Lesson 24

Read chapter 8

Proof of Eckard Young th

Best fit k spaces

Recall $\| A - A_k \|_2 = \sigma_{k+1}$

Th (Eckard-Young) For any B of rank k
$\| A - B \|_2 \geq \sigma_{k+1}$

Proof: given $A \in R^{m \times n}$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ $B \in R^{m \times n}$ of rank $k < r$ we need to show
$\| A - B \|_2 \geq \sigma_{k+1}$ i.e
$\max_{\|w\|=1} \| (A-B) w \| \geq \sigma_{k+1}$

So we want to find a vector w, $\|w\|=1$
s.t $\| (A-B) w \| = \| Aw - Bw \| \geq \sigma_{k+1}$
choose w such that:
1) $\|w\|=1$ 2) $w \in Null(B)$ a subspace of R^n of dimension $n-k$ 3) $w \in span(v_1 \cdots v_{k+1})$
(The first $k+1$ right singular vectors of A)
a subspace of R^n of dimension $k+1$
Why do I know such a w exists?
Then $w = x_1 v_1 + x_2 v_2 + \cdots + x_{k+1} v_{k+1}$ with
$x_1^2 + x_2^2 + \cdots + x_{k+1}^2 = 1$, $\| (A-B) w \| =$
$\| U \Sigma V^T w \| = \| U \Sigma \begin{pmatrix} x_1 \\ \vdots \\ x_{k+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \| = \sqrt{\sigma_1^2 x_1^2 + \cdots + \sigma_{k+1}^2 x_{k+1}^2}$
$\geq \sqrt{\sigma_{k+1}^2 (x_1^2 + \cdots + x_{k+1})^2} = \sigma_{k+1}$

Example

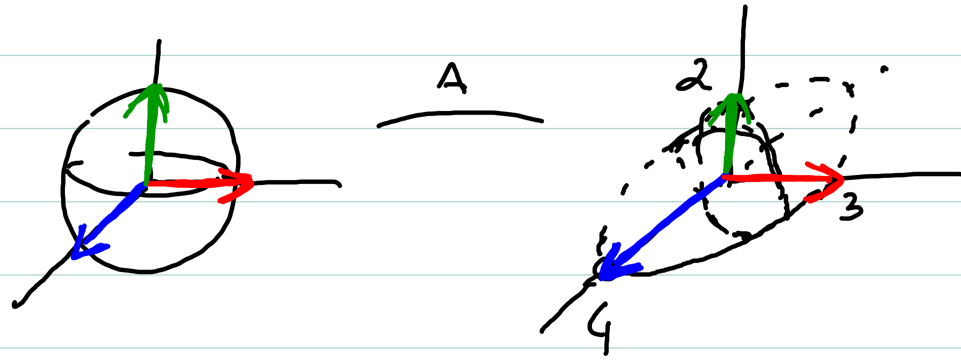
$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} = I \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} I^T =$$

$$= 4 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} [1\ 0\ 0] + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} [0\ 1\ 0] + 2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} [0\ 0\ 1]$$

$$A_1 = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 3.5 & 3.5 & 0 \\ 3.5 & 3.5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We know $\| A - A_1 \|_2 = 3$ so we must have $\| A - B \|_2 \geq 3$

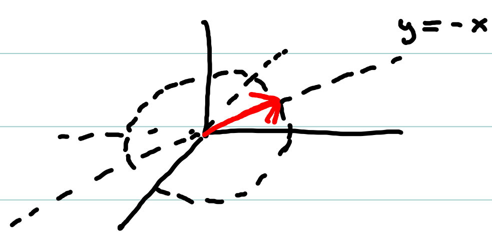


want w, $\|w\|=1$ s.t $\|Aw - Bw\| \geq 3$

Null (B) is The plane $x+y=0$

Span$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\right)$ is the plane $z=0$

