Data $x_1 — x_n$

Mean $\langle x \rangle = \frac{1}{n} \sum_{i=1}^{n} x_i$

Variance $\sigma^2 = \frac{1}{n} \sum_{i=1}^{N} (x_i - \langle x \rangle)^2$

Std. deviation σ

mean centering / standardisation

$$x_i' = \frac{x_i - \langle x \rangle}{\sigma}$$

$\rightsquigarrow$ mean 0, std dev of one

variation of zero-mean data $\sigma^2 = \frac{1}{n} \sum_i x_i$

covariance of two zero-mean measurements
$A = a_1, —, a_n$ and $B = b_1, —, b_n$

$$\sigma_{AB} = \frac{1}{n} \sum_i a_i b_i$$

if positive $\rightsquigarrow$ correlated data
if negative $\rightsquigarrow$ negat. correlated data
if $\emptyset$ $\rightsquigarrow$ uncorrelated

vector notation

$a = (a_1 —— a_n)$
$b = (b_1 —— b_n)$

$\sigma_{ab}^2 = \frac{1}{n} ab^T$

Data given as $X = \begin{pmatrix} x_1 \\ | \\ x_m \end{pmatrix}$ ($m \times n$ matrix)

Covariance matrix $C_X = \frac{1}{n} X X^T$ ($m \times m$ matrix)

- symmetric
- diagonal elements indicate variances
- off-diagonal elements are co-variances

$\rightsquigarrow$ example in R

Principal Component Analysis

Intro ~ Slides

Note: now $\vec{x}_i$ are columns, still a $m \times n$ matrix

A basis $\mathcal{B} = \begin{pmatrix} \vec{b}_1 \\ \vec{b}_2 \\ \vdots \\ \vec{b}_m \end{pmatrix}$ Our data $X = \begin{pmatrix} \vec{x}_1 & \vec{x}_2 & \dots & \vec{x}_n \end{pmatrix}$

(I won't use $\rightarrow$)

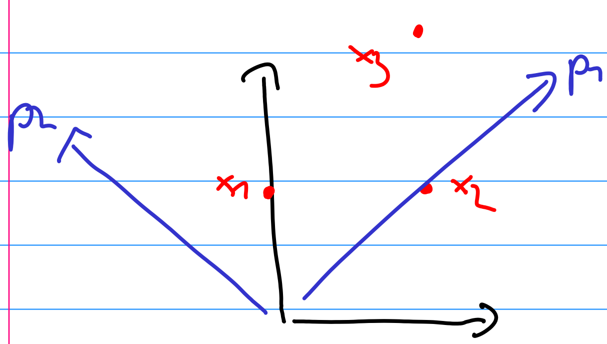
Change of basis via transformation P

$$\underset{m \times m}{P} \cdot \underset{m \times n}{X} = \underset{m \times n}{Y} \qquad P = \begin{pmatrix} \text{---}\; p_1 \;\text{---} \\ p_2 \\ \vdots \\ \text{---}\; p_m \;\text{---} \end{pmatrix} \qquad X = \begin{pmatrix} x_1 & | & x_2 & | & \dots & | & x_n \end{pmatrix}$$

P transforms X into Y (rotation, stretch)

$p_1, \dots, p_m$ are a new basis

sketchy example: (p_i are not (pairwise) orthonormal, having unit length i.e. $\|p_i\| = 1$, only orthogonal, i.e. $p_i \cdot p_j^T = 0$ for $i \neq j$)



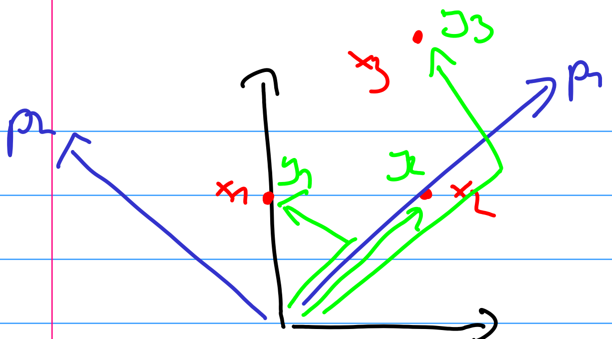
$x_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$ $\quad p_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

$x_2 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$ $\quad p_2 = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$

$x_3 = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$

$$PX = \begin{pmatrix} p_1 \\ | \\ p_m \end{pmatrix} \begin{pmatrix} x_1 & \text{---} & x_n \end{pmatrix} = \begin{pmatrix} p_1 \cdot x_1 & \cdots & p_1 \cdot x_n \\ \vdots & \ddots & \vdots \\ p_m \cdot x_1 & \cdots & p_m \cdot x_n \end{pmatrix}$$

$$= \begin{pmatrix} y_1 & | & \dots & | & y_n \end{pmatrix}$$



$$x_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \qquad p_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$x_2 = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \qquad p_2 = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$x_3 = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

$$y_1 = \begin{pmatrix} p_1 \cdot x_1 \\ p_2 \cdot x_1 \end{pmatrix} = \begin{pmatrix} 1.5 \\ -1.5 \end{pmatrix}$$

$$y_2 = \begin{pmatrix} p_1 \cdot x_2 \\ p_2 \cdot x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

$$y_3 = \begin{pmatrix} p_1 \cdot x_3 \\ p_2 \cdot x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ -2 \end{pmatrix}$$

How to choose P?

(1) minimise redundancy (measured as covariance)

(2) maximize the "signal" (measured as variance)

$$C_X \overset{P}{\curvearrowright} C_Y = \begin{pmatrix} \ddots & 0 \\ 0 & \ddots \end{pmatrix} \quad \text{possible?}$$

Find some (orthonormal) matrix P with $Y = P \cdot X$, such that $C_Y = \frac{1}{n} Y Y^T$ is a diagonal matrix.

$$C_Y = \frac{1}{n} Y Y^T$$
$$= \frac{1}{n} (PX)(PX)^T$$
$$= \frac{1}{n} P X X^T P^T$$
$$= P \frac{1}{n} X X^T P^T$$
$$= P \cdot C_X \cdot P^T$$

Some more linear algebra needed:
any symmetric matrix A can be written as $A = E \cdot D \cdot E^T$

$(e_1, \dots)$ e_i orthonormal eigenvectors

diagonal matrix of eigenvalues

choose P, such that each row p_i of P is an eigenvector of $\frac{1}{n} X X^T = C_X$ (i.e. $C_X \cdot p_i^T = \lambda_i \cdot p_i^T$)

$$C_Y = P \cdot C_X \cdot P^T$$
$$= P \cdot E \cdot D \cdot E^T \cdot P$$
$$= E^T \cdot E \cdot D \cdot E^T \cdot E$$

even more linear algebra:

1.) E is orthonormal

2.) the inverse of an orthonormal matrix is the inverse, i.e. $E^{-1} = E^T$

Proof: $(E^T \cdot E)_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases} = I$

$\rightsquigarrow \quad E^T = E^{-1}$

$= E^T \cdot E \cdot D \cdot E^T \cdot E$

$= E^{-1} \cdot E \cdot D \cdot E^{-1} \cdot E$

$= D$

cool.