ČVUT
v Praze
in Prague

Least squares

Václav Hlaváč
Czech Technical University in Prague (ČVUT)
Czech Institute of Informatics, Robotics,
and Cybernetics (CIIRC)
Prague 6, Jugoslávských partyzánů 1580/3
Czech Republic
vaclav.hlavac@cvut.cz
http://people.ciirc.cvut.cz/hlavac/

m p

Courtesy: Fred Pighin and J.P. Lewis, SIGGRAPH 2007 Course

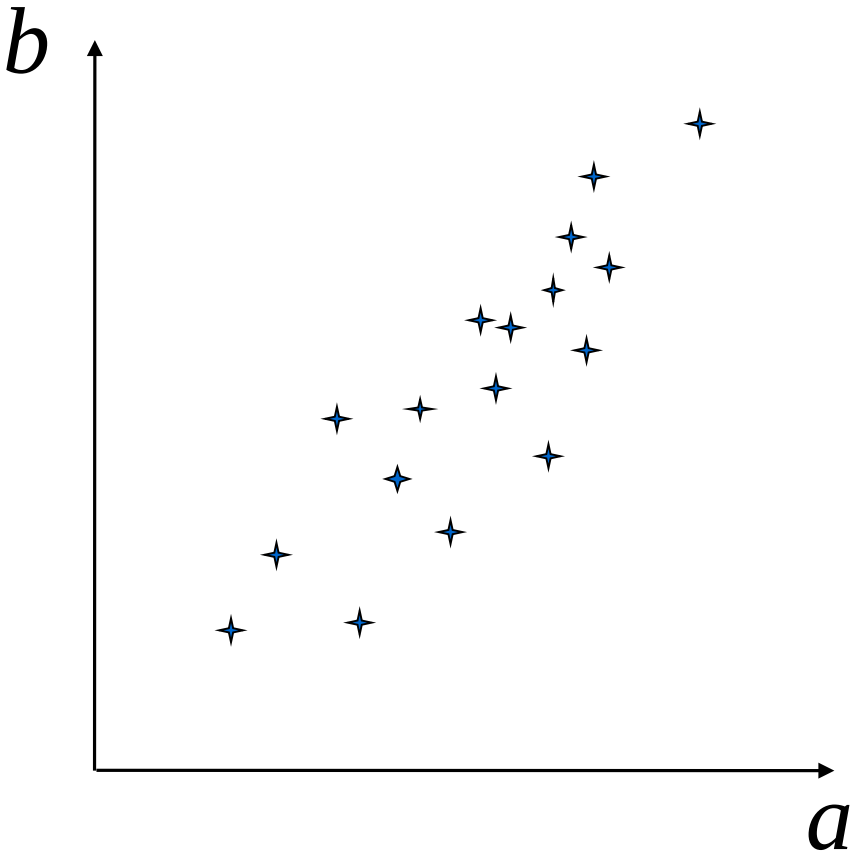
Outline

ČVUT
v Praze
in Prague

- Linear regression
- Geometry of least-squares
- Discussion of the Gauss-Markov theorem

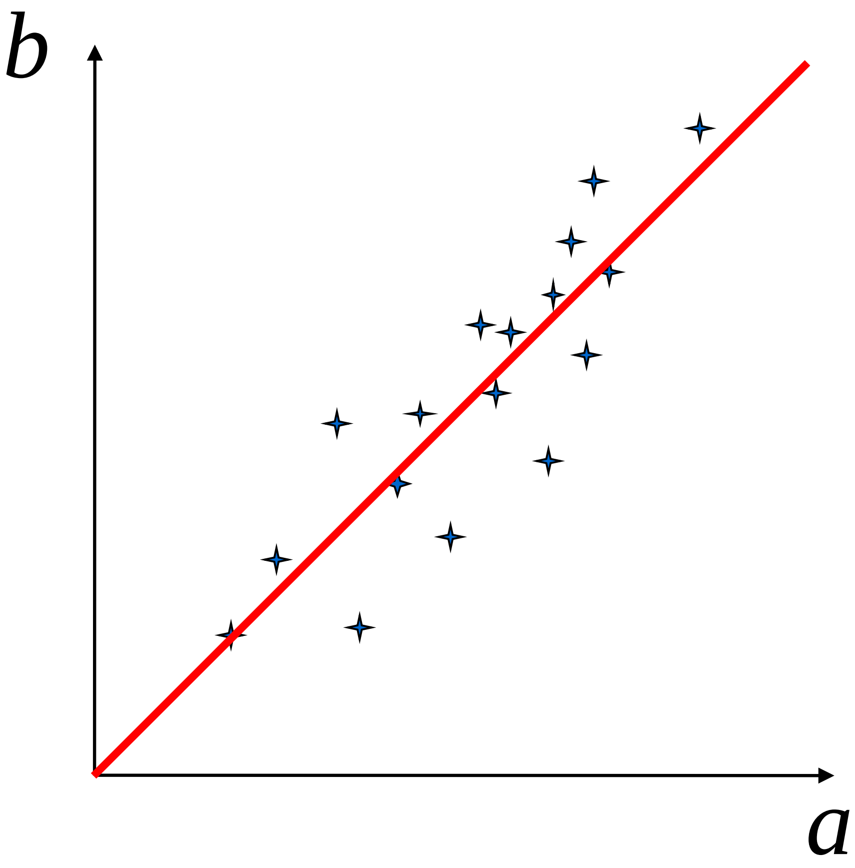
One-dimensional regression

ČVUT
v Praze
in Prague



One-dimensional regression

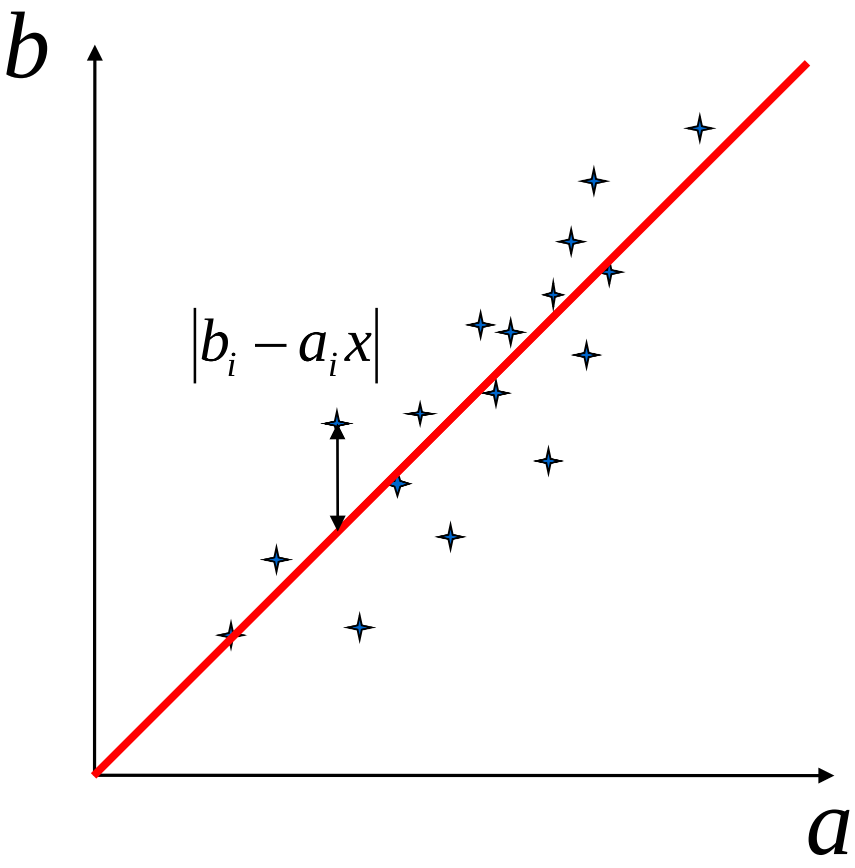
ČVUT
v Praze
in Prague



Find a line that represent the ”best” linear relationship:

