Session 10 PCA

In the last session we tackled SVD:

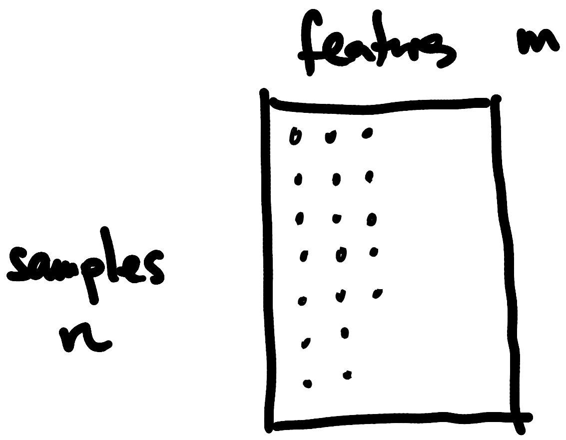
$$A = U \Sigma V^t$$

U, V: orthogonal

Σ: diagonal

$V = [v_1 | \cdots | v_m]$ eigenvectors of $A^tA \sim$ "covariance" → orthonormal basis of $\mathbb{R}^m$

$U = [u_1 | \cdots | u_n]$ eigenvectors of AA^t → orthonormal basis of $\mathbb{R}^n$



$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \\ & & \end{bmatrix}$$

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m > 0$$

$rk(A) = m$

A^tA is definite-positive

Remark:

U, V do not necessarily encode rotations; in general, we can say they encode "rigid" transformations

→ lengths, angles } does not change

$AV \longrightarrow$ cols = linear comb. of the columns of A

$$\left[\, v_1 \,\middle|\, v_2 \,\middle|\, \cdots \,\middle|\, v_m \,\right]$$

$\tilde{A} = AV$

expression

	MFG	YAG	MFG2
	feat 1	feat 2	feat 3
s1			
s2			
⋮			
sn			

A

$\rightsquigarrow$

	$\frac{1}{\sqrt{2}}(f_1+f_2)$	$\frac{1}{\sqrt{2}}(f_2+f_3)$	$\frac{1}{\sqrt{2}}(f_1+f_3)$
	↑	↑	↑
	c1	c2	c3
s_1			
⋮			
s_n			

$\tilde{A}$

$$\tilde{A}^t\tilde{A} = (AV)^t(AV)$$

$\boxed{(xy)^t = y^tx^t}$

$(xyz)^t = z^t(xy)^t = z^ty^tx^t$

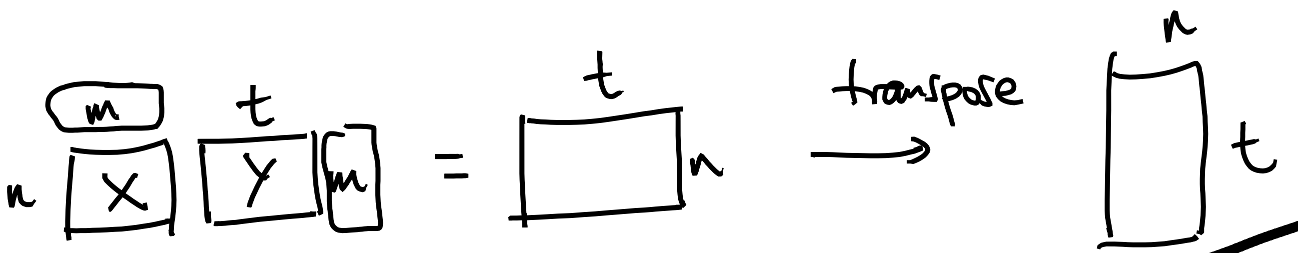
$$= V^tA^tAV$$

SVD

$$= V^t(U\Sigma V^t)^t(U\Sigma V^t)V$$

$$= V^tV\Sigma^tU^tU\Sigma V^tV =$$

$V^t = V^{-1}$
$U^t = U^{-1}$



$x^ty^t \longrightarrow ??$

$y^tx^t \longrightarrow t \times n$ size

$$= \Sigma^t\Sigma$$

$$= \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_m & \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ & & \sigma_m \\ & 0 & \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix}$$

