Dimensionality Reduction

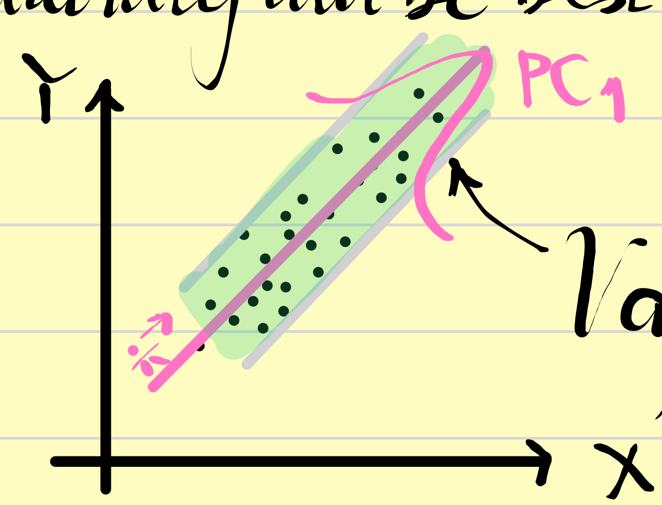
Examples Multidimensional Systems

- Management System
- Images / Videos

Principal Component Analysis (PCA)

Definition: a PC is the Eigenvector of the Covariance Matrix.

Intuition: the PC$_1$ is the direction in which variability can be best explained.



Variability is smallest in this direction

Covariance Matrix 3x3

	KPI_1	KPI_2	KPI_3
cv_1	=	=	=
cv_2	=	=	=
cv_3	=	=	=
...			

$$A = \begin{bmatrix} VAR(X) & cov(X,Y) & cov(X,Z) \\ cov(X,Y) & VAR(Y) & cov(Y,Z) \\ cov(X,Z) & cov(Y,Z) & VAR(Z) \end{bmatrix}$$

$$VAR(X) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad cov(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$\boxed{\det(A - \lambda I) = 0 \rightarrow \lambda \rightarrow \vec{v}}$$

Example for 2 KPIs:

- $KPI_1 = €/Unit = [17, 19, 23, 22] = X$
- $KPI_2 = Revenue = [34, 41, 46, 45] = Y$

$$\bar{x} = \frac{17+19+23+22}{4} = 20'25 \qquad \bar{y} = \frac{34+41+46+45}{4} = 41'5$$

$$VAR(X) = \frac{(17-20'25)^2 + (19-20'25)^2 + (23-20'25)^2 + (22-20'25)^2}{3} = 7'67$$

$$VAR(Y) = \frac{(34-41'5)^2 + (41-41'5)^2 + (46-41'5)^2 + (45-41'5)^2}{3} = 29'67$$

$$\mathrm{cov}(X,Y)=\frac{(17-20'25)(34-41'5)+(19-20'25)(41-41'5)+(23-20'25)(46-41'5)+(22-20'25)(45-41'5)}{3}$$

$$=14'5$$

$$\boxed{\mathcal{A}=\begin{bmatrix} 7'67 & 14'5 \\ 14'5 & 29'67 \end{bmatrix}}$$

$$\det(\mathcal{A}-\lambda I)=0 \rightarrow \det\left[\begin{bmatrix} 7'67-\lambda & 14'5 \\ 14'5 & 29'67-\lambda \end{bmatrix}\right]=0 \rightarrow$$

$$\rightarrow (7'67-\lambda)(29'67-\lambda)-14'5^2=0 \rightarrow$$

$$\rightarrow 7'67.29'67-(7'67+29'67)\lambda+\lambda^2-14'5^2=0$$

$$\rightarrow \lambda^2-37'34\lambda+17'32=0$$

$$\rightarrow \lambda=\frac{37'34\pm\sqrt{37'34^2-4.17'32}}{2}= \begin{cases} \lambda_1=36'87 \\ \lambda_2=0'47 \end{cases}$$

$$\boxed{\begin{array}{c} ax^2+bx+c=0 \\ x=\dfrac{-b\pm\sqrt{b^2-4ac}}{2a} \end{array}}$$

$\boxed{\lambda_1 = 36'87} \rightarrow A \cdot \vec{v_1} = \lambda_1 \vec{v_1} \rightarrow \begin{bmatrix} 7'67 & 14'5 \\ 14'5 & 29'67 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = 36'87 \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix}$

- $7'67 \cdot v_{11} + 14'5 \cdot v_{12} = 36'87 v_{11}$
- $14'5 \cdot v_{11} + 29'67 \cdot v_{12} = 36'87 \cdot v_{12}$