$$b = ax$$

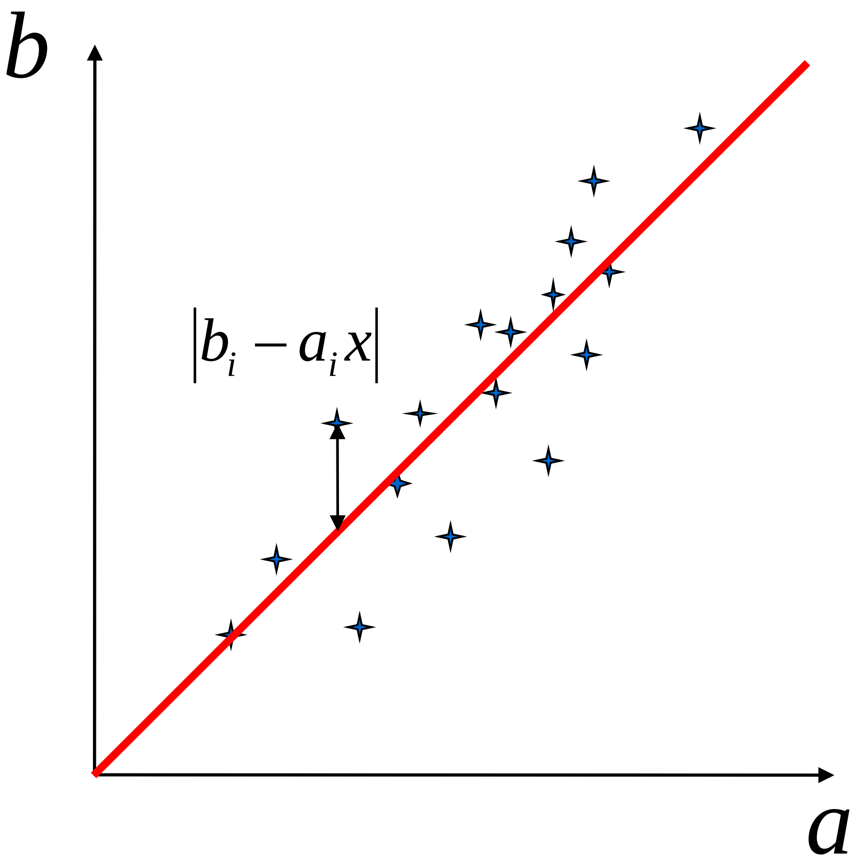
One-dimensional regression



- Problem: the data does not go through a line

$$e_i = b_i - a_i x$$

One-dimensional regression



- Problem: the data does not go through a line

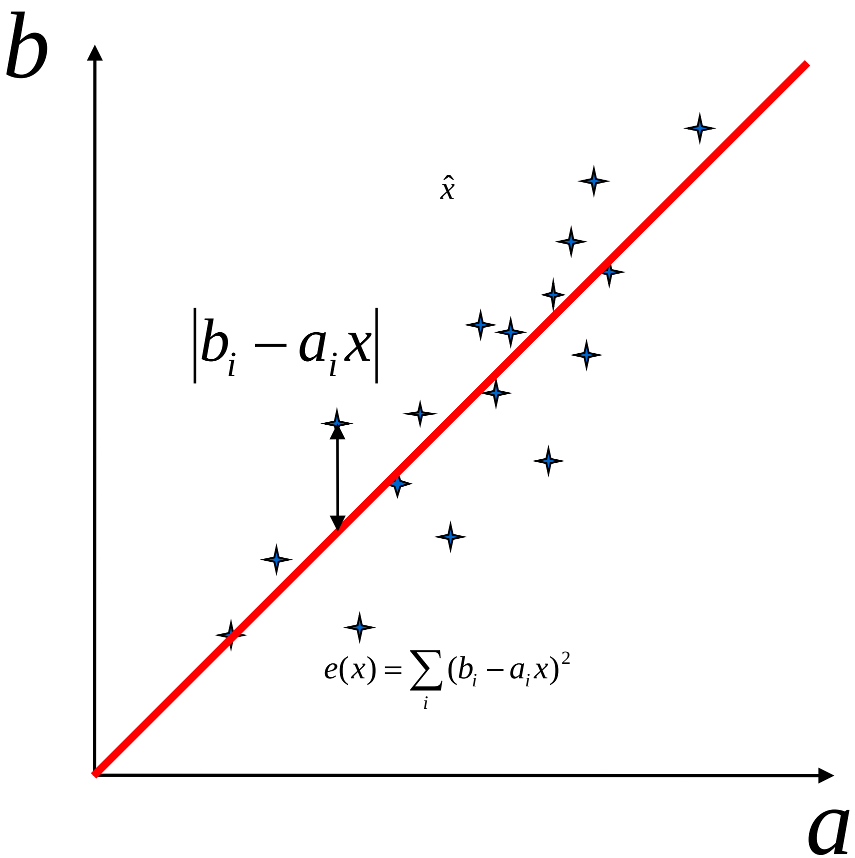
$$e_i = b_i - a_i x$$

- Find the line that minimizes the sum:

$$\sum_i (b_i - a_i x)^2$$

One-dimensional regression

ČVUT
v Praze
in Prague



- Problem: the data does not go through a line

$$e_i = b_i - a_i x$$

- Find the line that minimizes the sum:

$$\sum_i (b_i - a_i x)^2$$

- We are looking for that minimizes

Least squares example

ČVUT
v Praze
in Prague

There are 3 mountains u, y, z that from one site have been measured as 2474 m, 3882 m, and 4834 m. But from u, y looks 1422 m taller and the z looks 2354 m taller, and from y, z looks 950 m taller. Set up the overdetermined system.

$$Ax = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} u \\ y \\ z \end{pmatrix} \sim \begin{pmatrix} 2474 \\ 3882 \\ 4834 \\ 1422 \\ 2354 \\ 950 \end{pmatrix} = b$$

