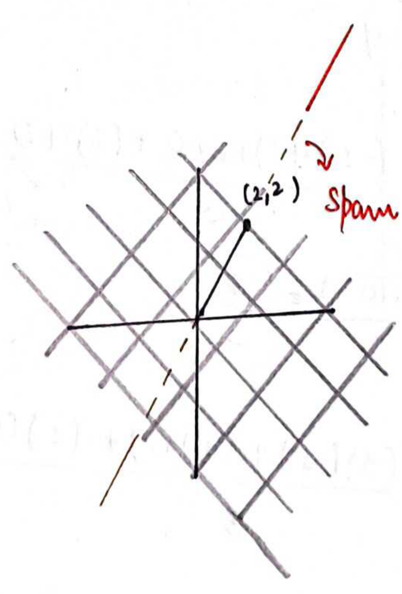
Square symmetrix

$x \mid y \mid z$

$$\begin{array}{c|c} & \begin{array}{ccc} x & y & z \end{array} \\ \hline \begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix} Var\, x & C_{x,y} & C_{x,z} \\ C_{y,x} & Var\, y & C_{y,z} \\ C_{z,x} & C_{z,y} & Var\, z \end{bmatrix} \end{array}$$

Span:-

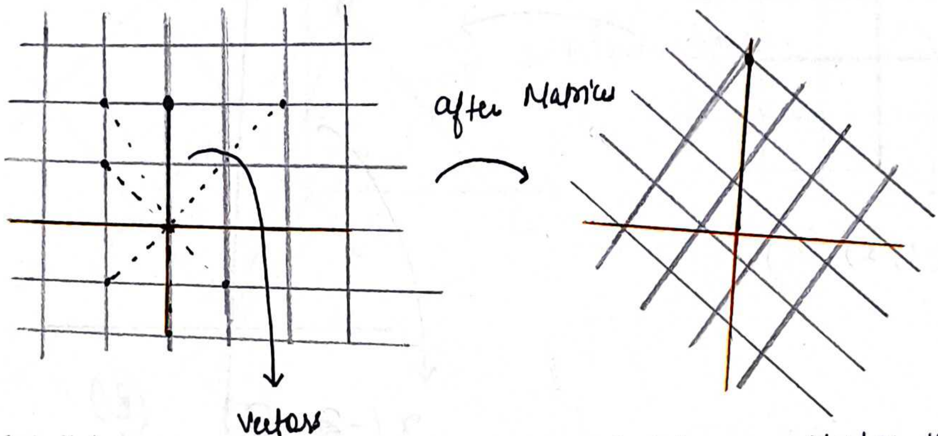




Linear Transformation, Eigen Vectors and Eigen Values

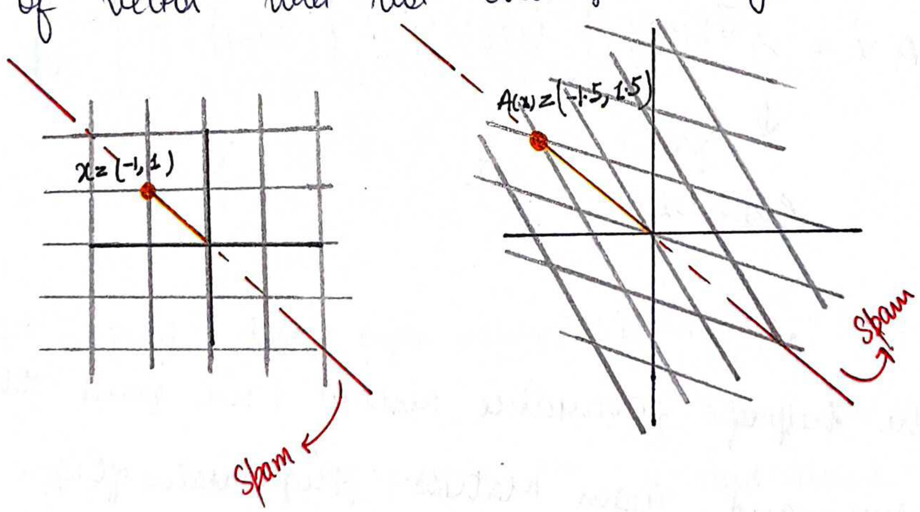
$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$ Matrix

When we apply matrix then ~~direction~~ span of vector same and magnitude of vector will be change.



