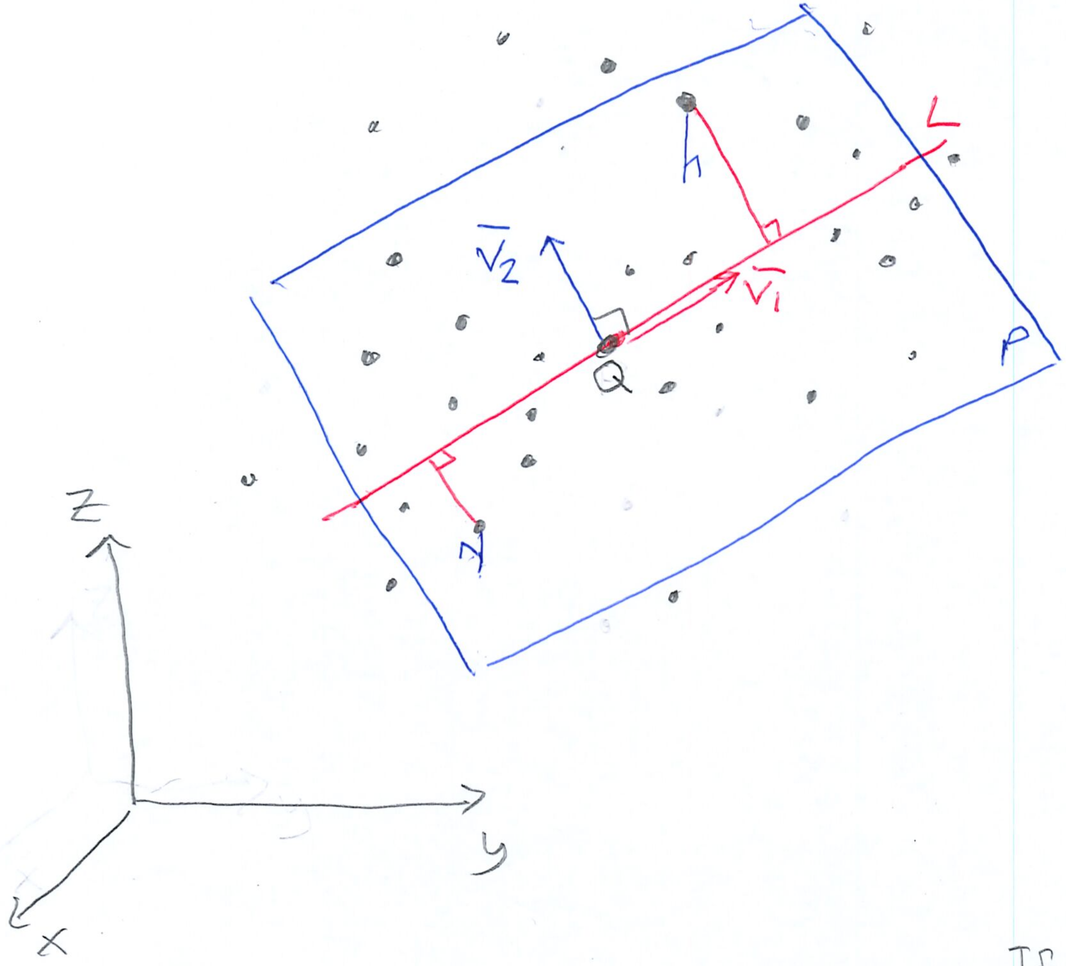
SVD and PCA



Given these data points in $\mathbb{R}^3$, how do we find the best line or plane that fits data?

Minimize sum of distances squared from data points to line or plane. Called total least squares.

How do we find a good new coordinate system? A centre point Q (new origin) and a new (ON) basis $\bar{v}_1, \bar{v}_2, \ldots$?

$\bar{v}_1, \bar{v}_2, \ldots$ are called principal components.

(and L will be inside P!)