Want to minimize $\|Ax-b\|_2$

Approaches to solve $Ax \approx b$

- Normal equations-quick and dirty
- QR- standard in libraries uses orthogonal decomposition
- SVD - decomposition which also gives indication how linear independent columns are
- Conjugate gradient - no decompositions, good for large sparse problems

Matrix notation

ČVUT
v Praze
in Prague

Using the following notations

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

we can rewrite the error function using linear algebra as:

$$e(x) = \sum_i (b_i - a_i x)^2$$

$$= (\boldsymbol{b} - x\boldsymbol{a})^T (\boldsymbol{b} - x\boldsymbol{a})$$

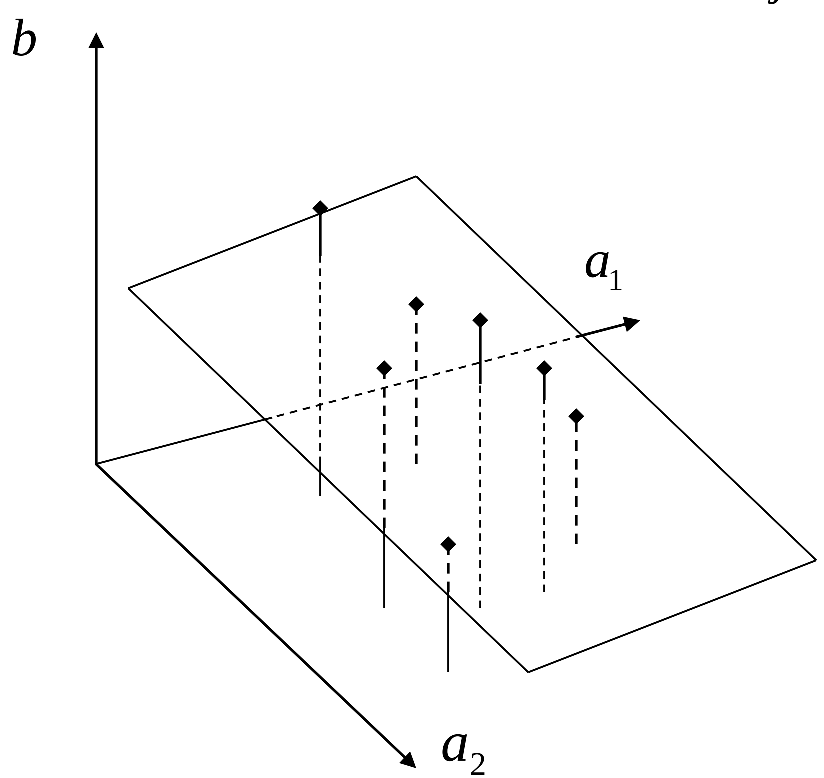
$$\boxed{e(x) = \|\boldsymbol{b} - x\boldsymbol{a}\|^2}$$

Multidimentional linear regression

ČVUT
v Praze
in Prague

Using a model with *m* parameters

$$b = a_1 x_1 + \ldots + a_m x_m = \sum_j a_j x_j$$



Multidimentional linear regression

ČVUT
v Praze
in Prague

Using a model with *m* parameters

$$b = a_1 x_1 + \ldots + a_m x_m = \sum_j a_j x_j$$

and *n* measurements

$$e(\boldsymbol{x}) = \sum_{i=1}^{n} \left(b_i - \sum_{j=1}^{m} a_{i,j} x_j\right)^2$$

$$= \left\| \boldsymbol{b} - \left[\sum_{j=1}^{m} a_{i,j} x_j \right] \right\|^2$$

$$e(\boldsymbol{x}) = \left\| \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \right\|^2$$

Matrix representation

ČVUT
v Praze
in Prague

parameter 1

$$\boldsymbol{b} - \boldsymbol{Ax} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} - \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

measurement *n*

$$= \begin{bmatrix} b_1 - (a_{1,1}x_1 + \ldots + a_{1,m}x_m) \\ \vdots \\ b_n - (a_{n,1}x_1 + \ldots + a_{n,m}x_m) \end{bmatrix}$$

Minimizing $e(\boldsymbol{x})$

ČVUT
v Praze
in Prague

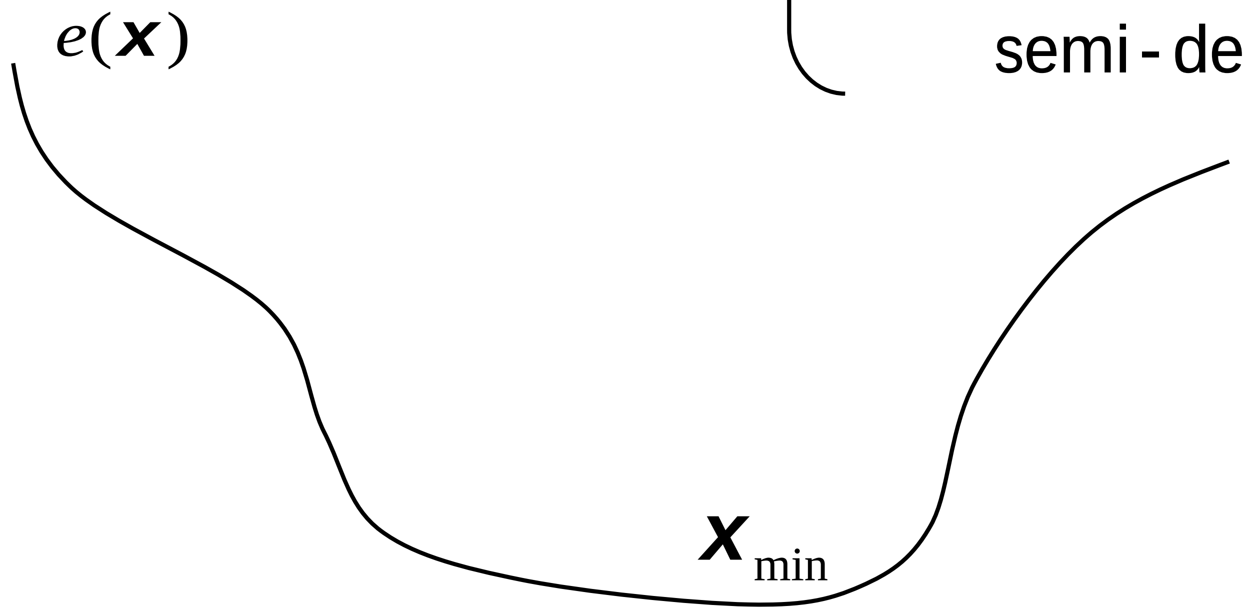
$\boldsymbol{x}_{\min}$ minimizes $e(\boldsymbol{x})$ if