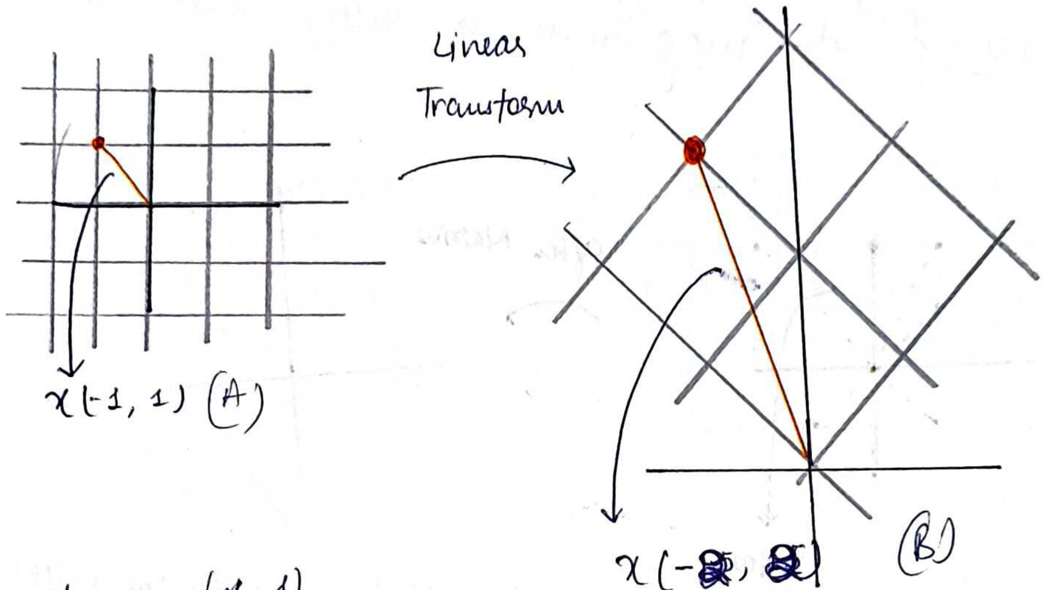
* after Matrix → Vector will be expand or shrink.

Eigen Vectors :- Eigen Vectors are special vectors when we apply transform then direction of vector will not change. Magnitude will change



Eigen Value : Eigen value means difference between the before vector magnitude and after vector magnitude.

Factor which vector magnitude increase and decrease.



Eigen Value = (1, 1) ← difference betn (A) and (B)

Eigen Vector

$$A\vec{v} = \lambda\vec{v}$$

A → Matrix

λ → Eigen Values

In short :-

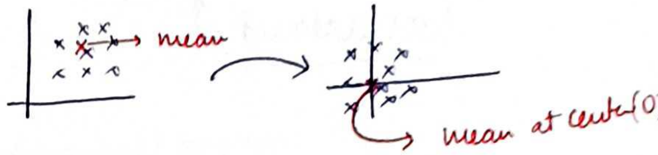
With the help of covariance Matrix, we find the Eigen values and Eigen Matrix. High value of Eigen value indicates high variance.