If we have 10,000 data points in $\mathbb{R}^{20}$, how do we find the best 4-dim "plane" in $\mathbb{R}^{20}$ that fits to data?

Unsupervised learning → find structure in data

$\underline{\text{Ex}}$ 10 students used $\left\{\begin{array}{l} a \text{ time units at lectures} \\ b \text{ time units studying the course at home} \\ c \text{ time units surfing} \end{array}\right.$ to prepare an exam.

The result was d. The data observed was

$$\begin{array}{r} \text{Student } 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 10 \end{array}\begin{pmatrix} 10 & 16 & 5 & 15 \\ 12 & 12 & 15 & 10 \\ 14 & 9 & 12 & 16 \\ 9 & 10 & 14 & 8 \\ 13 & 18 & 10 & 17 \\ 15 & 16 & 11 & 19 \\ 8 & 9 & 17 & 5 \\ 6 & 14 & 17 & 7 \\ 11 & 12 & 11 & 11 \\ 12 & 14 & 8 & 12 \end{pmatrix} = \underset{\substack{\text{data} \\ \text{matrix}}}{X}$$

$$\underbrace{\begin{matrix} a & b & c & d \end{matrix}}_{\text{features}}$$

10 points in $\mathbb{R}^4$. To find patterns in data, find line, 2D-plane, 3D-plane best fitted to data

$\underline{\text{Theorem}}$ (proof on separate page)

The centre point (average) $\bar{m}$ of datapoints belongs to these lines, 2D-planes, ---

$\Rightarrow$ Our new O (origin)!

In the example:

$$\bar{m} = \frac{1}{10}\left[(10,16,5,15) + \dots + (12,14,8,12)\right] = (11,13,12,12)$$

Subtract $\bar{m}$ from each data point to get $\underline{\text{centered data}}$ (average 0 in columns):

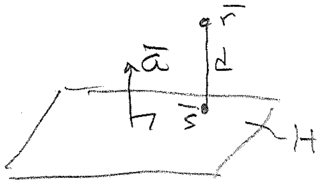
$$Y = \begin{pmatrix} -1 & 3 & -7 & 3 \\ 1 & -1 & 3 & -2 \\ 3 & -4 & 0 & 4 \\ -2 & -3 & 2 & -4 \\ 2 & 5 & -2 & 5 \\ 4 & 3 & -1 & 7 \\ -3 & -4 & 5 & -7 \\ -5 & 1 & 5 & -5 \\ 0 & -1 & -1 & -1 \\ 1 & 1 & -4 & 0 \end{pmatrix}$$

Best line, 2D-plane, -- fitted to these data go through the origin.

$\Rightarrow$ $\underline{\text{subspaces}}$.

Proof, $\bar{m}$ belongs to best line, 2D-plane, ...

Hyperplane in $\mathbb{R}^N$ has dim $N-1$:



$$\underbrace{a_1x_1+\dots+a_nx_n}_{\bar{a}\cdot\bar{x}}+b=0$$

$b\neq 0 \Rightarrow H$ not through $\bar{0} \Rightarrow$ not a subspace of $\mathbb{R}^N$

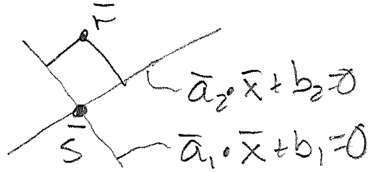
If $|\bar{a}|=1$, $\bar{r}\in\mathbb{R}^N$, $d=$ distance $\bar{r}$ to H, $\bar{r}-\bar{s}=\pm d\cdot\bar{a}$, set H

$$\Rightarrow \pm d\,\bar{a}\cdot\bar{a}=(\bar{r}-\underbrace{\bar{s})\cdot\bar{a}}_{-b}=\bar{r}\cdot\bar{a}+b \Rightarrow d=|\bar{r}\cdot\bar{a}+b|$$

