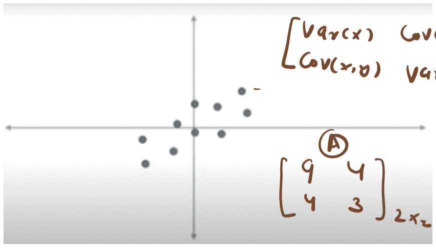
PCA

Saturday, October 23, 2021 8:29 AM

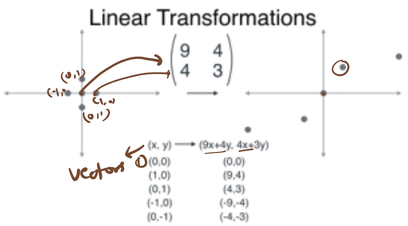


$$\begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{bmatrix} = \text{Var-cov Matrix} \qquad V = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

Ⓐ $\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}_{2\times 2}$

$[[1\ 3],\ [2\ 4]]$ numpy $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}_{2\times 2}$

pd.DataFrame(x) [ ]



vector ①

What is a linear transformations ??

$y = \beta_1 x_1 + \beta_2 x_2$ $\boxed{y = 2x}$

$2\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$ $\beta_1 \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} + \beta_2 \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$

$y = f(x)$ — ①

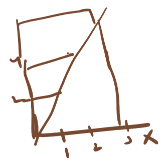
Transformation

$= (9x+4y)\ (4x+3y)$

$2\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ f will be linear

$T(u+v) = T(u) + T(v)$

[



x $(a+b)(x)$

$g(f(x))$

$2\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}$

$2\begin{bmatrix} 9 \\ 4 \end{bmatrix} + 2\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ $3\begin{bmatrix} 18 & 8 \\ 8 & 6 \end{bmatrix} = 3T\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}$ $\begin{bmatrix} 1 \\ 2 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix}$

$3\begin{bmatrix} 18 & 8 \\ 8 & 6 \end{bmatrix}$ $= ⓪$

$\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}$

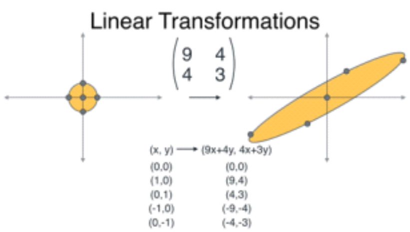
$(9\times 3 - 4\times 4) = \boxed{11}$

$27 - 16$

$\boxed{\lambda x}$ ei

$\boxed{A}\boxed{x} = \boxed{\lambda}\boxed{x}$ eigen value

Square (eigen vector)



$Ax = \lambda x$ A- var-cov Matrix

x- eigen vector

λ-eigen value (scaler)

$$\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}_{2\times 2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$A \cdot x = \lambda \cdot x$$

$$(Ax - \lambda x) = 0$$

$$x(A - \lambda I) = 0 \quad \text{if } x \neq 0 \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$(A - \lambda I) = 0$$

$$\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = 0$$

$$\Rightarrow \begin{pmatrix} 9-\lambda & 4-0 \\ 4-0 & 3-\lambda \end{pmatrix} = 0 \quad * \text{Determinant}$$

$$(9-\lambda)\cdot(3-\lambda) - 16 = 0$$

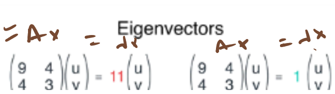
$$27 - 3\lambda - \lambda^2 - 9\lambda - 16 = 0$$

$$11 - 12\lambda - \lambda^2 = 0$$

$$(\lambda - 11)\cdot(\lambda - 1) = 0$$

Two solution $\lambda = 11,\ \lambda = 1$

$\Rightarrow Ax = \lambda x$ $\quad Ax = \lambda x$



where $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$= Ax = \lambda x$

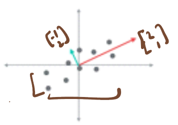
Eigenvectors

$\begin{pmatrix}9 & 4\\4 & 3\end{pmatrix}\begin{pmatrix}u\\v\end{pmatrix} = 11\begin{pmatrix}u\\v\end{pmatrix}$ $\quad$ $\underbrace{\begin{pmatrix}9 & 4\\4 & 3\end{pmatrix}\begin{pmatrix}u\\v\end{pmatrix}}_{Ax} = 1\begin{pmatrix}u\\v\end{pmatrix}$ $\quad$ where $\begin{pmatrix}u\\v\end{pmatrix} = \begin{pmatrix}x_1\\x_2\end{pmatrix}$