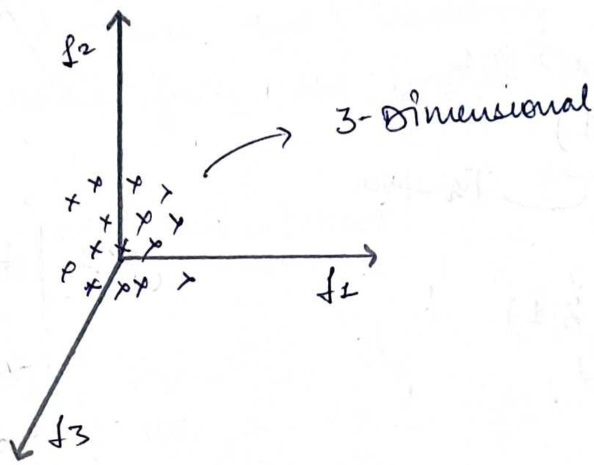
$\boxed{\text{Covariance Matrix}}$ → $\boxed{\text{Eigen Matrix and Eigen Value}}$ → $\boxed{\text{Variance}}$

Step by step Solution

$f_1 \mid f_2 \mid f_3$

Step 1: Mean centered →





Step 2: find Covariance matrix

$$
\begin{array}{c|ccc|}
 & & & \\
f_1 & Var(f_1) & C(f_1 f_2) & C(f_1 f_3) \\
f_2 & C(f_1 f_2) & Var(f_2) & C(f_2 f_3) \\
f_3 & C(f_3 f_1) & C(f_3 f_2) & Var(f_3) \\
 & f_1 & f_2 & f_3
\end{array}
\rightarrow \text{cov Matrix}
$$

Step 3: find the eigen value / vector

3D → 3 Vectors

λ_1 λ_2 λ_3

PC1 PC2 PC3

1D

2D

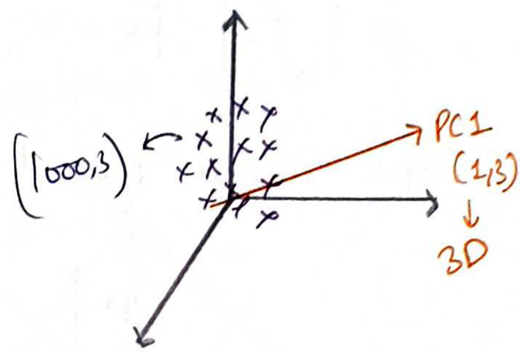
* Max eigen value is PC1 and second Max eigen value is PC2 and PC3

* We can make 2D data and 1 D data

How to transform points?

f1 | f2 | f3 | target

1000 dataset



$U^T \cdot X$

$(1000,3) \cdot \underbrace{(3,1)}_{\text{Transpose}}$ → PC1

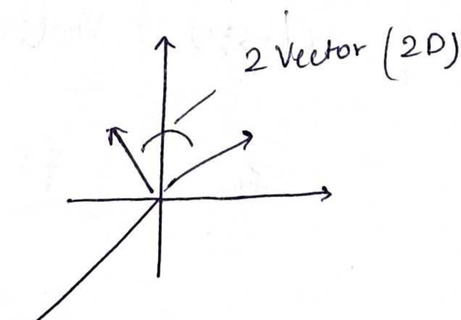
PC1 | target

$(1000,\underbrace{3) \cdot (3}_{\text{same}},1)$

$(1000,1) \rightarrow$ 1-D

for 2-D

$(1000,\underbrace{3) \cdot (3}_{\text{same}},2) = (1000,2)$



PC1 | PC2 | target

Finding optimum number of Principle Components

how percent explain

$eV = \lambda$

$784 \rightarrow \lambda_1, \lambda_2, \lambda_3 \ldots\ldots \lambda_{784}$

convert %

λ means that this principal component individually how much variance explain from the original value.

$$\left(\frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots \lambda_{784}}\right) \times 100 \rightarrow \text{percent}$$