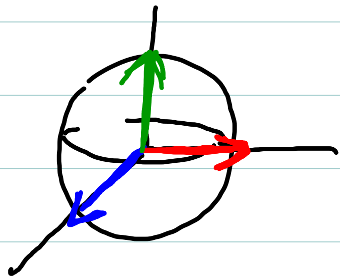
want w on both of these planes



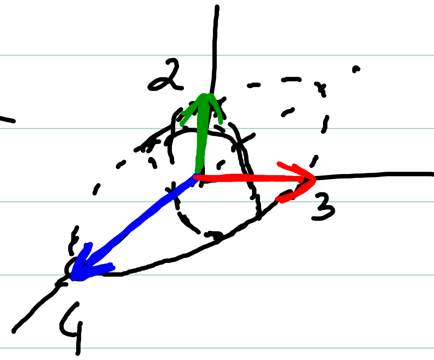
$$w = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}$$

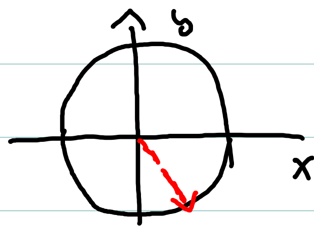
$$\left\| \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 4/\sqrt{2} \\ -3/\sqrt{2} \\ 0 \end{bmatrix} \right\| = \sqrt{\frac{16}{2} + \frac{9}{2}} = \frac{5}{\sqrt{2}}$$

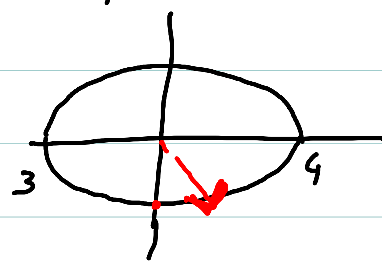
≈ 3.54



$\xrightarrow{A}$







Best fit k-spaces

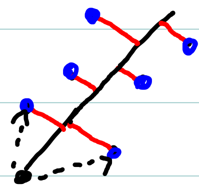
Given $\underset{m\times n}{A} = U\Sigma V^T = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$

Our goal is to show span(v_1) is the line through the origin in $\mathbb{R}^n$ that is closest to the rows of A.

and span(u_1) is the line through the origin in $\mathbb{R}^m$ that is closest to the columns of A.

What does "closest" mean?

Closest line is the line that minimizes the sum of the squares of the (orthogonal) distances of data points (row/columns of A) from line.



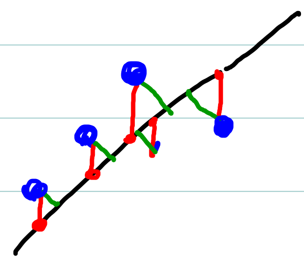
data points
columns/rows of A

Def: The best fit 1-space to the rows of $A = \begin{bmatrix} r_1^T \\ \vdots \\ r_n^T \end{bmatrix}$ is span(v) where

v is a unit vector that minimizes.

$$\sum_{i=1}^{n} \| v - r_i \|^2 .$$

Compare with least squares linear regression:



Least squares line minimizes vertical distances

Best fit line minimizes (orthogonal) distances.

Suppose $\underset{m\times n}{A} = \begin{bmatrix} r_1^T \\ \vdots \\ r_m^T \end{bmatrix}$ $\quad r_i \in \mathbb{R}^n$

Linear regression : $r_i = (a_1^i \dots a_{n-1}^i \ \ b^i)$

$$B = \begin{bmatrix} 1 & a_1^1 & \cdots & a_{n-1}^1 \\ 1 & & & \\ \vdots & & & \\ 1 & a_1^m & & a_{n-1}^m \end{bmatrix}$$

suppose it has rank n

$$B(B^T B)^{-1} B^T \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} d \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}$$