$\rightarrow v_{11} = \dfrac{14'5}{29'2} v_{12} = 0'49 v_{12}$

$v_{11} = 1 \rightarrow v_{12} = \dfrac{1}{0'49} = 2'04$

$\boxed{\vec{v_1} = \begin{bmatrix} 1 \\ 2'04 \end{bmatrix}}$

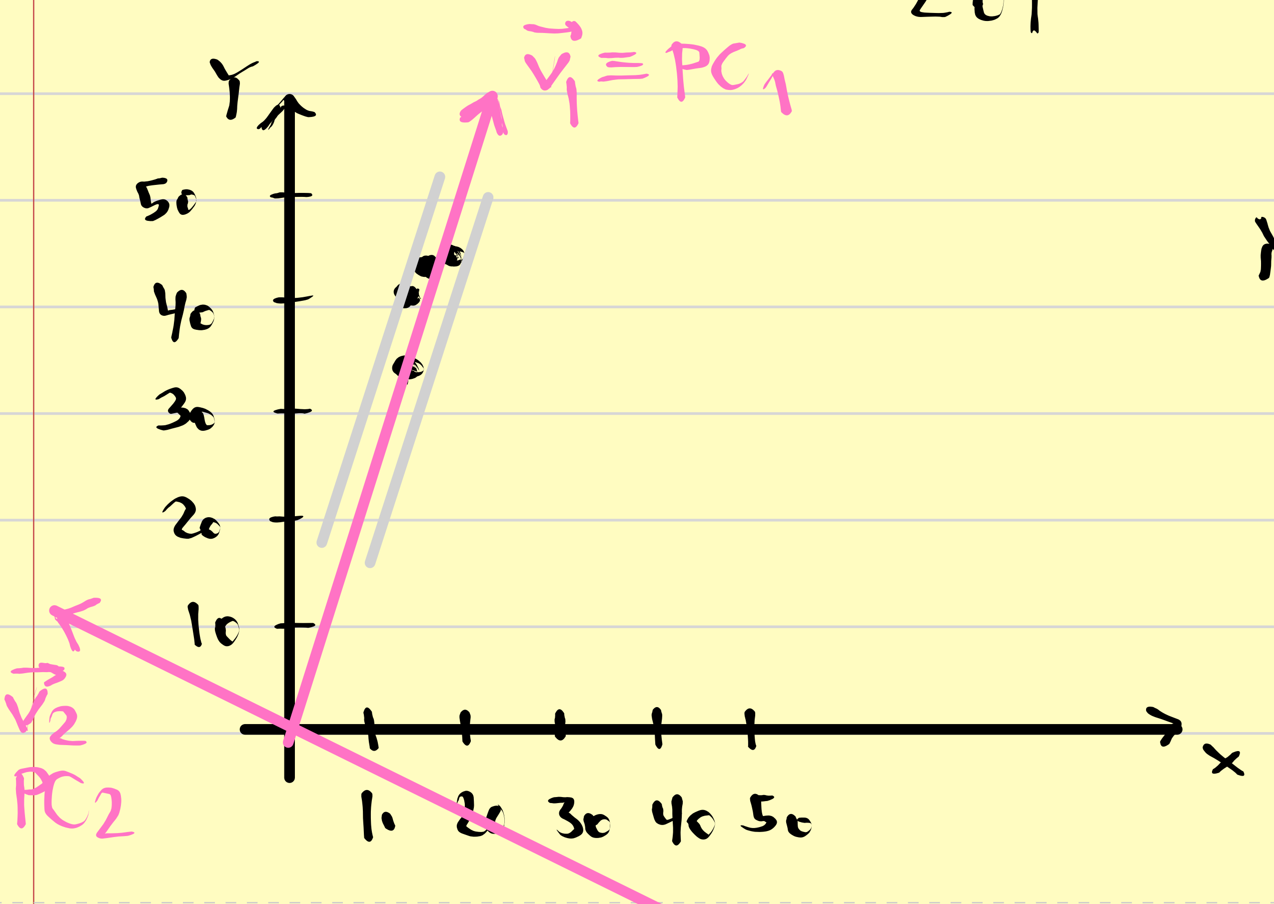
$\boxed{\lambda_2 = 0'47} \rightarrow A \cdot \vec{v_2} = \lambda_2 \cdot \vec{v_2} \rightarrow \begin{bmatrix} 7'67 & 14'5 \\ 14'5 & 29'67 \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = 0'47 \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix}$

- $7'67 \cdot v_{21} + 14'5 \cdot v_{22} = 0'47 v_{21}$
- $14'5 \cdot v_{21} + 29'67 \cdot v_{22} = 0'47 \cdot v_{22}$