let λ_1, λ_2, λ_3 ---- λ_{15}

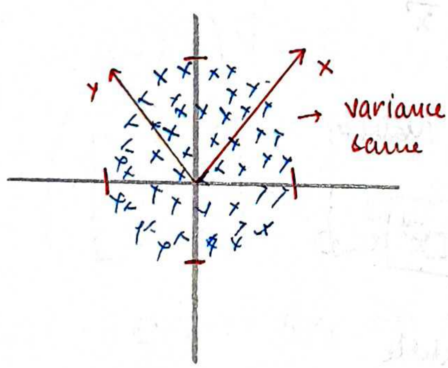
30% 25% 15% ------

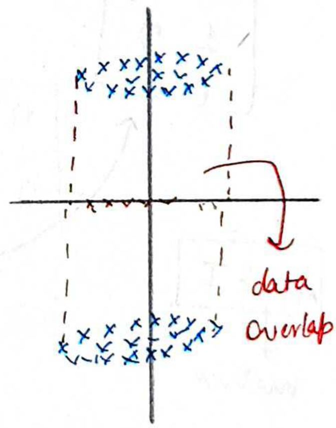
all add %

λ_1 — λ_{15}

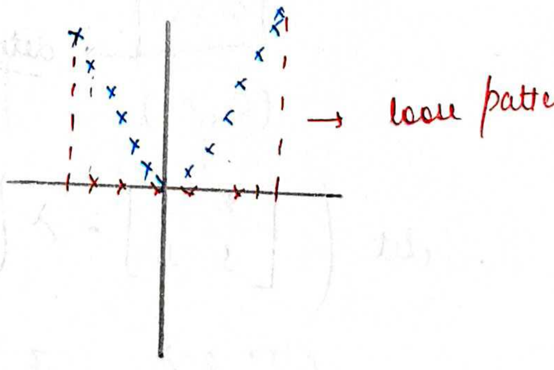
90% complete

When PCA does not Work



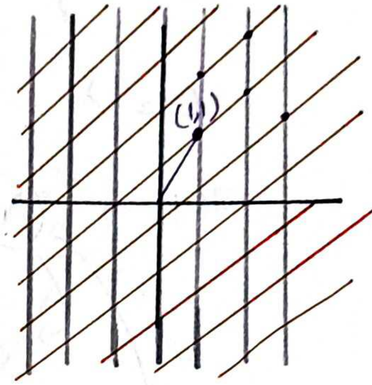


* If data make special form or pattern and convert or transform into lower Dimension might chance to loose the information of pattern.



What is Matrices?

$$\begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix} \rightarrow \text{Linear Transform}$$



$$\begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

How to calculate Eigen Vectors and Eigen Values

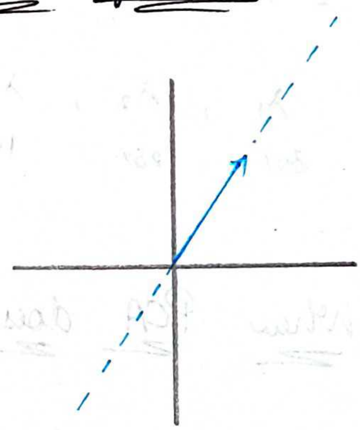
$$\begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}$$

$A\vec{x} = \lambda \vec{x}$

$A \rightarrow$ matrix

$\vec{x} =$ Vector

$\lambda \rightarrow$ Scaler



$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$A\vec{x} = \lambda I \vec{x}$

$\uparrow$

Identity Matrix

$\boxed{A - \lambda I}$

$\downarrow$

matrix

$\boxed{A\vec{x} - \lambda I \vec{x}} = 0$

$\uparrow$

non-Invertable

$\hookrightarrow$ determinate $= 0$

$\hookrightarrow$ 2d $\rightarrow$ 1D

$$\det\left(\begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) = 0$$

$$\det\left(\begin{bmatrix} 2-\lambda & 3 \\ 1 & 1-\lambda \end{bmatrix} \right) = 0$$

$$(2-\lambda)(1-\lambda)+3=0$$

$$2-2\lambda-\lambda+\lambda^2+3=0$$

$$\boxed{\lambda^2-3\lambda+5=0}$$ ← after solve this eqn

↑ Not eigen value

Another matrix

$$\det\left(\begin{bmatrix}2 & 3\\ 0 & 1\end{bmatrix}-\begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}\right)=0$$