gives us "hyper plane" $y = d + c_1 x_1 + \cdots + c_{n-1} x_{n-1}$

such that

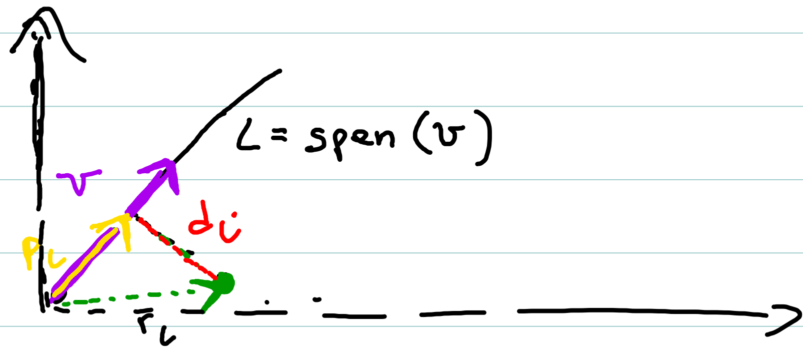
$$\sum_{i=1}^{m} \| d + c_1 a_1^i + c_2 a_2^i + \cdots + c_{n-1} a_{n-1}^i - b^i \|^2$$

is minimum

So the problem of linear regression is to find:

$$\min_{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{n-1} \end{bmatrix} \in \mathbb{R}^n} \sum_{i=1}^{m} \| w_0 + w_1 a_1^i + w_2 a_2^i \cdots + w_{n-1} a_{n-1}^i - b^i \|^2$$

Best fit line. Find



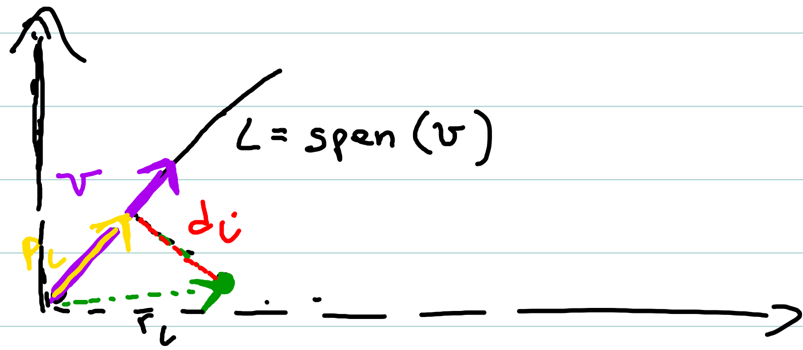
$$\min_{\substack{\|v\|=1 \\ \|v\| \in \mathbb{R}^n}} \sum_{i=1}^{m} \| r_i - p_i \|^2 \qquad p_i = \text{proj}_{\text{span}(v)}(r_i)$$

Note: Span(v) is a line through origin so first we went to center our data.

$$\bar{a}^k = \frac{1}{m} \sum_{i=1}^{m} a^i_k \quad : \qquad \text{Consider}$$

$$B = A - \begin{bmatrix} \bar{a}^1 & \bar{a}^2 & \cdots & \bar{a}^n \\ \bar{a}^1 & \bar{a}^2 & & \vdots \\ \vdots & & & \\ \bar{a}^1 & \bar{a}^2 & & \bar{a}^n \end{bmatrix}$$

$$= \begin{bmatrix} \underbrace{\begin{pmatrix} \; \end{pmatrix} \begin{matrix} -\bar{a}^1 \\ -\bar{a}^1 \\ \vdots \\ -\bar{a}^1 \end{matrix}}_{\text{sums to } 0} & \underbrace{\begin{pmatrix} \; \end{pmatrix} \begin{matrix} -\bar{a}^2 \\ -\bar{a}^2 \\ \\ -\bar{a}^2 \end{matrix}}_{\text{sums to } 0} & \cdots \end{bmatrix}$$



$$\min_{\substack{\|v\|=1 \\ \|v\| \in \mathbb{R}^n}} \sum_{i=1}^{m} \| \overbrace{r_i - p_i}^{d_i} \|^2 \qquad p_i = \text{proj}_{\text{span}(v)}(r_i).$$

