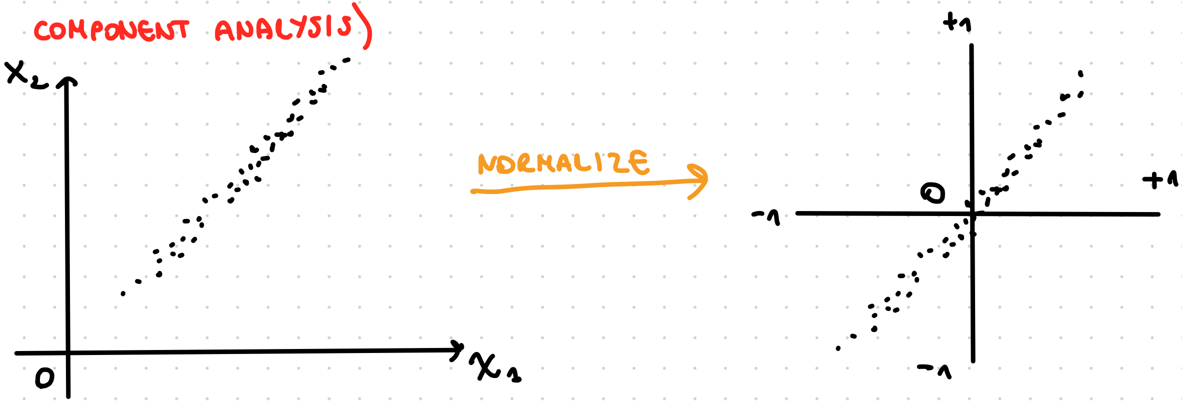
RECALL REGRESSION 7/28/2023

$x \in \mathbb{R}^d, \quad y \in \mathbb{R}$ y OBSERVED

IN CLUSTERING (PCA) WE NEED TO FIND SUBSPACES

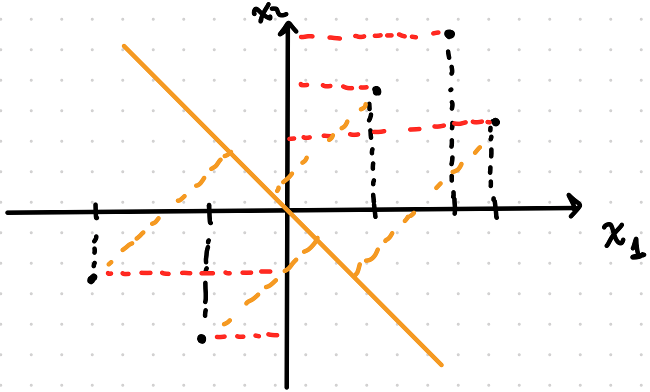
$x \in \mathbb{R}^d, \quad y \in \mathbb{R}^k \ (k < d)$ y IS NOT OBSERVED $\longrightarrow$ $y^{(i)} = \begin{bmatrix} \mu_1^T x^{(i)} \\ \mu_2^T x^{(i)} \\ \vdots \\ \mu_k x^{(i)} \end{bmatrix} \in \mathbb{R}^k$

PCA (PRINCIPAL COMPONENT ANALYSIS)

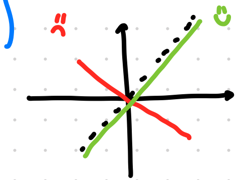


WE START BY NORMALIZING OUR DATA (CENTER, MEAN 0), WE BECOME IMMUNE TO UNIT OF REPRESENTATION (SHIFT AND SCALE AXIS)

