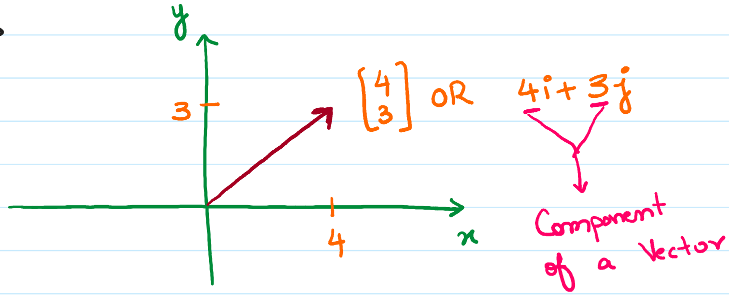
PCA - Tutorial

Monday, February 10, 2025 3:00 AM

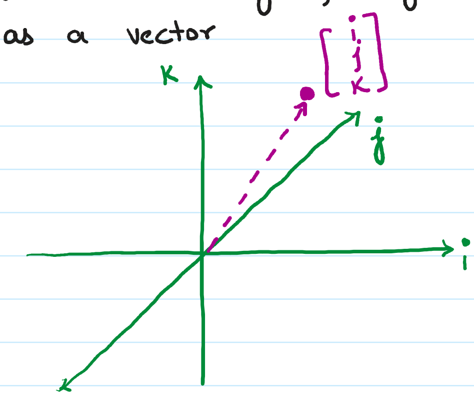
Mathematical Basics

Vectors



A Vector has both Magnitude + Direction
It helps represent Data Points in many dimensions.
for example a person with height, weight and age can be stored as a vector

$$\begin{bmatrix} \text{height} \\ \text{Weight} \\ \text{Age} \end{bmatrix}$$



Matrix

A big table of Numbers

$$\underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}}_{\text{Columns}} \left.\vphantom{\begin{bmatrix} 1 \\ 3 \end{bmatrix}}\right\} \text{Rows}$$

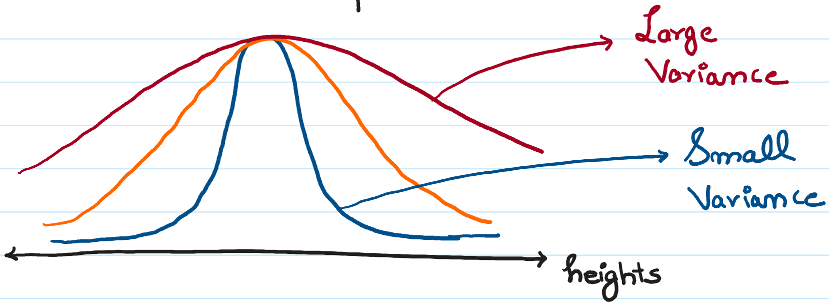
A lot of Vectors can be stacked in one matrix

$$\begin{bmatrix} | & | & | & | \\ v_1 & v_2 & v_3 & v_4 \\ | & | & | & | \end{bmatrix}$$

Important Quantities like the Covariance Matrix are also matrices

Variances & Covariance

Variance :- A fancy word for "how spread our data is"



If you have a list of Numbers, (like heights of 1000 people), variance tells you whether those heights are generally close to each other or all over the place

$$(X - \text{mean}(X))^2$$

Covariance :- Measures how to 2 variables move together

Suppose 2 variables
i) height ii) Weight

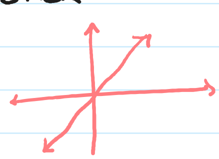
If taller people tend to be heavier, then weight goes up, height goes up.
Covariance is Positive