A k-dim plane U is the intersection of $N-k$ hyperplanes,

$$\begin{cases}\bar{a}_1\cdot\bar{x}+b_1=0\\ \vdots\\ \bar{a}_{N-k}\cdot\bar{x}+b_{N-k}=0\end{cases}$$

whose normals $\bar{a}_1,\dots,\bar{a}_{N-k}$ may be assumed orthogonal



U orth. to picture

$$\Rightarrow \text{dist}(\bar{r},U)^2=$$
$$=(\bar{a}_1\cdot\bar{r}+b_1)^2+\dots+(a_{N-k}\cdot\bar{r}+b_{N-k})^2$$

$$\Rightarrow \sum_{j=1}^{N}\text{dist}(\bar{x}_j,U)^2=\sum_{j=1}^{N}\sum_{i=1}^{N-k}(\bar{a}_i\cdot\bar{x}_j+b_i)^2=f(\bar{a}_1,\dots,\bar{a}_{N-k},b_1,\dots,b_{N-k})$$

Find minimum of f with conditions $\bar{a}_i\cdot\bar{a}_j=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$

$\Rightarrow \nabla f$ and gradients of conditions lin. dependent, but conditions do not include $b_1,\dots,b_{N-k} \Rightarrow$

$$\frac{\partial f}{\partial b_i}=0 \text{ for all } i=1,\dots,N-k \Rightarrow$$

$$2\sum_{j=1}^{N}(\bar{a}_i\cdot\bar{x}_j+b_i)=0 \Rightarrow$$

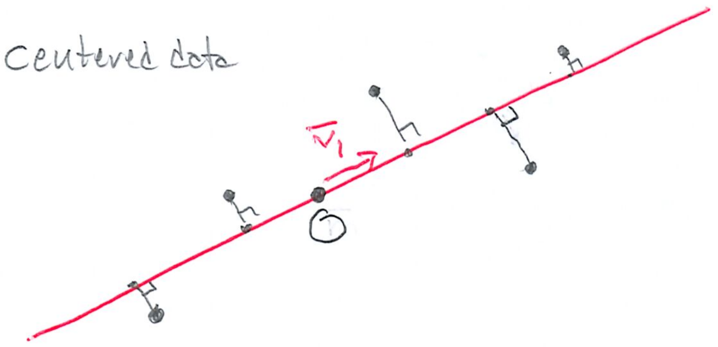
$$\bar{a}_i\cdot(\underbrace{\sum_{j=1}^{N}\bar{x}_j}_{=N\cdot\bar{m}})+Nb_i=0 \Rightarrow$$

$$\Rightarrow \bar{a}_i\cdot\bar{m}+b_i=0,\quad i=1,\dots,N-k$$

$$\Rightarrow \underline{\underline{\bar{m}\in U}}$$

QED

centered data



If all our points are approximated by points on the line, all rows in Y are approximated by vectors parallel to $\bar{v}_1$ $\Rightarrow$ Y is approximated by a rank-1 matrix

rank = dimensions spanned by rows or columns of a matrix

If we approximate with points on a 2D-plane spanned by $\bar{v}_1$ and $\bar{v}_2$, all rows in Y are approximated by linear combinations of $\bar{v}_1$ and $\bar{v}_2$ $\Rightarrow$ Y is approximated by a rank-2 matrix

The sum of row lengths squared in Y is

$$(y_{1,1}^2+\dots+y_{1,4}^2)+\dots+(y_{10,1}^2+\dots+y_{10,4}^2) =$$

$$= \sum_{i=1}^{10}\sum_{j=1}^{4} y_{ij}^2$$

The difference between Y and another matrix $\tilde{Y}$ can be measured by

$$\sum_{i=1}^{10}\sum_{j=1}^{4}(y_{ij}-\tilde{y}_{ij})^2 = \|Y-\tilde{Y}\|^2 \quad \text{(Frobenius norm)}$$

How do we minimize this if we require that $\tilde{Y}$ has rank 1? Or rank 2? Or ...