- $e(\boldsymbol{x})$ is flat at $\boldsymbol{x}_{\min}$

  $$\nabla e(\boldsymbol{x}_{\min}) = \boldsymbol{0}$$

- $e(\boldsymbol{x})$ does not go down around $\boldsymbol{x}_{\min}$

  $H_e(\boldsymbol{x}_{\min})$ is positive semi-definite



Positive semi-definite

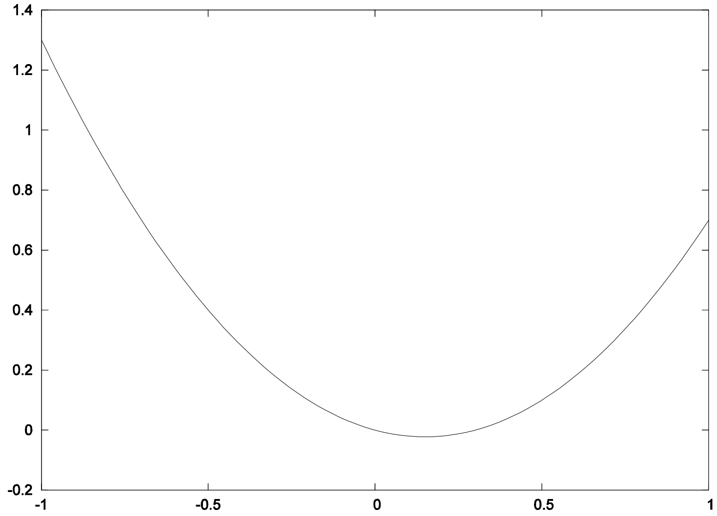
ČVUT
v Praze
in Prague

$$\boldsymbol{A} \text{ is positive semi-definite}$$

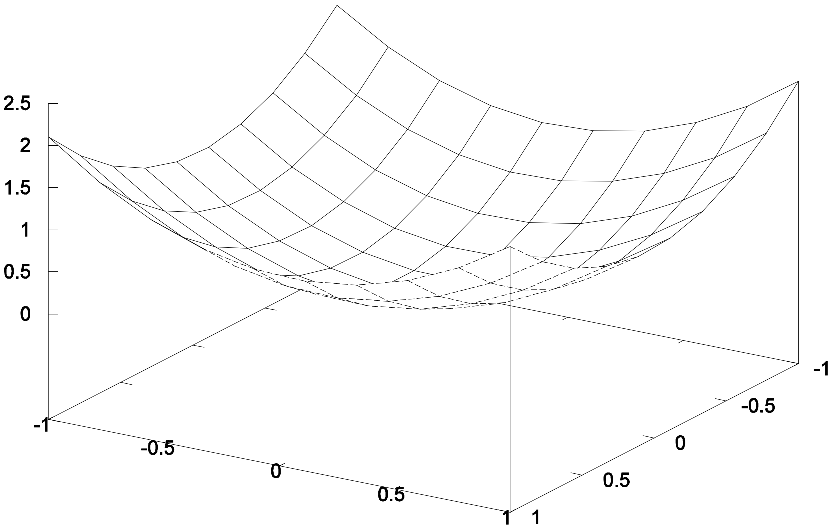
$$\Leftrightarrow$$

$$\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{0}, \text{ for all } \boldsymbol{x}$$