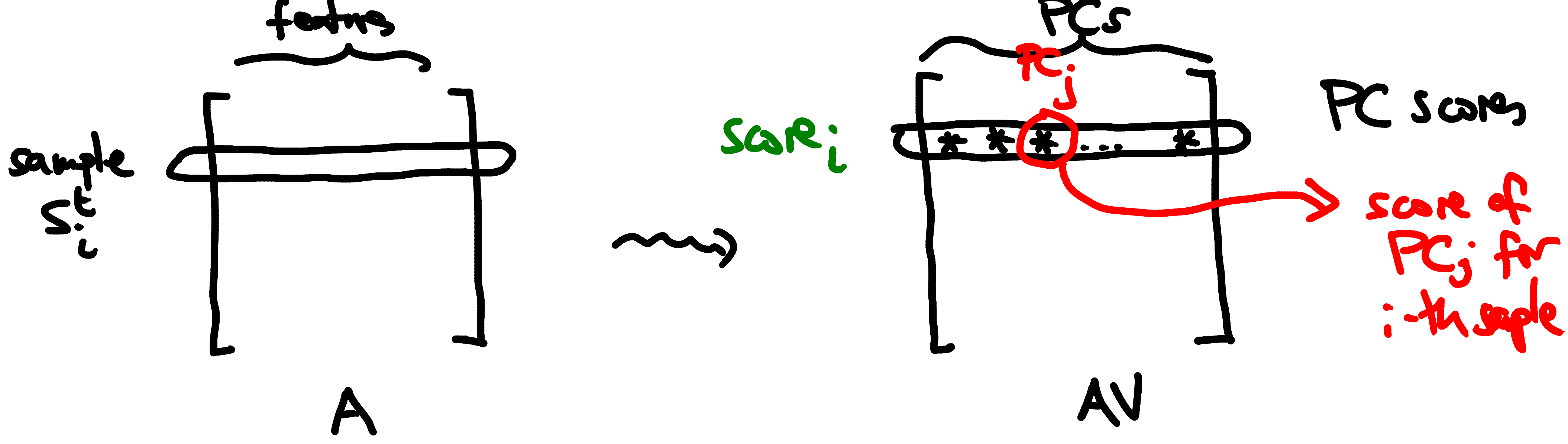
↳ diagonal matrix

variance of the transformed columns

Idea: new columns are uncorrelated

$$cov(C_i, C_j) = 0$$
for $i \neq j$

Def:
The columns of V are known Principal Components (PC)



original data

PC representation

$V = [v_1 | \dots | v_m]$

1) features ⇝ PCs

2) scores = entries of the transformed matrix AV

3) statistical information:
variance of the scores in each PC
⇝ σ_i = std,
σ_i^2 = variance
$\sum \sigma_i^2 \sim TSS$ total variance

$\sigma_1^2 \geq \dots \geq \sigma_m^2 > 0$
variance of the transformed columns

singular values σ_i are to be interpreted as std's.

Remark :

PC1 $\rightsquigarrow$ Of all the vectors in the "feature space" the one which yields maximum variance of the resulting scores.

The next PC's are chosen to fulfill this condition while being orthogonal to the previously computed PCs

What are the scores of a sample?

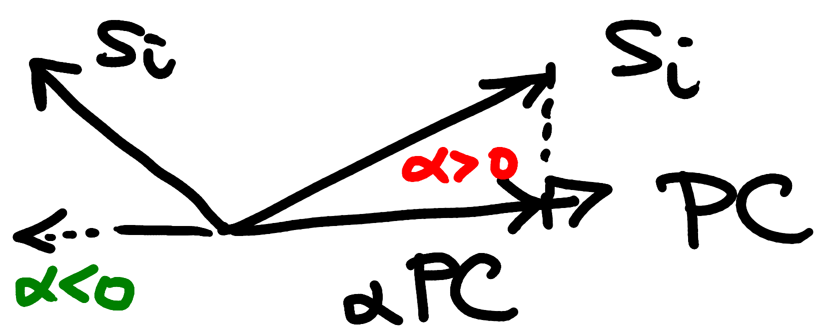
$\mathbb{R}^m \ni s_i^t$ original data vector for a given sample (entries of the i-th sample in the original data)