The solution is given by the Eckart-Young theorem (the most applied mathematical theorem in modern applications?). If your linear algebra course had 2 more weeks, you would have seen it.

A $\underline{\text{any}}$ m×n matrix (real here) $\Rightarrow$ A^TA n×n symmetric : $(A^TA)^T = A^T(A^T)^T = A^TA$

$(A^TA)\bar{v} = \lambda\bar{v} \Rightarrow \lambda\underbrace{\bar{v}^T\bar{v}}_{=|\bar{v}|^2>0} = \bar{v}^TA^TA\bar{v} = (A\bar{v})^T(A\bar{v}) = |A\bar{v}|^2 \geq 0 \Rightarrow \underline{\lambda \geq 0}$ non-neg. eigenvalues

If $\bar{v}_j$ eigenvectors to λ_j (form ON-basis), form columns of a matrix V $(\Rightarrow V^{-1} = V^T)$,

$A^TA = VDV^T$, $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ is the diagonalization of A^TA (spectral theorem in lin.alg.)

Scalar prod: $(A\bar{v}_j)^T(A\bar{v}_k) = \bar{v}_j^TA^TA\bar{v}_k = \lambda_k\bar{v}_j^T\bar{v}_k = \begin{cases} \lambda_k, & j=k \\ 0, & j \neq k \end{cases} \Rightarrow \underbrace{A\bar{v}_1}_{\in\mathbb{R}^m}, \dots, A\bar{v}_n$ orthogonal in $\mathbb{R}^m$ of length $\sqrt{\lambda_n}$, or some $= \bar{0}$

Put $\sigma_k = \sqrt{\lambda_k} \geq 0$ $\underline{\text{singular values of } A}$, and $\bar{w}_k = \frac{1}{\sigma_k}A\bar{v}_k \Rightarrow \bar{w}_1, \dots, \bar{w}_m$ ON basis in $\mathbb{R}^m$ (if $\sigma_1, \dots, \sigma_m$ all $\neq 0$, still works if some $= 0$)

$$\Rightarrow \underbrace{AV}_{m\times n} = \begin{pmatrix} | & & | \\ A\bar{v}_1 & \cdots & A\bar{v}_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \sigma_1\bar{w}_1 & \cdots & \sigma_n\bar{w}_n \\ | & & | \end{pmatrix} = \underbrace{\begin{pmatrix} | & & | \\ \bar{w}_1 & \cdots & \bar{w}_n \\ | & & | \end{pmatrix}}_{\substack{m\times m \\ = W}} \underbrace{\begin{pmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \end{pmatrix}}_{\substack{m\times n \\ = \Sigma}} \qquad \Rightarrow A = W\Sigma V^T$$

In general: $\underline{A = W\Sigma V^T}$ with $\underset{n\times n}{V^{-1} = V^T}$, $\underset{m\times m}{W^{-1} = W^T}$, $\Sigma = \underset{m\times n}{\begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ & & \sigma_r \\ 0 & & & 0 \\ & & & & \ddots \end{pmatrix}}$

the $\underline{\text{singular value decomposition of } A \text{ (SVD)}}$ columns of V = eigenvectors of A^TA

— " — W — " — AA^T

Usually we order eigenvectors s.t. $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots$

r = number of non-zero σ_j:s = rank of A (= rank A^TA, since $A\bar{v} = \bar{0} \Rightarrow A^TA\bar{v} = \bar{0}$ and $A^TA\bar{v} = \bar{0} \Rightarrow \bar{v}^TA^TA\bar{v} = 0 \Rightarrow |A\bar{v}|^2 = 0 \Rightarrow A\bar{v} = 0 \Rightarrow$ nullspaces of A and A^TA equal)