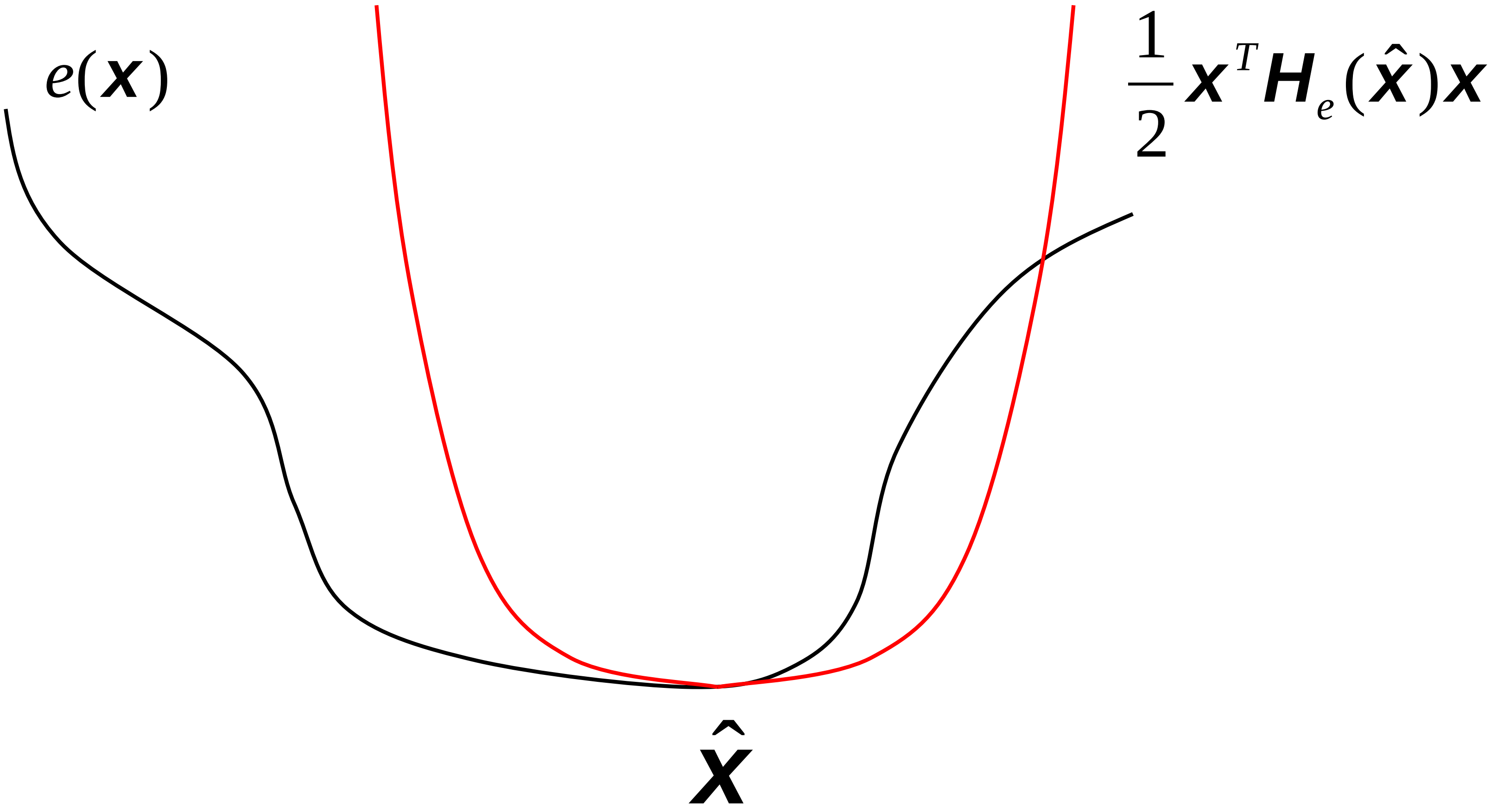
In 1-D



In 2-D

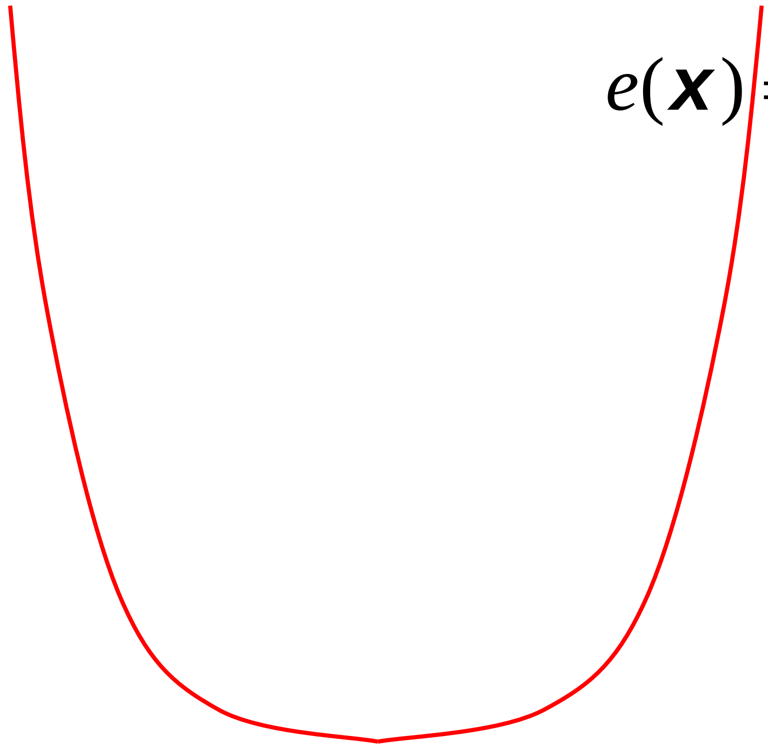


Minimizing $e(\boldsymbol{x})$



Minimizing $e(\boldsymbol{x}) = \left\| \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \right\|^2$

$$e(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{H}_e(\hat{\boldsymbol{x}})\boldsymbol{x}$$



$\hat{\boldsymbol{x}}$

Minimizing $e(\boldsymbol{x}) = \left\| \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \right\|^2$

ČVUT
v Praze
in Prague

$\hat{\boldsymbol{x}}$ minimizes $e(\boldsymbol{x})$ if

$$\boldsymbol{A}^T \boldsymbol{A} \hat{\boldsymbol{x}} = \boldsymbol{A}^T \boldsymbol{b}$$

The ***normal equation***

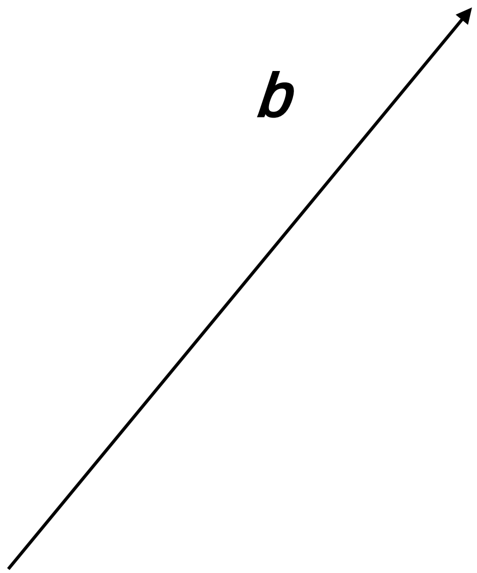
$2\boldsymbol{A}^T \boldsymbol{A}$ is positive semi-definite