$\text{score}_i = [\alpha_1, \alpha_2, \dots, \alpha_j, \dots, \alpha_m]$

"orthogonal projection" of s_i^t onto PC_j

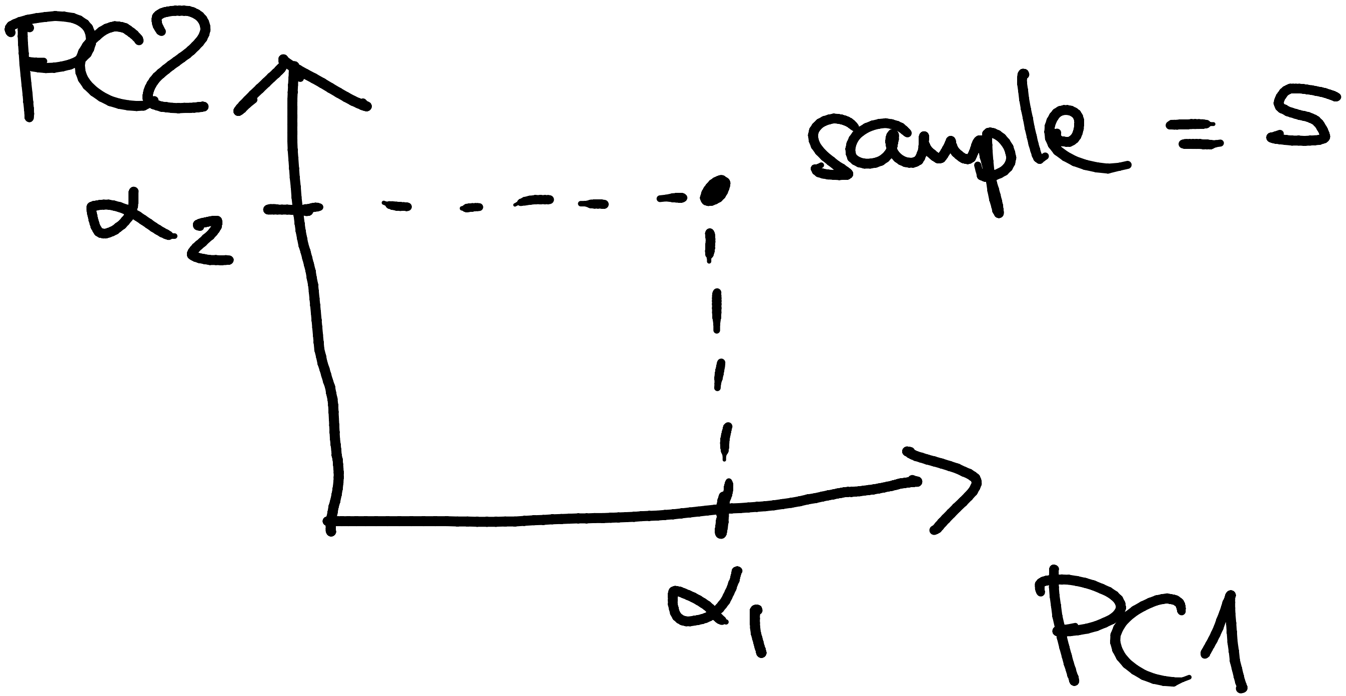
α_1 is the "orthogonal projection" of s_i^t onto the PC1

$PC1 \dots PCm$
$\cap$
$\mathbb{R}^m$



$\text{score}_i = [\alpha_1, \dots, \alpha_j, \dots \alpha_m]$

these are the coordinates of s_i^t in the basis $\{PC1, \dots, PCm\}$


PC2
sample = s
α_2
α_1
PC1