Py the goreen th seys $\|r_i\|^2 = \|p_i\|^2 + \|d_i\|^2$

$$\|d_i\|^2 = \|r_i\|^2 - \|p_i\|^2$$

We can minimize $\|d_i\|^2$ by maximizing $\|p_i\|$

want v s.t $\sum_i \|p_i\|^2$ is max

p_i is the projection of r_i on $V = \text{span}(v)$

$p_i = r_i^T \cdot v \cdot v$ $\qquad \|p_i\|^2 = (r_i^T \cdot v)^2$

want $v = \underset{\|v\|=1}{\arg\max} \sum_{i=1}^{n} (r_i^T v)^2 = \|Av\|^2$

since $\quad Av = \begin{bmatrix} r_1^T \\ r_2^T \\ \vdots \\ r_m^T \end{bmatrix} v = \begin{bmatrix} r_1^T \cdot v \\ r_2^T v \\ \vdots \\ r_m^T v \end{bmatrix}$

So we are looking for a unit vector s.t $\|Av\|$ is max

Recall $\|A\|_2 = \max_{\|v\|=1} \|Av\| = \sigma_1$, where

σ_1 is the first singular value of A, and

$\|Av_1\| = \sigma_1$ where v_1 is the

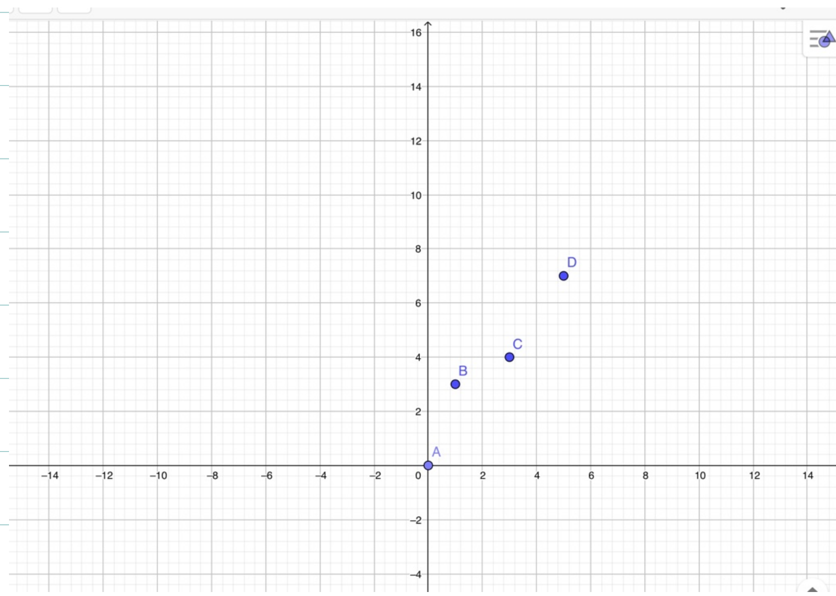
first right singular value of A.

So $\boxed{v = v_1}$ and $\sum_{i=1}^{n} \|p_i\|^2 = \sigma_1$

Ex: In hw 5 you found best fit linear regression line to

$(0,0)\ (1,3)\ (3,4)\ (5,7)$: $y = 0.68 + 1.25x$

Best fit line? we find a line through origin



Note: maybe no line through origin is very good fit to data, if data is not centered around origin, so we may want to shift our data first.