If height & weight have no relationship
Covariance is around Zero

Note :- Variance is basically the Covariance of a variable with itself

Covariance Matrix

If you have several Variables, you can stick pairwise Covariances into a Matrix



Lets take the same example

Lets take the same example

$$
\begin{array}{c} \\ \text{height}(H) \\ \text{Weight}(W) \\ \text{Age}(A) \end{array}
\begin{array}{c}
\begin{array}{ccc} \text{height}(H) & \text{Weight}(W) & \text{Age}(A) \end{array} \\
\begin{bmatrix}
Var(H) & Cov(H,W) & Cov(H,A) \\
Cov(W,H) & Var(W) & Cov(W,A) \\
Cov(A,H) & Cov(A,W) & Var(A)
\end{bmatrix}
\end{array}
$$

Why Repeat this LA Jargon :-

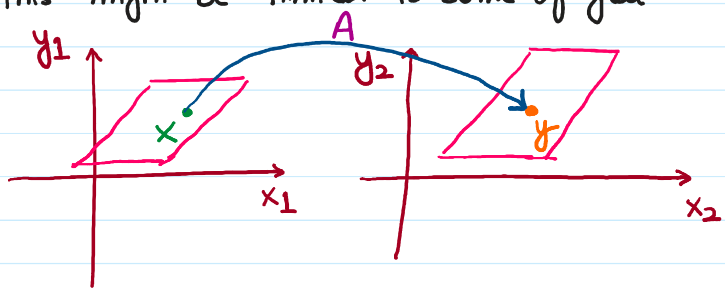
1- For PCA, Everything revolves around this Covariance Matrix.
2- It tells us which variables move together & which are independent.
3- Some of you really don't know LA.

Eigen Value & Vectors (The heart of PCA)

$$x \longrightarrow \boxed{A} \longrightarrow y$$

$$y = Ax$$

This might be familiar to some of you



Projecting from One space to Another

This is a basic 2D to 2D transformation
When you have a Matrix A, you can think of it as a machine that transforms any vector you pass into it

→ An Eigen Vector :- A special vector (x_E) that, when you feed it into Matrix (A), it doesn't change direction _ it might get 'shrunk' or 'extended' or 'extended'

→ The Amount it has shrunked or Extended is called Eigen Value

Formally :-

$$Av = \lambda v$$

Transformation Matrix

Eigen Value

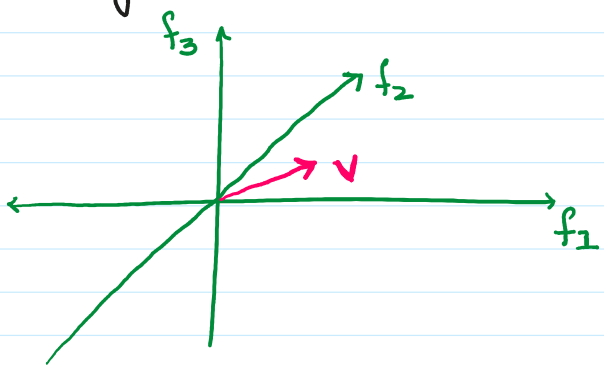
Eigen Vector

(In PCA, A is the Covariance Matrix we will discuss that below)

Connecting Covariance with Eigen Value & Vectors

Remember Covariance Matrix (Σ) tells us how Variables/ features / column correlates with every other feature

Variance along a direction :-



$f_i \rightarrow i^{th}$ feature for $i \in [1, 3]$

Lets say you pick a direction (unit vector v), you can measure how spread out (Variance) the data is in that direction

Mathematically :- $v^T \Sigma v$

Covariance Matrix

→ Scalar Value that tells you how large the spread is when you look at data along v

★ If you want v direction to capture the Greatest variance (the direction in which data is most spread out), you want to pick v to Maximise $v^T \Sigma v$

★ How to do that ? Eigen Vectors.
A fundamental result in Linear Algebra (look it up)

★ How to do that? Eigen Vectors.