A view of matrix multiplication: $A=\begin{pmatrix}a_1 & c_1\\ a_2 & c_2\end{pmatrix}$, $B=\begin{pmatrix}b_1 & b_2\\ d_1 & d_2\end{pmatrix}$ $\Rightarrow AB=\begin{pmatrix}a_1b_1+c_1d_1 & a_1b_2+c_1d_2\\ a_2b_1+c_2d_1 & a_2b_2+c_2d_2\end{pmatrix}=$

$$=\underbrace{\begin{pmatrix}a_1b_1 & a_1b_2\\ a_2b_1 & a_2b_2\end{pmatrix}}_{\text{rank}=1}+\underbrace{\begin{pmatrix}c_1d_1 & c_1d_2\\ c_2d_1 & c_2d_2\end{pmatrix}}_{\text{rank}=1}=\underset{\substack{\uparrow\\ \text{col 1}\\ \text{of } A}}{\begin{pmatrix}a_1\\ a_2\end{pmatrix}}\underset{\substack{\uparrow\\ \text{row 1}\\ \text{of } B}}{(b_1\ b_2)}+\underset{\substack{\uparrow\\ \text{col 2}\\ \text{of } A}}{\begin{pmatrix}c_1\\ c_2\end{pmatrix}}\underset{\substack{\uparrow\\ \text{row 2}\\ \text{of } B}}{(d_1\ d_2)}$$

similar for larger matrices

rows parallel to (b_1, b_2)

col. parallel to $\begin{pmatrix}a_1\\ a_2\end{pmatrix}$

SVD: $A=\underbrace{\begin{pmatrix}| & & |\\ \sigma_1\bar{w}_1 & \cdots & \sigma_m\bar{w}_n\\ | & & |\end{pmatrix}}_{m\times n}\underbrace{\begin{pmatrix}-\ \bar{v}_1^T\ -\\ \vdots\\ -\ \bar{v}_n^T\ -\end{pmatrix}}_{n\times n}=\underset{\substack{\uparrow\\ \sigma_j>0\\ \text{up to } j=r}}{\sigma_1}\underbrace{\bar{w}_1\bar{v}_1^T}_{\text{rank}=1}+\dots+\sigma_r\underbrace{\bar{w}_r\bar{v}_r^T}_{\text{rank}=1}$

<u>SVD as a sum</u>

useful!

<u>Ex</u> $A=\underset{2\times 3}{\begin{pmatrix}1 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}}$ has $\underset{(A \text{ has rank } 2)}{\sigma_1=\sqrt{3},\ \sigma_2=1}$, $\underbrace{\bar{v}_1=\frac{1}{\sqrt{6}}\begin{pmatrix}1\\ 2\\ 1\end{pmatrix},\ \bar{v}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 0\\ -1\end{pmatrix},\ \bar{v}_3=\frac{1}{\sqrt{3}}\begin{pmatrix}1\\ -1\\ 1\end{pmatrix}}_{\text{ON of } \mathbb{R}^3}$, $\underbrace{\bar{w}_1=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 1\end{pmatrix},\ \bar{w}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -1\end{pmatrix}}_{\text{ON of } \mathbb{R}^2}$

(Exercise: calculate A^TA and find all these)

$$\Rightarrow A=W\Sigma V^T=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}\sqrt{3} & 0 & 0\\ 0 & 1 & 0\end{pmatrix}\begin{pmatrix}1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6}\\ 1/\sqrt{2} & 0 & -1/\sqrt{2}\\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\end{pmatrix}=$$

$$=\underset{\sigma_1}{\sqrt{3}}\underset{\bar{w}_1}{\begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}}\underset{\bar{v}_1^T}{\left(\frac{1}{\sqrt{6}}\ \frac{2}{\sqrt{6}}\ \frac{1}{\sqrt{6}}\right)}+\underset{\sigma_2}{1}\cdot\underset{\bar{w}_2}{\begin{pmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\end{pmatrix}}\underset{\bar{v}_2^T}{\left(\frac{1}{\sqrt{2}}\ 0\ \frac{-1}{\sqrt{2}}\right)}=\underbrace{\begin{pmatrix}1/2 & 1 & 1/2\\ 1/2 & 1 & 1/2\end{pmatrix}}_{\text{rank } 1}+\underbrace{\begin{pmatrix}1/2 & 0 & -1/2\\ -1/2 & 0 & 1/2\end{pmatrix}}_{\text{rank } 1}$$