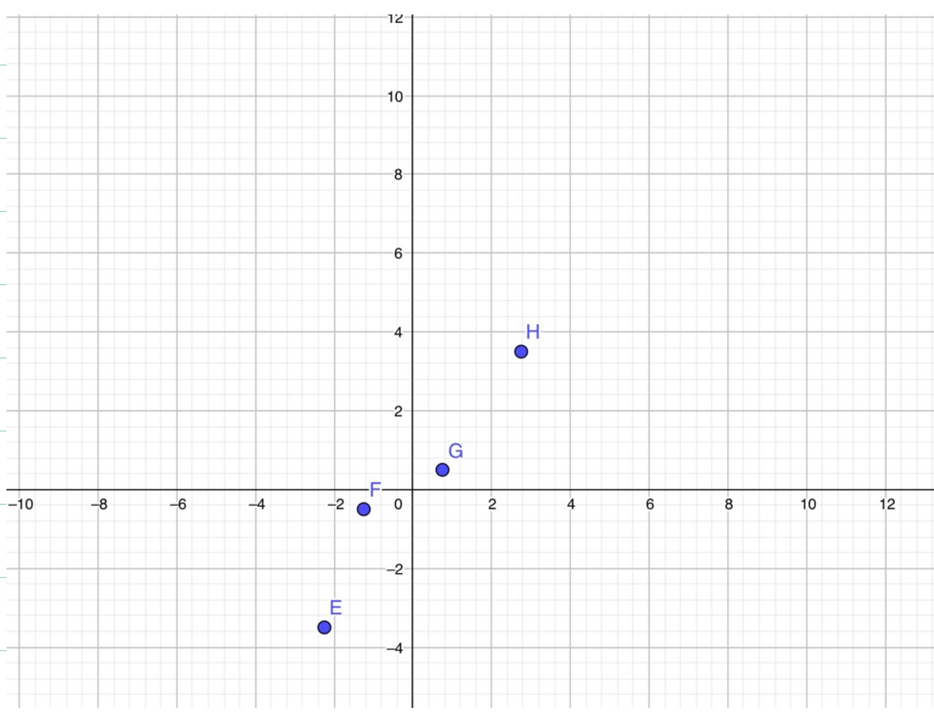
Data is in rows : subtract mean of each column to center around origin.

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 3 \\ 3 & 4 \\ 5 & 7 \end{bmatrix}$$

$$x \text{ mean} = \frac{0+1+3+5}{4} = 9/4$$

$$y \text{ mean} = \frac{0+3+4+7}{4} = \frac{14}{4} = \frac{7}{2}$$

$$B = \begin{bmatrix} -9/4 & -7/2 \\ 1-9/4 & 3-7/2 \\ 3-9/4 & 4-7/2 \\ 5-9/4 & 7-7/2 \end{bmatrix} = \begin{array}{c} \begin{bmatrix} -9/4 & -7/2 \\ -5/4 & -1/2 \\ 3/4 & 1/2 \\ 11/4 & 7/2 \end{bmatrix} \end{array} \begin{array}{c} E \\ F \\ G \\ H \end{array}$$