Always true

Geometric interpretation

ČVUT
v Praze
in Prague

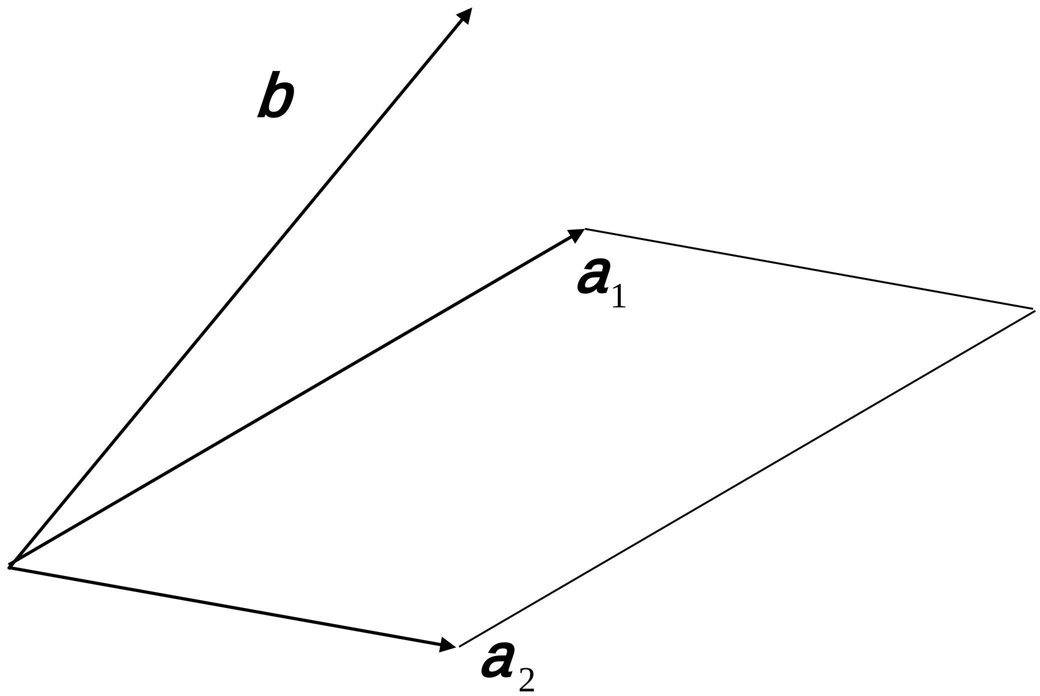
- $\boldsymbol{b}$ is a vector in R^n



Geometric interpretation

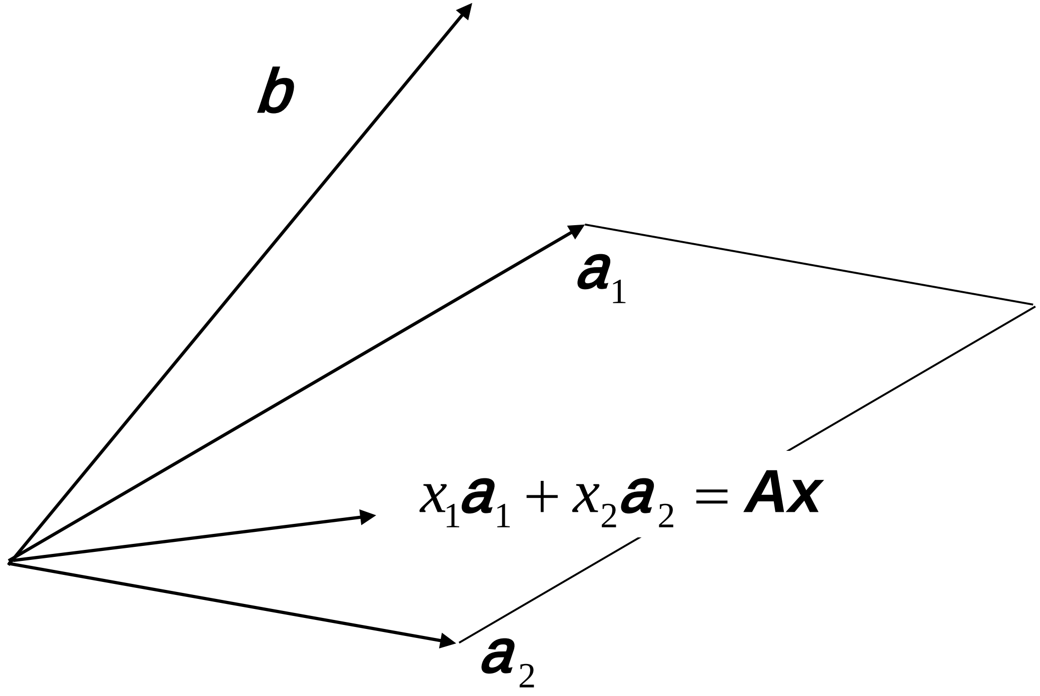
ČVUT
v Praze
in Prague

- ***b*** is a vector in R^n
- The columns of ***A*** define a vector space *range(**A**)*



Geometric interpretation

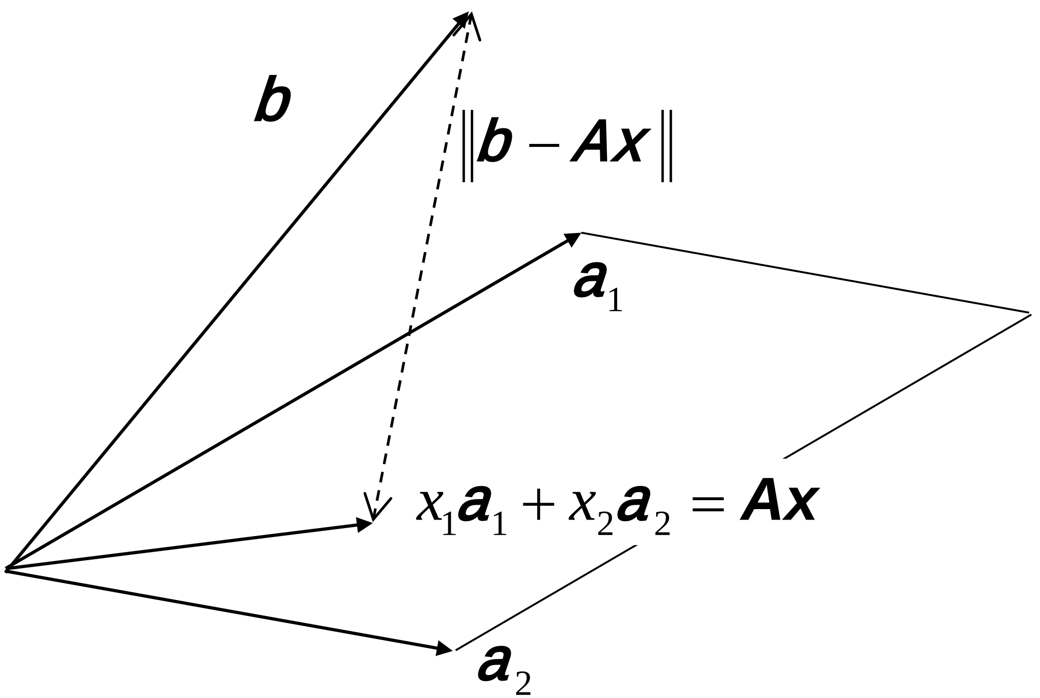
- ***b*** is a vector in R^n
- The columns of ***A*** define a vector space *range(**A**)*
- ***Ax*** is an arbitrary vector in *range(**A**)*



Geometric interpretation

ČVUT
v Praze
in Prague

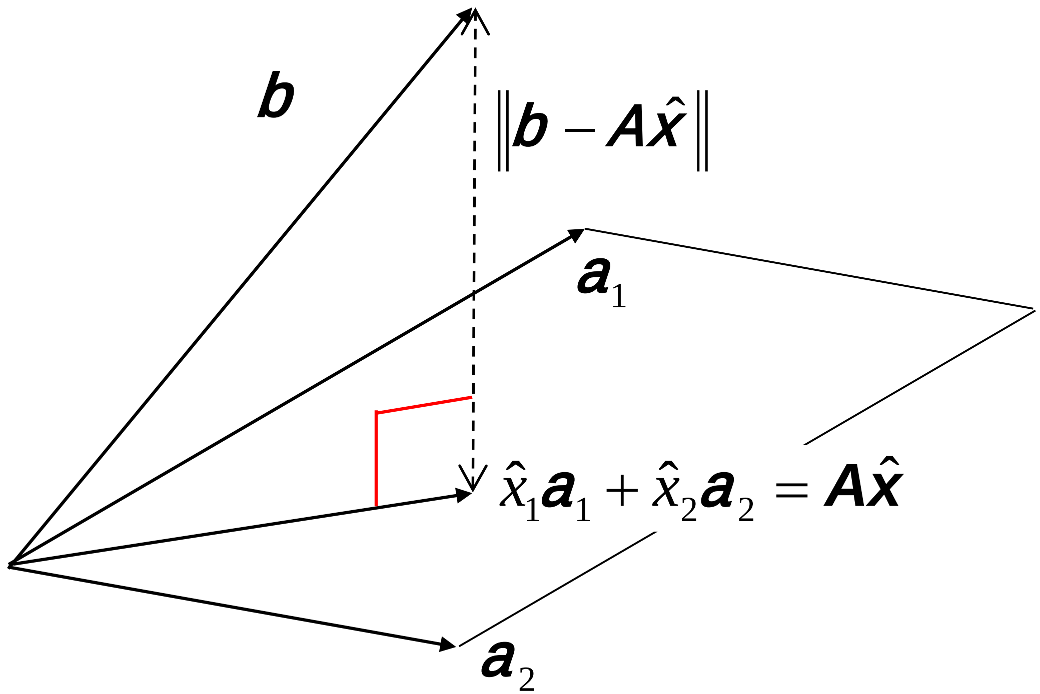
- ***b*** is a vector in R^n
- The columns of ***A*** define a vector space *range(**A**)*
- ***Ax*** is an arbitrary vector in *range(**A**)*



Geometric interpretation

- $\boldsymbol{A\hat{x}}$ is the orthogonal projection of $\boldsymbol{b}$ onto *range(**A**)*

$$\Leftrightarrow \boldsymbol{A}^T\left(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}\right) = \boldsymbol{0} \Leftrightarrow \boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$$