$$\det\left(\begin{bmatrix}2-\lambda & 3\\ 0 & 1-\lambda\end{bmatrix}\right)=0$$

2 eigen values

$$(2-\lambda)(1-\lambda)=0 \qquad \boxed{\lambda=2} \qquad \boxed{\lambda=1}$$

Eigen matrix

$$(A-\lambda I)\vec{V}=0$$

$$\left(\begin{bmatrix}2 & 3\\ 0 & 1\end{bmatrix}-\begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix}\right)\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$$

$$0x_1+3x_2=0 \qquad \boxed{3x_2=0}$$

$$0x_1-x_2=0 \qquad \boxed{x_2=0}$$

$$\boxed{x_2=0},\ \boxed{x_1=1} \qquad (1,0)$$

$$\begin{bmatrix}2 & 3\\ 0 & 1\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}=\begin{bmatrix}2\\ 0\end{bmatrix}$$

$$V \xrightarrow[\text{double}]{} 2V \qquad \boxed{\lambda=2}$$

$$\left(\begin{bmatrix}2 & 3\\0 & 1\end{bmatrix} - \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

$$\begin{bmatrix}1 & 3\\0 & 0\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

$x_1 + 3x_2 = 0$

$\boxed{x_1 = -3x_2}$ → If $x_2 = 1$ $x_1 = -3$

$[-3, 1]$

Proof

$$\begin{bmatrix}2 & 3\\0 & 1\end{bmatrix}\begin{bmatrix}-3\\1\end{bmatrix} = \begin{bmatrix}-6+3\\0+1\end{bmatrix} = \begin{bmatrix}-3\\1\end{bmatrix}$$

Same and ($\lambda = 1$) because $\lambda = 1$

* After Using Eigen Value and Eigen Vect Matrix the span will be same and magnitude change.

Properties

1. **Sum of Eigen Values:** The sum of all the Eigen values of a matrix is equal to its trace (the sum of the ~~degree~~ diagonal elements of the matrix). This holds true regardless of whether the matrix is square or not.

$$\begin{bmatrix}2 & 3\\0 & 1\end{bmatrix}$$

$\boxed{\lambda = 2}$ $\boxed{\lambda = 1}$

2. Product of Eigenvalues : The product of all ~~element~~ eigenvalues of a matrix is equal to its determinant. This also holds for square matrices.

$$\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} \rightarrow 2 \times 1 - 0 \times 3 = \boxed{2} \leftarrow \text{determinant}$$

Matrics

same.

$\boxed{\lambda = 2}$ $\boxed{\lambda = 1}$

eigen values

Product of eigen values $= 2 \times 1 = \boxed{2}$

Imp

3. Eigenvectors correspondings to different eigenvalues are ~~orga~~ orthogonal: If a matrix A is symmetric (i.e., A = A^T), the eigenvectors corresponding to distinct eigenvalues are orthogonal to each other.

orthogonal ⇒

not correlated

symmetric ⇒ $\begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$

After Transpose ⇒ result same $A^T = A$

4. Eigenvalue of a Identity Matrix : For an identity matrix, the eigenvalues are all 1, regardless of the dimension of the matrix.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \longrightarrow \boxed{\lambda = 1} \quad \boxed{\lambda = 1}$$

5. **Eigenvalue of a Scale Multiple:** If B is a matrix obtained by multiplying a scaler c to a matrix A (i.e $B = cA$) then the eigenvalues of B are just the eigenvalues of A each multiplied by c.

$$\underset{\text{Scaler}}{c}\ \underset{\lambda_1\ \lambda_2}{\begin{bmatrix} 0 & 1 \\ 3 & 4 \end{bmatrix}} \rightarrow \boxed{c\lambda_1,\ c\lambda_2}$$

6. **Eigenvalues of a Diagonal Matrix:** For a diagonal matrix, the eigenvalue are the diagonal element themselves.

$$\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} \longrightarrow \boxed{\lambda = 3} \quad \boxed{\lambda = 4}$$

7. **Eigenvalues of a Transposed Matrix:** The eigenvalues of a matrix and its transpose are the same.

$$A \text{ and } A^T \rightarrow \text{same } \lambda$$