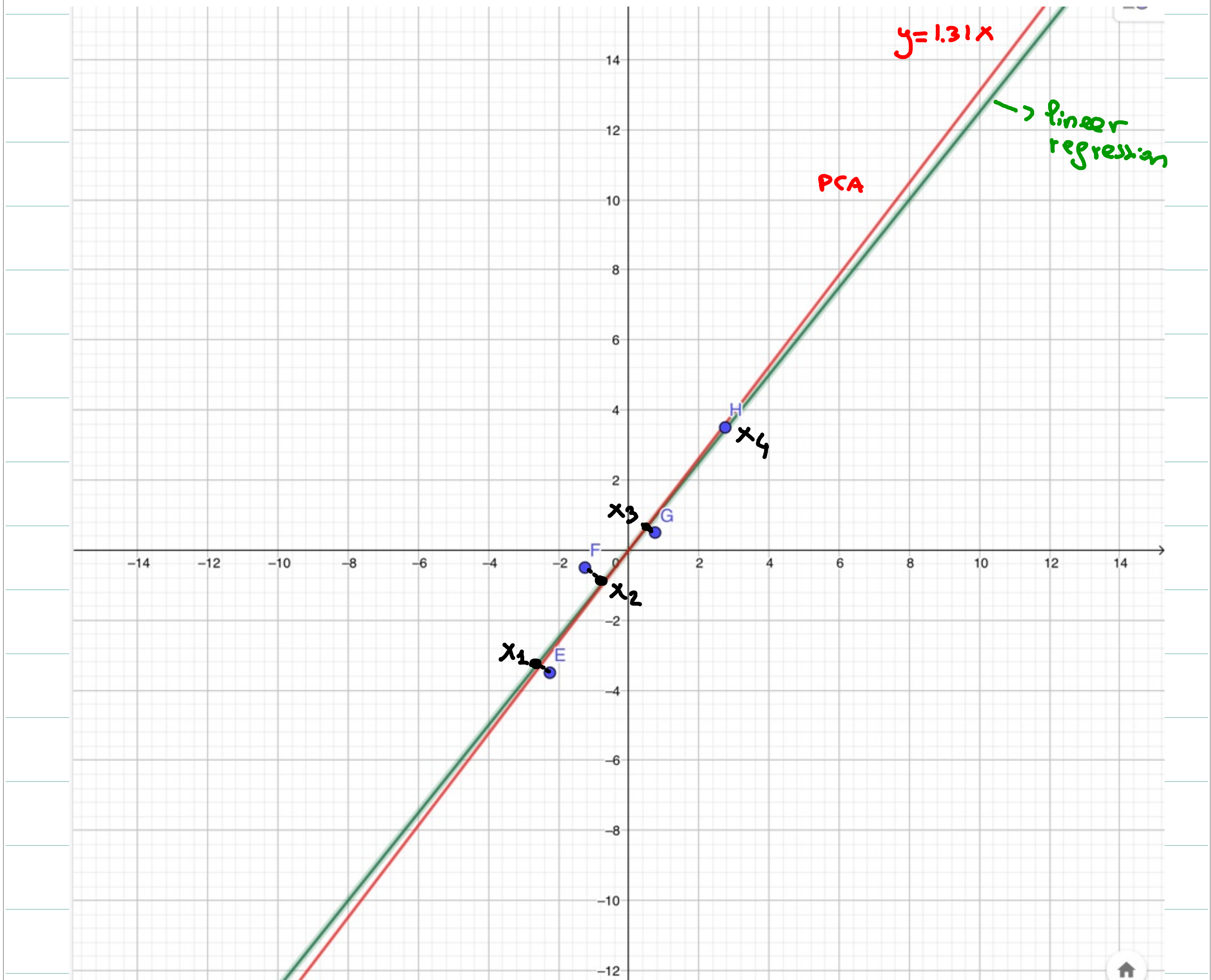


```
 [1] top-level scope
   @ none:1

julia> using LinearAlgebra

julia> a=[-9/4 -7/2; -5/4 -1/2; 3/4 1/2; 11/4 7/2]
4×2 Matrix{Float64}:
 -2.25  -3.5
 -1.25  -0.5
  0.75   0.5
  2.75   3.5

julia> svd(a)
SVD{Float64, Float64, Matrix{Float64}}
U factor:
4×2 Matrix{Float64}:
 -0.663581   0.398393
 -0.184732  -0.8405
  0.136306   0.357277
  0.712008   0.0848297
singular values:
2-element Vector{Float64}:
 6.25074081965229
 0.8235527946237668
Vt factor:
2×2 Matrix{Float64}:
 0.605404   0.795918
 0.795918  -0.605404

julia>
```

* v_1

$$\begin{bmatrix} x \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} 0.605 \\ 0.796 \end{bmatrix}}_{v_1} \qquad y = \frac{0.796}{0.605} x \qquad y = 1.31x$$

Best fit line is line through origin in the direction of v_1.



$y + \frac{7}{2} = 0.68 + 1.25(x + 9/4)$ Linear regression line

$y = 1.25x$

$y = 1.31x$ Best fit line