Eckart-Young theorem (low-rank approximation of a matrix)

The matrix A_s of rank $\leq s$ $(s<r=\text{rank}\,A)$ that minimizes $\|A-A_s\|$, $A=W\Sigma V^T$, is

$$A_s=\sum_{j=1}^{s}\sigma_j \bar{w}_j v_j^T = W\begin{pmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_s & \\ & & & 0 \ddots\end{pmatrix}V^T$$

, truncation of SVD

/Proof on separate page/

Ex The best rank-1 approximation of $A=\begin{pmatrix}1 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}$ is $A_1=\sigma_1\bar{w}_1\bar{v}_1^T=\begin{pmatrix}1/2 & 1 & 1/2\\ 1/2 & 1 & 1/2\end{pmatrix}$

Back to our example (10 points in $\mathbb{R}^4$) $[W, S, V]=\text{svd}(A)$ in Matlab gives all we need.

First columns of W and V ($\bar{w}_1$ and $\bar{v}_1$) and σ_1 from S $\Rightarrow$

Best rank-1 approx Y_1 of Y is $Y_1=\sigma_1\bar{w}_1\bar{v}_1^T \approx 19\cdot\begin{pmatrix}0.34\\ -0.16\\ \vdots\\ -0.03\\ 0.14\end{pmatrix}\underbrace{(0.35, 0.34, -0.51, 0.71)}_{1^{st}\text{ principal component}}$

Each row in Y_1 gives $\approx 19\,(\bar{w}_1)_j\,(0.35, 0.34, -0.51, 0.71)$

positive $1^{st}, 2^{nd}, 4^{th}$ comp. $\Rightarrow$ Attending lectures and studying home good for exam result

↑ exam result

$\rightarrow$ >0 for students 1,3,5,6,10 that did well on exam.

Best rank-2 approx Y_2 of Y is

$$Y_2=\sigma_1\bar{w}_1\bar{v}_1^T+\sigma_2\bar{w}_2\bar{v}_2^T$$

$\sigma_2 \approx 8.5$

$\bar{v}_2 \approx (\underbrace{0.55}_{>0}, \underbrace{-0.58}_{<0}, 0.48, \underbrace{0.35}_{>0})$ 2^{nd} princ. comp.

$\sigma_3\approx 6.9$, $\sigma_4\approx 2.8$ $\Rightarrow$ data points have, in general, smaller coordinates in last $\bar{v}_j$:s

Proof of Eckart-Young $\quad A$ $m\times n$, rank $A = r$

If A_s (rank $\le s$) minimizes $\|A-A_s\|$ and $A = W\Sigma V^T = W\begin{pmatrix}\sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_s & & \\ & & & \ddots & \\ & & & & \sigma_r \\ & & & & & 0 \\ & & & & & & \ddots\end{pmatrix}V^T$, show $A_s = W\begin{pmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_s & \\ & & & 0 \\ & & & & \ddots\end{pmatrix}V^T$

SVD of A_s : $A_s = \hat{W}\hat{\Sigma}\hat{V}^T$, $\hat{\Sigma} = \begin{pmatrix}\hat\sigma_1 & & & \\ & \ddots & & \\ & & \hat\sigma_s & \\ & & & 0 \\ & & & & \ddots\end{pmatrix}$. Put $\hat{\Sigma}_s = \begin{pmatrix}\hat\sigma_1 & & \\ & \ddots & \\ & & \hat\sigma_s\end{pmatrix}$ $(s\times s)$ and $C = \hat{W}^T A\hat{V} \Leftrightarrow A = \hat{W}C\hat{V}^T \Rightarrow$