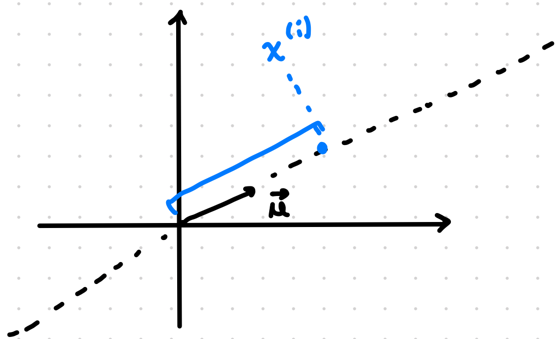
$$\boxed{x_j^{(i)} \leftarrow \frac{x_j^{(i)} - \mu_j}{\sigma_j} \qquad \mu_j = \frac{1}{n} \sum_{i=1}^{n} x_j^{(i)} \qquad \sigma_j^2 = \frac{1}{n} \sum_{i=1}^{n} (x_j^{(i)} - \mu_j)^2}$$



WE NEED TO FIND PLANE OF PROJECTION S.T. THE SPREAD OF PR. POINTS IS MAXIMUM (MAX VARIANCE)



WE NEED TO FIND $\vec{\mu}$ AND CALCULATE $x^{(i)^T}\vec{\mu} \quad \forall i$



$$= x^{(i)^T}\mu$$

$$x^{(i)^T}\mu \in \mathbb{R}$$

PSD, SIMM. MAT A

MAXIMIZE

$$\mathrm{VAR} = \frac{1}{n}\sum_{i=1}^{n}\left(\mu^T x^{(i)}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\mu^T x^{(i)} x^{(i)^T}\mu = \mu^T \overbrace{\left(\sum_{i=1}^{n} x^{(i)} x^{(i)^T}\right)}^{A}\mu$$

MAXIMIZE $\vec{\mu}$ S.T.

$$\hat{\mu} = \arg\max_{\mu} \mu^T A \mu \quad , \quad A = \frac{1}{n}\sum_{i=1}^{n} x^{(i)} {x^{(i)}}^T$$

MAXIMIZED

$\hat{\mu}$:= EIGENVECTOR CORRESPONDING TO THE LARGEST EIGENVECTOR

$$A = X^T X \quad , \quad X$$ DESIGN MATRIX (NORMALIZED)

MAXIMIZE EIGENVECTORS OF $A \longrightarrow \hat{\mu}, \dots, \hat{\mu}_k$

$$\tilde{x} : y^{(i)} = \begin{bmatrix} \hat{\mu}_1 x^{(i)} \\ \hat{\mu}_2 x^{(i)} \\ \hat{\mu}_k x^{(i)} \end{bmatrix} \in \mathbb{R}^k$$

THEN APPLY WHATEVER MODEL, ALGORITHM

ICA (INDIPENDENT COMPONENT ANALYSIS)

$s \in \mathbb{R}^d$ SIGNAL COMING OUT OF SOURCES (d SOURCES)

$A \in \mathbb{R}^{d \times d}$ MIXING MATRIX

$x \in \mathbb{R}^d$ OUR DATA

$s^{(i)}$ $\quad$ $s^{(t)}$ $\quad$ EX SOUND IN TIME, SAMPLING $\rightarrow$ EXAMPLE $s^{(i)}$

GIVEN OBSERVATION $x^{(i)}$, WE NEED TO ESTIMATE s, THIS NEEDS TO CALCULATE A

MIXED AUDIO $\quad$ SEPARATED SOUND

EX

d

	s_1	s_2	s_3	s_d
$s^{(t)}$	1			
$s^{(t+1)}$	2			
$\vdots$	$\vdots$			
$s^{(T)}$	-2			

NOTE

WE NEED TO EST. A

THEN $A^{-1} = W$ AND

$$s = Wx$$

$W = A^{-1}$ UNMIXING MATRIX

$$W = \begin{bmatrix} - w_1^T - \\ - w_2^T - \\ \vdots \\ - w_d^T - \end{bmatrix}, \quad s_j^{(i)} = w_j^T x^{(i)}$$

BUT HOW TO FIND W?