$\rightarrow v_{21} = \dfrac{14'5}{-7'2} v_{22} = -2'01 v_{22}$

$v_{21} = 1 \rightarrow v_{22} = \dfrac{-1}{2'01} = -0'497$

$\boxed{\vec{v_2} = \begin{bmatrix} 1 \\ -0'49 \end{bmatrix}}$

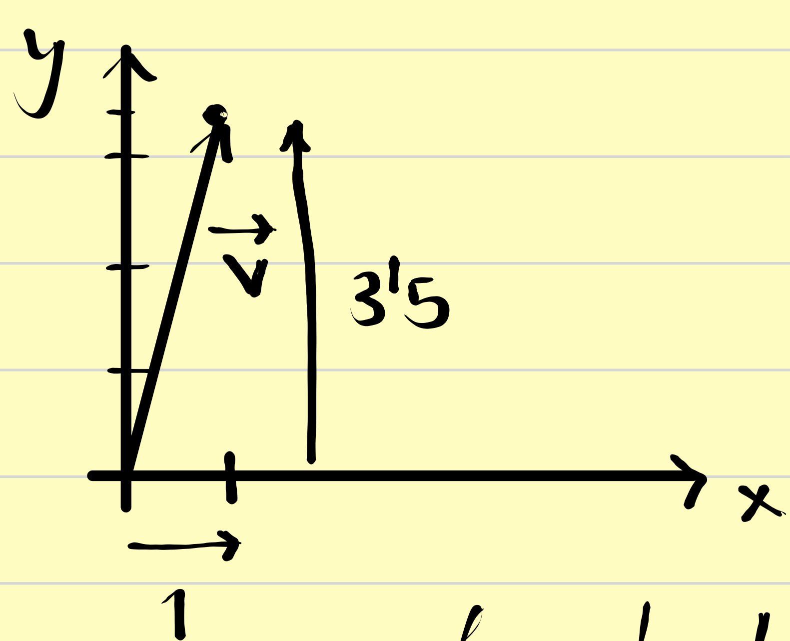


$\lambda_1 = 36'87$
$\lambda_2 = 0'47$

The amount of variability explained by each PC is given by the eigenvalues:

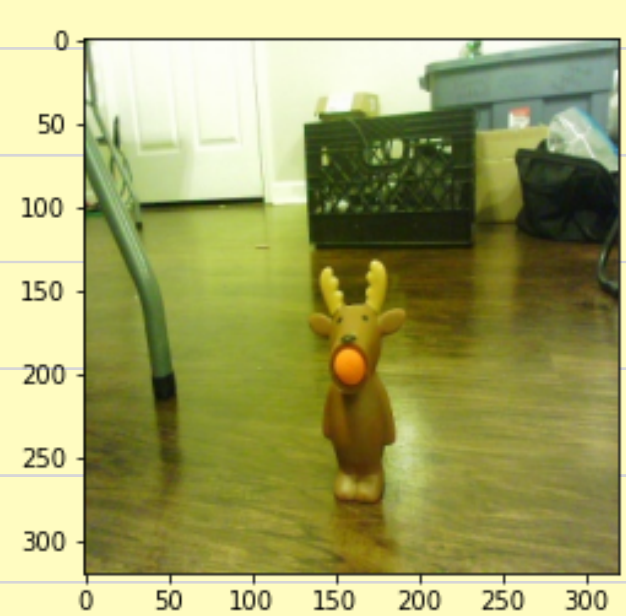
$$\%VAR(\vec{v_1}) = \frac{\lambda_1}{\sum \lambda_i} \cdot 100 = \frac{36'57 \cdot 100}{36'87 + 0'47} = 95'7\%$$

$$\%VAR(\vec{v_2}) = \frac{\lambda_2}{\sum \lambda_i} \cdot 100 = \frac{0'47 \cdot 100}{36'87 + 0'47} = 1'3\%$$

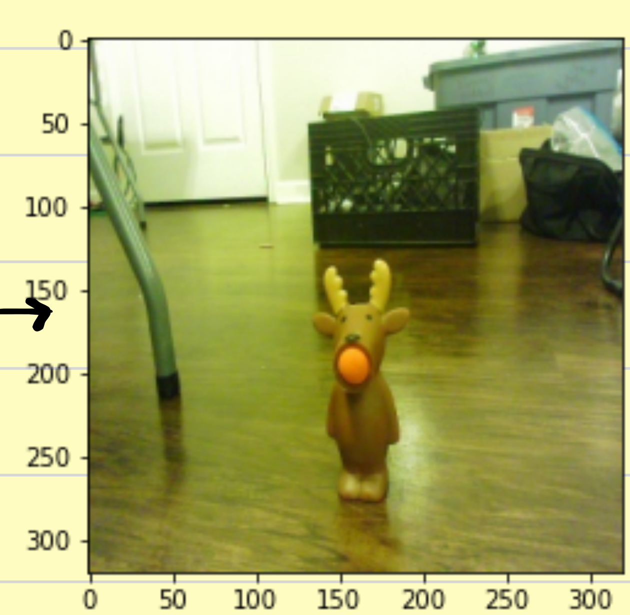


$\vec{v} \begin{bmatrix} 1 \\ 3'5 \end{bmatrix}$

how to draw a vector



310 x 310



20 PCAs

.zip Datacompression