$\|A-A_s\| = \|\underset{\uparrow}{\hat{W}}(C-\hat{\Sigma})\hat{V}^T\| = \|C-\hat{\Sigma}\|$. Write $C = \begin{pmatrix}C_1 & C_2 \\ C_3 & C_4\end{pmatrix}\begin{matrix}\}s \\ \}m-s\end{matrix}$ blockmatrix $\Rightarrow$

isometries (length-preserving)

$\qquad\qquad\qquad\qquad\qquad\qquad\quad \underbrace{\quad}_{s}\ \underbrace{\quad}_{n-s}$

$\|A-A_s\|^2 = \|C-\hat{\Sigma}\|^2 = \|C_1-\hat{\Sigma}_s\|^2 + \|C_2\|^2 + \|C_3\|^2 + \|C_4\|^2$ ⊛

Now $Y_1 = \hat{W}\left(\begin{array}{c|c}C_1 & C_2 \\ \hline 0 & 0\end{array}\right)\hat{V}^T$ and $Y_2 = \hat{W}\left(\begin{array}{c|c}C_1 & 0 \\ \hline C_3 & 0\end{array}\right)\hat{V}^T$ have rank $\le s \Rightarrow \|A-A_s\|^2 \le \begin{cases}\|A-Y_1\|^2 = \|C_3\|^2 + \|C_4\|^2 \\ \|A-Y_2\|^2 = \|C_2\|^2 + \|C_4\|^2\end{cases}$

$\overset{⊛}{\Rightarrow} C_1 = \hat{\Sigma}_s,\ C_2 = 0,\ C_3 = 0 \Rightarrow \|A-A_s\| = \|C_4\|$ and

$$C = \left(\begin{array}{c|c}\hat{\Sigma}_s & 0 \\ \hline 0 & C_4\end{array}\right) \underset{\substack{\uparrow \\ \text{SVD of } C_4}}{=} \left(\begin{array}{c|c}\hat{\Sigma}_s & 0 \\ \hline 0 & \tilde{W}\tilde{\Sigma}\tilde{V}^T\end{array}\right) = \left(\begin{array}{c|c}I_s & 0 \\ \hline 0 & \tilde{W}\end{array}\right)\left(\begin{array}{c|c}\hat{\Sigma}_s & 0 \\ \hline 0 & \tilde{\Sigma}\end{array}\right)\left(\begin{array}{c|c}I_s & 0 \\ \hline 0 & \tilde{V}^T\end{array}\right) \Rightarrow$$

$$A = \hat{W}C\hat{V}^T = \underbrace{\hat{W}\left(\begin{array}{c|c}I_s & 0 \\ \hline 0 & \tilde{W}\end{array}\right)}_{\substack{\text{orthogonal} \\ = W}}\underbrace{\left(\begin{array}{c|c}\hat{\Sigma}_s & 0 \\ \hline 0 & \hat{\Sigma}\end{array}\right)}_{\substack{\text{"diagonal"} \\ = \Sigma = \begin{pmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_s & \\ & & & \sigma_{s+1} \\ & & & & \ddots\end{pmatrix}}}\underbrace{\left(\begin{array}{c|c}I_s & 0 \\ \hline 0 & \tilde{V}^T\end{array}\right)\hat{V}^T}_{\substack{\text{orthogonal} \\ = V^T}} \Rightarrow$$

$\longleftarrow$ by SVD uniqueness

$$\underline{\underline{A_s}} = \hat{W}\hat{\Sigma}\hat{V}^T = W\left(\begin{array}{c|c}I_s & 0 \\ \hline 0 & \tilde{W}\end{array}\right)^T\left(\begin{array}{c|c}\hat{\Sigma}_s & 0 \\ \hline 0 & 0\end{array}\right)\left(\begin{array}{c|c}I_s & 0 \\ \hline 0 & \tilde{V}\end{array}\right)V^T = \underline{\underline{W\begin{pmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_s & \\ & & & 0 \\ & & & & \ddots\end{pmatrix}V^T}}$$

