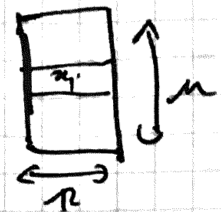
Principal Component Analysis (PCA).

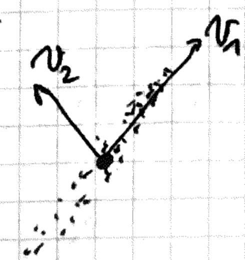
- Input: $(x_i)_{i=1}^n$ $\quad x_i \in \mathbb{R}^p \rightsquigarrow X =$ 

- Centering: $x_i \leftarrow x_i - \underbrace{\frac{1}{n}\sum_j x_j}_{\hat{m}}$

- Covariance: $C \triangleq \frac{1}{n}\sum_{i=1}^n x_i x_i^T = \frac{1}{n} X^T X \in \mathbb{R}^{p\times p}$

- SVD: $X = U\,\mathrm{diag}(\sigma_s)\,V^T \cdot \sqrt{n}$ $\quad U \in \mathbb{R}^{n\times r}$ $\quad \underbrace{V \in \mathbb{R}^{p\times r}}_{\text{important}}$.

typically $r = p \ll n$

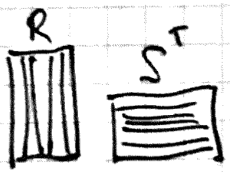


$$C = \frac{1}{n} X^T X = V\,\mathrm{diag}(\underset{\lambda_s}{\sigma_s^2})\,V^T = V \Lambda V^T$$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$$

$$V = [v_1 \dots v_r]$$

- PCA pbm: solve compression/decompression pbm



$$\min_{R,S} \Big\{ \sum_{i=1}^n \underbrace{\|x_i - R\,S^T x_i\|^2}_{f(R,S)} \; : \; R, S \in \mathbb{R}^{p\times k} \Big\} \quad (*)$$

<u>Thm</u>: A sol° of $(*)$ is $S = R = [v_1, \dots, v_k] = V_{1:k}$

<u>Rmq</u>: $f(\cdot, S)$ and $f(R, \cdot)$ are convex (least square) but f is not

<u>Heuristic</u>: minimize iteratively on $S \rightsquigarrow R \rightsquigarrow S \longrightarrow \dots$

$\Rightarrow$ in general, does not work $\Rightarrow$ Thm gives global minimum

<u>Lemma</u>: One can suppose $S = R$ and $S^T S = \mathrm{Id}_{k\times k}$

<u>Proof</u>:
- $x \mapsto (RS)x$ has rank $k' \leq k$, let W be a bon of $\mathrm{Im}(RS)$, $W^T W = \mathrm{Id}_{k'}$
- $\operatorname{argmin}_z \|x - Wz\|^2 = W^T x$ because 1st order cond°: $W^T(x - Wz) = 0$
- <u>Hence</u>: $\sum_i \|x_i - \underbrace{RS x_i}_{= W z_i \text{ for some } z_i}\|^2 = \sum_i \|x_i - W z_i\|^2 \geq \sum_i \|x_i - \underbrace{W W^T x_i}_{\text{better than } R,S}\|^2$

Lemma : $(*) \Leftrightarrow \max \left\{ \mathrm{tr}(S^T C S) : S \in \mathbb{R}^{p\times k},\ S^T S = Id_k \right\}$ $(**)$

Dem ! $(*) \underset{\text{lemma}\uparrow}{\Leftrightarrow} \min_{S^T S = Id} \sum_i \|x_i - S S^T x_i\|^2 =$

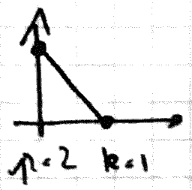
$$= \sum_i \underbrace{\|x_i\|^2}_{cst} - 2 x_i^T S S^T x_i + x_i^T S \underbrace{S^T S}_{Id_k} S^T x_i$$

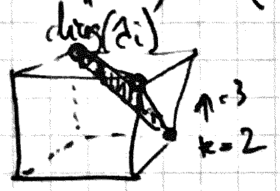
$$= - \sum_i \underbrace{x_i^T S S^T x_i}_{\in \mathbb{R}} = - \sum_i \mathrm{tr}(x_i^T S S^T x_i)$$

$$= - \sum_i \mathrm{tr}(S^T x_i x_i^T S) = - \mathrm{tr}\left(S^T \left(\sum_i x_i x_i^T\right) S\right)$$

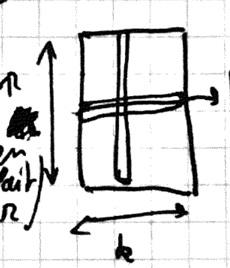
Lemma : denoting $C = V \underset{\mathrm{diag}(\lambda_i)}{\Lambda} V^T$, $\mathrm{tr}(S^T C S) \leq \max_{\substack{0 \leq \beta_i \leq 1 \\ \sum_i \beta_i \leq k}} \sum_{i=1}^{p} \lambda_i \beta_i = \sum_{i=1}^{k} \lambda_i$

$\beta = [\underbrace{1,1 \dots 1}_{k}, 0 \dots 0]$





Proof : $\mathrm{tr}(S^T C S) = \mathrm{tr}(\underbrace{S^T \tilde{V}}_{B^T} \tilde{\Lambda} \underbrace{\tilde{V}^T S}_{B})$ complete $p \times p$



$$= \mathrm{tr}(B^T \Lambda B) = \mathrm{tr}(\Lambda B B^T)$$

$$= \sum_{i=1}^{p} \lambda_i \underbrace{\|b_i\|^2}_{\beta_i}$$

- One has $B^T B = S^T \underbrace{\tilde{V} \tilde{V}^T}_{Id_p} S = S^T S = Id_k$ (complétion importante)

$$\sum_i \beta_i = \sum_i \|b_i\|^2 = \|B\|^2_{FRO} = \mathrm{tr}(B^T B) = \mathrm{tr}(Id_k) = k$$

- One extend $\tilde{B}$ in an ortho. basis  $\tilde{B}\tilde{B}^T = \tilde{B}^T \tilde{B} = Id_p$

$$0 \leq \|b_i\|^2 \leq \|\tilde{b}_i\|^2 = 1 \Rightarrow 0 \leq \beta_i \leq 1.$$

Conclusion : Setting $S = [v_1 \dots v_k]$ satisfies $CS = S\,\mathrm{diag}(\lambda_i)_{i=1}^{k}$

hence $\mathrm{tr}(S^T C S) = \mathrm{tr}(\underbrace{S^T S}_{Id_k} \mathrm{diag}(\lambda_i)_{i=1}^{k}) = \sum_{i=1}^{k} \lambda_i$ optimal