A fundamental result in Linear Algebra (look it up) says that the vector v that Maximises $v^T \Sigma v$ is the eigen Vector Corresponding to Largest eigen Value of covariance Matrix (Σ)

★ Hence

Largest Eigen Value → Maximum Variance

Associated Eigen Vector → Direction in which data most spread out

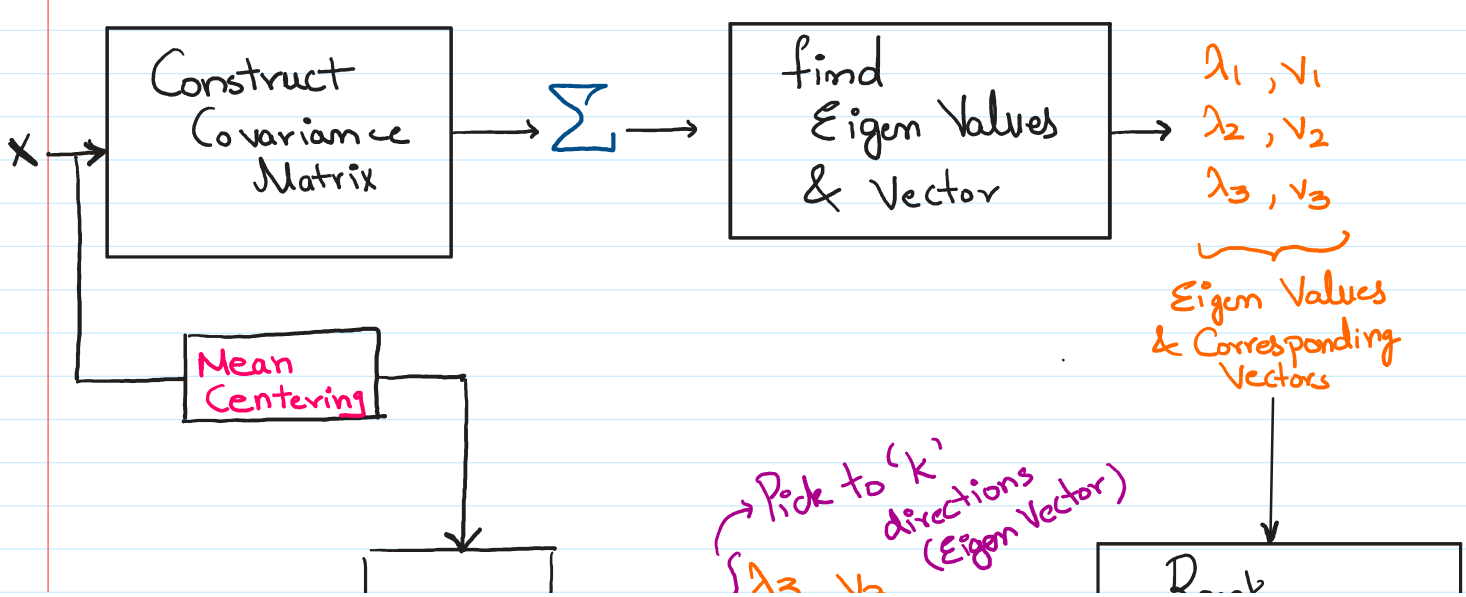
Naturally, we want to find directions (Eigen Vectors) along which our data is most spread out