QED

Remark: In statistics, $\frac{1}{n-1} Y^T Y$ (eigenvectors $\bar{v}_j$ = principal comp.) is called the covariance matrix

c_{ii} are variances of columns of Y, c_{ij} $(i \neq j)$ are covariances between columns

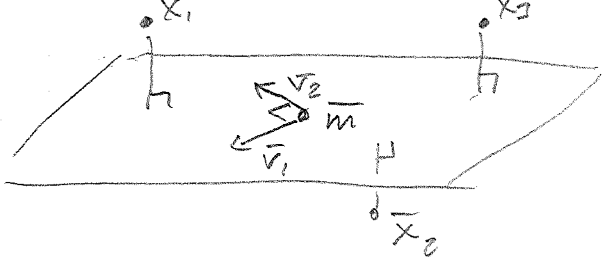
($Y^T Y \sim \begin{pmatrix} + & + & - & + \\ + & + & - & + \\ - & - & + & - \\ + & + & - & + \end{pmatrix}$ in ex.)

In general

For the points $\bar{x}_1, \dots, \bar{x}_N$ in $\mathbb{R}^M$, the closest k-dim "plane" in $\mathbb{R}^M$ is given by

$$\bar{m} + t_1 \bar{v}_1 + \dots + t_k \bar{v}_k \qquad , \quad t_1, \dots, t_k \in \mathbb{R} \text{ (parameters)}$$

where $\bar{m} = \frac{1}{N} \sum_{j=1}^{N} \bar{x}_j$ and $\bar{v}_1, \dots, \bar{v}_k$ are from the SVD of $Y = \begin{pmatrix} - \bar{y}_1 - \\ \vdots \\ - \bar{y}_N - \end{pmatrix}$ with $\bar{y}_j = \bar{x}_j - \bar{m}$

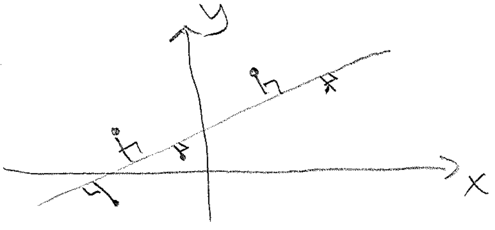


The total least square problem, or

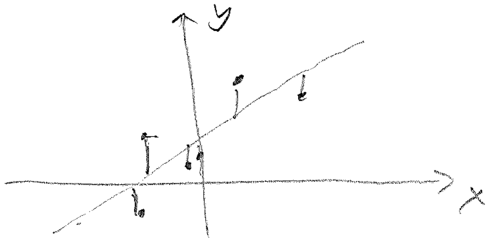
principal component analysis (PCA)

(SVD of centered data)

Note



total least squares



traditional least squares

Other applications of Eckart-Young

$$\underset{mn\text{ elements}}{A} = \sigma_1 \underbrace{\underset{m\text{ elements}}{\bar{w}_1}\,\underset{n\text{ elements}}{\bar{v}_1^T}}_{m+n} + \dots + \sigma_r \bar{w}_r \bar{v}_r^T$$

rank-s approx of A has $s \cdot (m+n)$ elements ⤳ data compression

ex $m=n=1000$ ⇒ A has 1000.000 elements to store

a rank-10 approx $A_{10} = \sigma_1 \bar{w}_1 \bar{v}_1^T + \dots + \sigma_{10} \bar{w}_{10} \bar{v}_{10}^T$ needs $10 \cdot 2000 = 20.000$ elements

98% reduction

common in, e.g., image processing

SVD/PCA is now a main tool in AI/ML (machine learning), in so-called unsupervised learning. Many applications in biology, chemistry, medicine.

(neural networks → supervised learning)