$\begin{pmatrix}u\\v\end{pmatrix} = \begin{pmatrix}2\\1\end{pmatrix}$ $\quad$ $\begin{pmatrix}u\\v\end{pmatrix} = \begin{pmatrix}-1\\2\end{pmatrix}$ — Two eigenvectors corresponding to two eigen values

$\lambda = ?$ $\quad$ $\begin{bmatrix}9 & 4\\4 & 3\end{bmatrix}\begin{bmatrix}-1\\2\end{bmatrix} = \lambda\begin{bmatrix}-1\\2\end{bmatrix} = 1$

Principal Component Analysis (PCA)



$\Sigma = \begin{pmatrix}9 & 4\\4 & 3\end{pmatrix}$ A

$\begin{pmatrix}2\\1\end{pmatrix}$ $\begin{pmatrix}-1\\2\end{pmatrix}$ Eigenvectors (direction)

11 $\quad$ 1 $\quad$ Eigenvalues (magnitude)

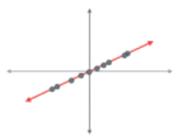
The Eigen vectors tell us which direction is our data is
The Eigen value tells - what is the magnitude of that spread is
Both vectors are perpendicular to each other - we designed this way because our co-variance matrix is symmetric matrix

$PC1 = A) = w_1x_1 + w_2x_2 + w_3x_3 \ldots w_{18}x_{18}$

$$\Sigma = \begin{pmatrix}9 & 4\\4 & 3\end{pmatrix}$$

Now the question is - which line is more important - The larger line or with a larger eigen value??

Principal Component Analysis (PCA)

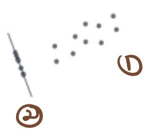


$\Sigma = \begin{pmatrix}9 & 4\\4 & 3\end{pmatrix}$

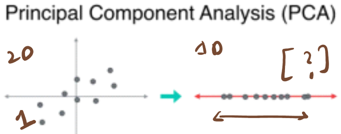
$\begin{pmatrix}2\\1\end{pmatrix}$ Eigenvectors (direction)

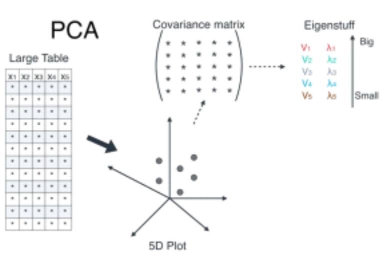
11 Eigenvalues (magnitude)

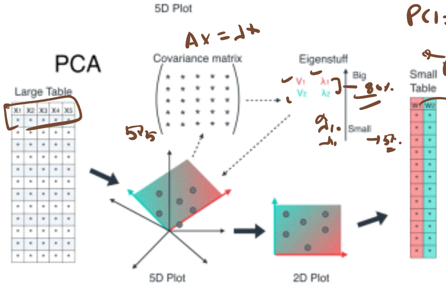
Dimensionality Reduction



Principal Component Analysis (PCA)







$PC1 = w_1x_1 + w_2x_2 + w_3x_3 - w_5x_5$

$PC2 = w_1x_1 \ldots w_5x_5$

Small Table W_1 W_2 → eigenvector

① Standardize
② Cov. matrix
③ eigenvalue, eigenvector
④ - fi) find
→ PCA

$\begin{bmatrix}cov(x,x) & cov(x,y)\\cov(y,x) & cov(y,y)\end{bmatrix}$