PROBLEMS

→ PERMUTATION (SWITCHING INDICES!, NOT A PROBLEM)

→ $x = As$, BUT $x = (2A)(\frac{1}{2}s)$, WE CAN RE-NORMALIZE ONCE WE GET s

→ $s \sim$ DISTRIBUTION

IF s COMES FROM $N(0, I)$ THIS IS A PROBLEM BECAUSE GIVEN A

$$x = As \longrightarrow \mathbb{E}[x] = \mathbb{E}[As] = A\,\mathbb{E}[s] = 0$$
$$Cov[x] = \mathbb{E}_{s \sim N(0,I)}[xx^T] = \mathbb{E}[Ass^TA^T] = A\,\mathbb{E}[ss^T]A^T = AA^T$$

THEN $x \sim N(0, AA^T)$ AND THAT'S A PROBLEM

SUPPOSE WE HAVE ORTHOGONAL MATRIX R S.T. $RR^T = R^TR = I$

NOW, WE CAN HAVE

$$A' = AR \quad \text{THEN} \quad x' = A's$$

$$\mathbb{E}[x'] = \mathbb{E}[A's] = A'\mathbb{E}[s] = 0 \quad \text{AGAIN}$$

$$Cov[x'] = \mathbb{E}_{x\sim(0,I)}[(x')(x')^T] = \mathbb{E}[A'ss^TA'^T] = A'\mathbb{E}[ss^T]A'^T$$

$$A'A'^T = ARR^TA^T = AA^T$$

A CAN BE ROTATED AND THE DIFFERENCE BETWEEN 2 DIFFERENT A s CAN BE INDISTINGUISHABLE LINEAR TRANSFORMATION → INDISTINGUISH IF DIST = GAUSSIAN

NOTE

$s \quad P_s(s)$

$A \quad x = As$

$P_x \quad W = A^{-1}$

$s = Wx$

$$P_x(x) \stackrel{?}{=} P_s(Wx) \quad \text{NO!}$$

DET. JACOBIAN

$$P_x(x) = P_s(Ws) \cdot |W|$$

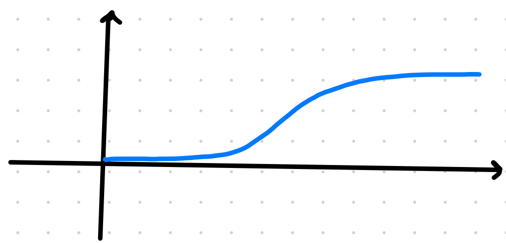
CHAIN RULE OF RANDOM VARIABLES

WITH THESE TOOLS WE CAN FIND OUR WAY OUT FOR DIST. WHICH ARE NOT NORMAL GAUSSIANS

$$p(s) = \prod_{j=1}^{n} p_s(s_j)$$

$$p(x) = \prod_{j=1}^{n} p_x(x_j) \quad (*)$$

RECALL CUMULATIVE DISTRIB. FUNCTION



SIGMOID? x_x

WE CAN USE $\sigma(z)$ AS CDF! EZ TO DIFFERENCIATE!

$$(*) \quad p(x) = \prod_{j=1}^{n} g'(w_j^T x)\,|W|$$

AND DEFINE A LIKELIHOOD

$$\boxed{\ell(W) = \sum_{i=1}^{n}\left(\sum_{j=1}^{d} \log g'(w_j^T x^{(i)}) + \log |W|\right)}$$

NOW, WE CAN TAKE THE GRADIENT W.R.T. W

$$\boxed{W \leftarrow W + \alpha \nabla_W \ell(W)}$$

WHICH, FOR THE i-TH EXAMPLE, LOOKS LIKE THIS

$$\boxed{W \leftarrow W + \alpha\left(\begin{bmatrix} \vdots \\ 1 - 2g(w_i^T x^{(i)}) \\ \vdots \end{bmatrix} x^{(i)^T} + (W^T)^{-1}\right)}$$