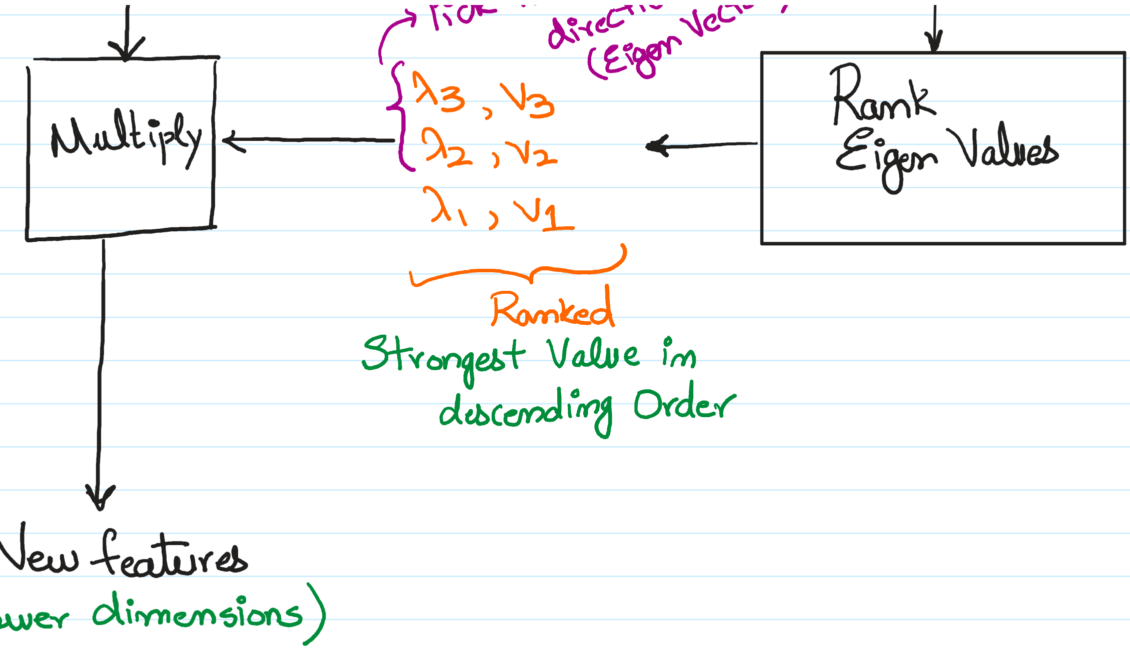
→ These directions become the Principal Components in PCA :)

Note :- Eigen Values tells us how important each direction is

By this point, I should tell you now this is PCA hidden under a fancy Name

Flow Chart





Why do all of this?

1. Dimensionality Reduction
2. Noise Reduction
3. Visualization

Lets do one Example

$$X = \begin{bmatrix} 2 & 4 & 5 & 3 & 2 \\ 2 & 3 & 4 & 4 & 3 \end{bmatrix}$$

6 Samples in 2D

Step (1) Calculate Mean

$$E(\vec{X}) = \frac{1}{6}\begin{bmatrix} 2+4+5+3+2 \\ 2+3+4+4+3 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 3.5 \end{bmatrix}$$

6 → Total # of samples

Step (2) Find Covariance Matrix

Step 2 Find Covariance Matrix

$$R_{xx} = \frac{1}{N-1}\sum_{i=0}^{N-1}\left(\vec{x}_i - E(\vec{x})\right)\left(x_i - E(\vec{x})\right)^T$$

$$= \frac{1}{6-1}\sum_{i=0}^{6-1}\left(\left(\vec{x}_i - E(\vec{x})\right)\left(x_i - E(\vec{x})\right)^T\right)$$

$$= \left\{\begin{bmatrix}2\\2\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}\left\{\begin{bmatrix}2\\2\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}^T + \left\{\begin{bmatrix}4\\3\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}\left\{\begin{bmatrix}4\\3\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}^T$$

$$\left\{\begin{bmatrix}5\\4\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}\left\{\begin{bmatrix}5\\4\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}^T + \left\{\begin{bmatrix}5\\5\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}\left\{\begin{bmatrix}5\\5\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}^T$$

$$\left\{\begin{bmatrix}3\\4\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}\left\{\begin{bmatrix}3\\4\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}^T + \left\{\begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}\left\{\begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}3.5\\3.5\end{bmatrix}\right\}^T$$

$$= \begin{bmatrix}1.9 & 1.1\\ 1.1 & 1.1\end{bmatrix}$$

Step 3 Find Eigen Values of Covariance

$$0 = \left|R_{xx} - \lambda I\right| = \begin{vmatrix}1.9-\lambda & 1.1\\ 1.1 & 1.1-\lambda\end{vmatrix}$$