The normal equation: $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$

The normal equation: $\boldsymbol{A}^T \boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T \boldsymbol{b}$

ČVUT
v Praze
in Prague

- Existence: has always a solution

$$\boldsymbol{A}^T \boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T \boldsymbol{b}$$

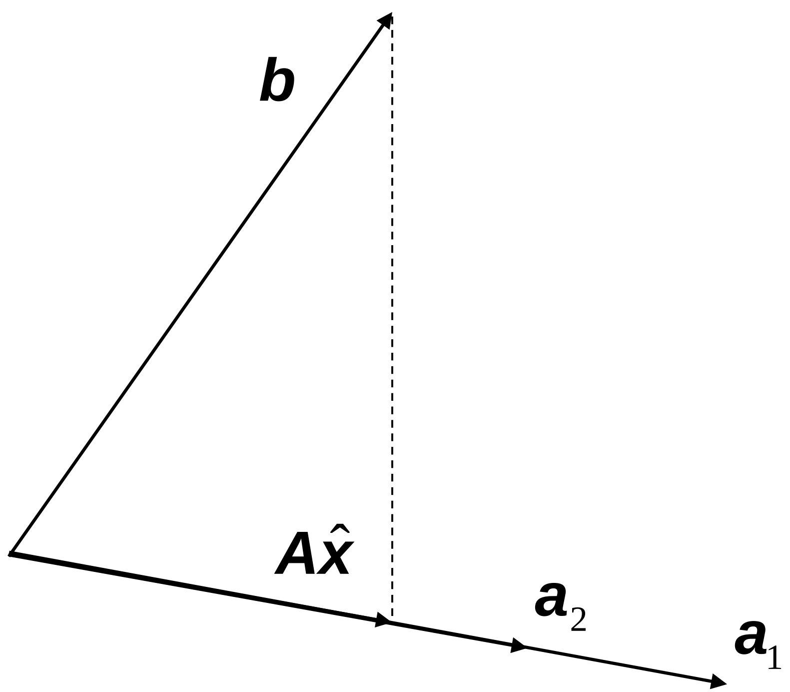
The normal equation: $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$

- Existence: $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$ has always a solution
- Uniqueness: the solution is unique if the columns of $\boldsymbol{A}$ are linearly independent

The normal equation: $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$

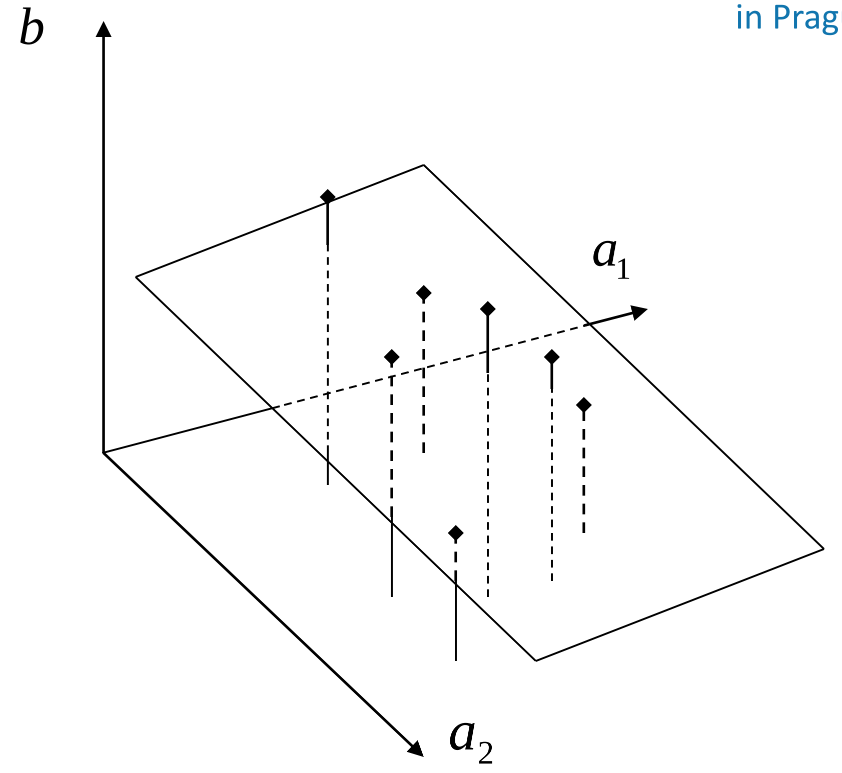
ČVUT
v Praze
in Prague

- Existence: $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$ has always a solution
- Uniqueness: the solution is unique if the columns of $\boldsymbol{A}$ are linearly independent