ad - bc

$$(1.9-\lambda)(1.1-\lambda) - (1.1)(1.1)$$

$$\lambda^2 - 3\lambda + 0.88 = 0$$

$$\lambda_1 = 2.67 \qquad \lambda_2 = 0.33$$

Step 4 Find Eigen Vectors of Covariance

$$\begin{bmatrix}1.9-2.67 & 1.1\\ 1.1 & 1.1-2.67\end{bmatrix}\begin{bmatrix}\phi_{11}\\ \phi_{21}\end{bmatrix} = 0$$

$$\phi_{11} = 1.43\,\phi_{21} \rightarrow (1)$$

$$\begin{bmatrix}1.9-0.33 & 1.1\\ 1.1 & 1.1-0.33\end{bmatrix}\begin{bmatrix}\phi_{12}\\ \phi_{22}\end{bmatrix} = 0$$

$$\phi_{12} = -0.7\,\phi_{22} \rightarrow (2)$$

Step 5 Normalize

$$\langle \vec{\phi}_1, \vec{\phi}_1 \rangle = 1 \rightarrow \phi_{11}^2 + \phi_{21}^2 = 1$$

$$\langle \vec{\phi}_1 , \vec{\phi}_1 \rangle = 1 \rightarrow \phi_{11}^2 + \phi_{21}^2 = 1$$

so $$(1.43\phi_{21})^2 + \phi_{21}^2 = 1$$

$$(1.43+1)^2 \phi_{21}^2 = 1$$

$$\phi_{21}^2 = \mp\sqrt{\frac{1}{3.0499}}$$

$$\approx \mp 0.57$$

$$\phi_{11} = 1.43 \times 0.57 \approx 0.82$$

Repeat for

$$\langle \phi_2 , \phi_2 \rangle = 1$$

So you get $$\begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} = \begin{bmatrix} 0.82 & -0.57 \\ 0.57 & 0.82 \end{bmatrix}$$

⑥ Transform the input

$$Y = \phi^T X = \begin{bmatrix} 0.82 & 0.57 \\ -0.57 & 0.82 \end{bmatrix} \begin{bmatrix} 2+4+5+3+2 \\ 2+3+4+4+3 \end{bmatrix}$$

$$Y = \begin{bmatrix} 2.78 & 4.99 & 6.38 & 6.95 & 4.74 & 3.35 \\ 0.5 & 0.18 & 0.43 & 1.25 & 1.57 & 1.32 \end{bmatrix}$$

⑦ Covariance of Transformed Data

$$R_{YY} = \begin{bmatrix} 2.67 & 0 \\ 0 & 0.33 \end{bmatrix}$$

$$\frac{2.67}{2.67+0.33} \times 100\% = 89\%$$

$$\frac{0.33}{2.67+0.33} \times 100\% = 11\%$$

⇒ You May Ask here, why didn't We remove the Second Dimension here

Let's do it :-

$$Y = \begin{bmatrix} 2.78 & 4.99 & 6.38 & 6.95 & 4.74 & 3.35 \end{bmatrix}$$

$$X' = \Phi Y'$$

$$\underset{2\times 1}{\begin{bmatrix} 0.82 \\ 0.57 \end{bmatrix}} \underset{1\times 6}{\begin{bmatrix} 2.78 & 4.99 & 6.38 & 6.95 & 4.74 & 3.35 \end{bmatrix}}$$

$$X' = \underset{2\times 6}{\begin{bmatrix} 2.28 & 4.1 & 5.23 & 5.7 & 3.89 & 2.75 \\ 1.58 & 2.84 & 3.64 & 3.96 & 2.7 & 1.91 \end{bmatrix}}$$

$\Downarrow$ Compare with

$$X = \begin{bmatrix} 2 & 4 & 5 & 5 & 3 & 2 \\ 2 & 3 & 4 & 5 & 4 & 3 \end{bmatrix}$$

$$J \Rightarrow E\left(|X - X'|^2 \right)$$