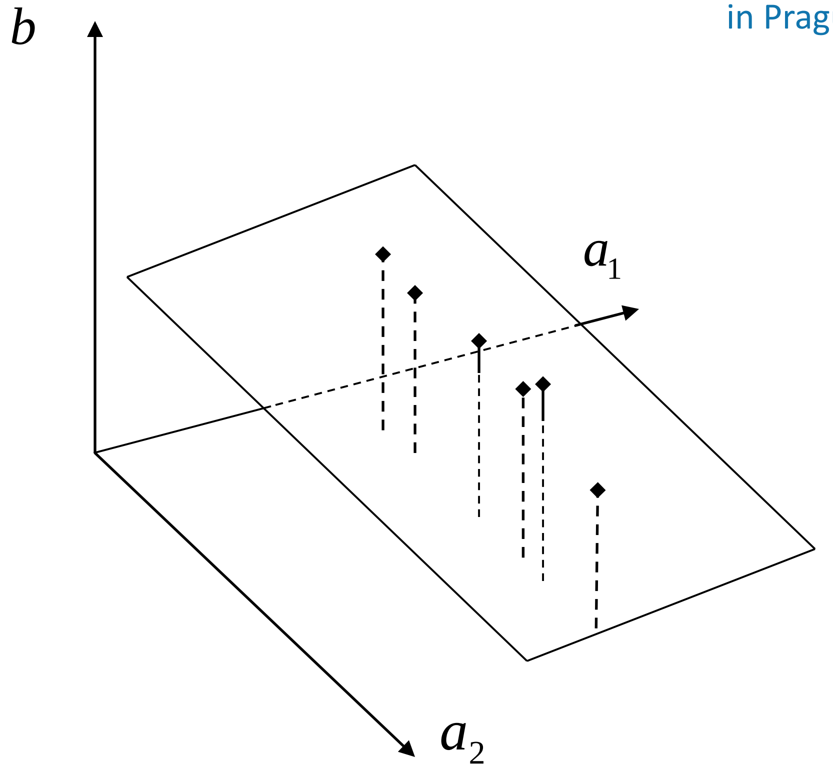


Under-constrained problem

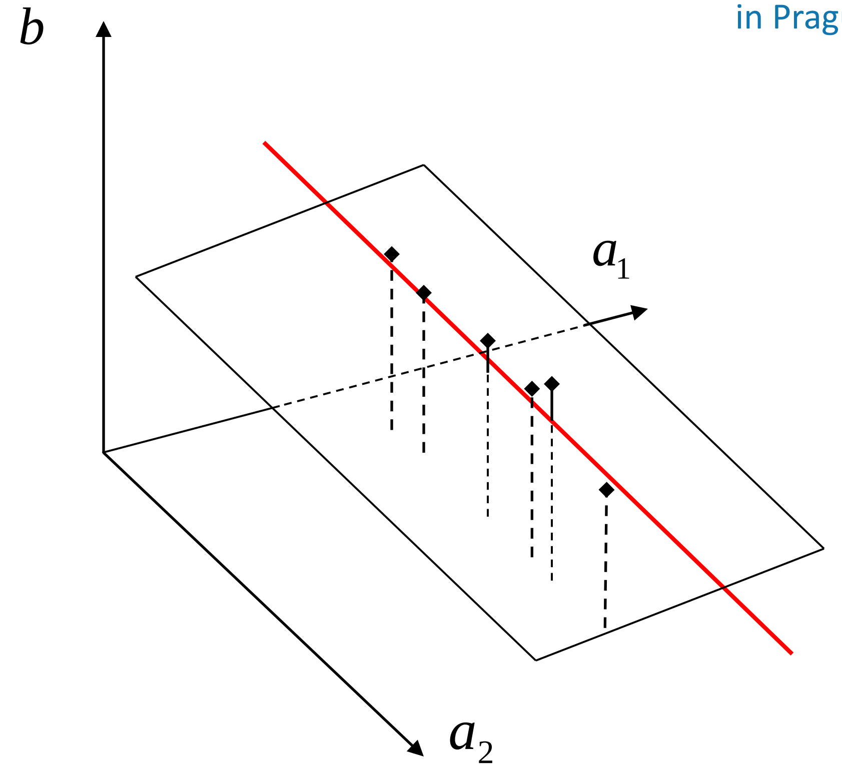
ČVUT
v Praze
in Prague



Under-constrained problem



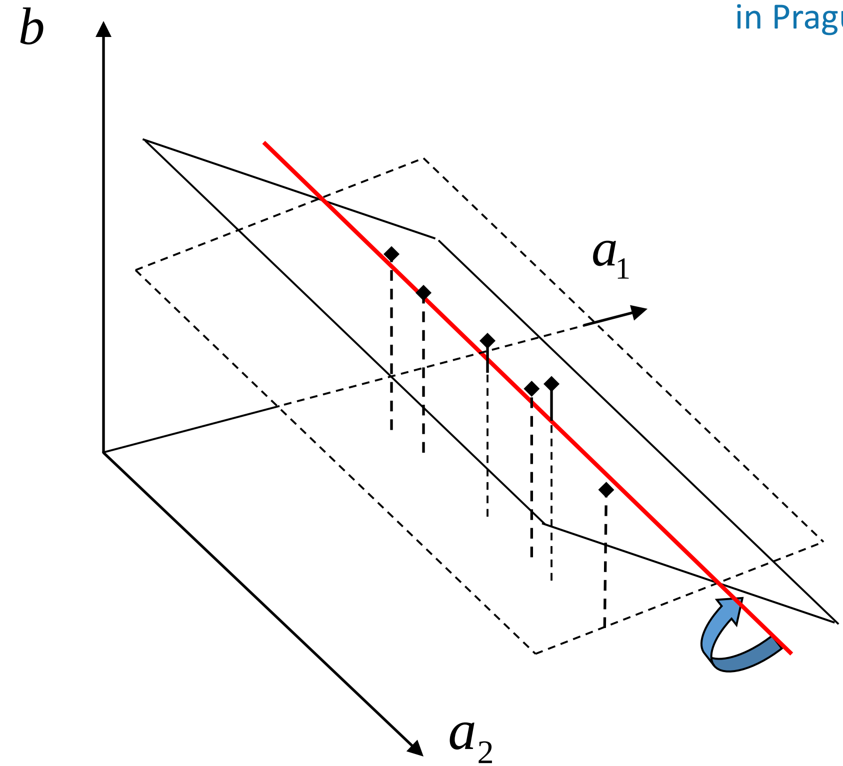
Under-constrained problem



Under-constrained problem

ČVUT
v Praze
in Prague

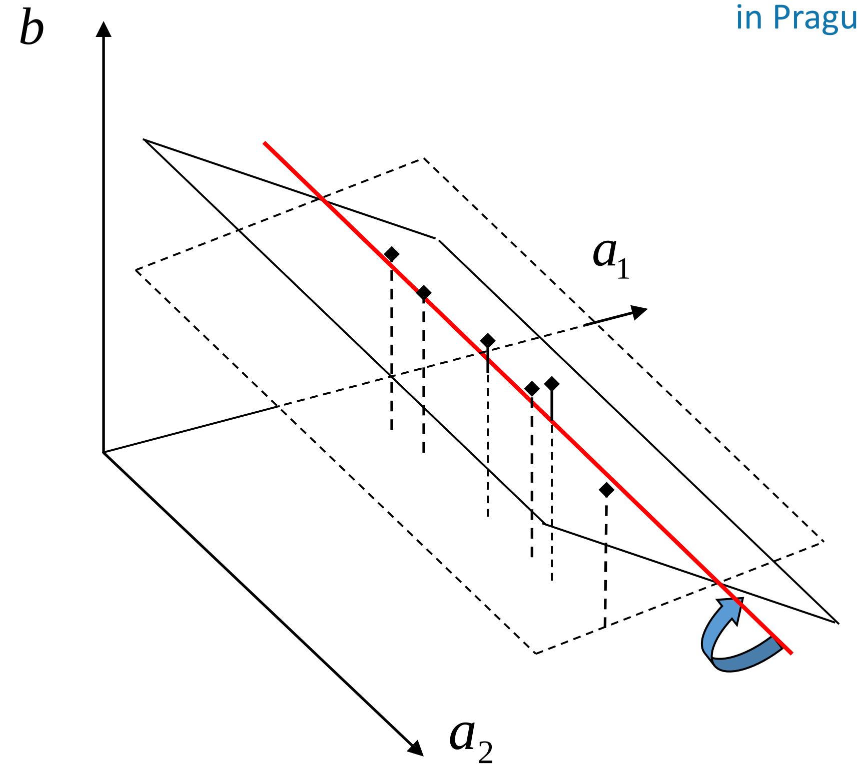
- Poorly selected data
- One or more of the parameters are redundant



Under-constrained problem

ČVUT
v Praze
in Prague

- Poorly selected data
- One or more of the parameters are redundant



Add constraints

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b} \text{ with } \min_{\boldsymbol{x}} \|\boldsymbol{x}\|$$

How good is the least-squares?

ČVUT
v Praze
in Prague

- Optimality: the Gauss-Markov theorem

  Let $\{b_i\}$ and $\{x_j\}$ be two sets of random variables

  and define: $$e_i = b_i - a_{i,1}x_1 - \ldots - a_{i,m}x_m$$

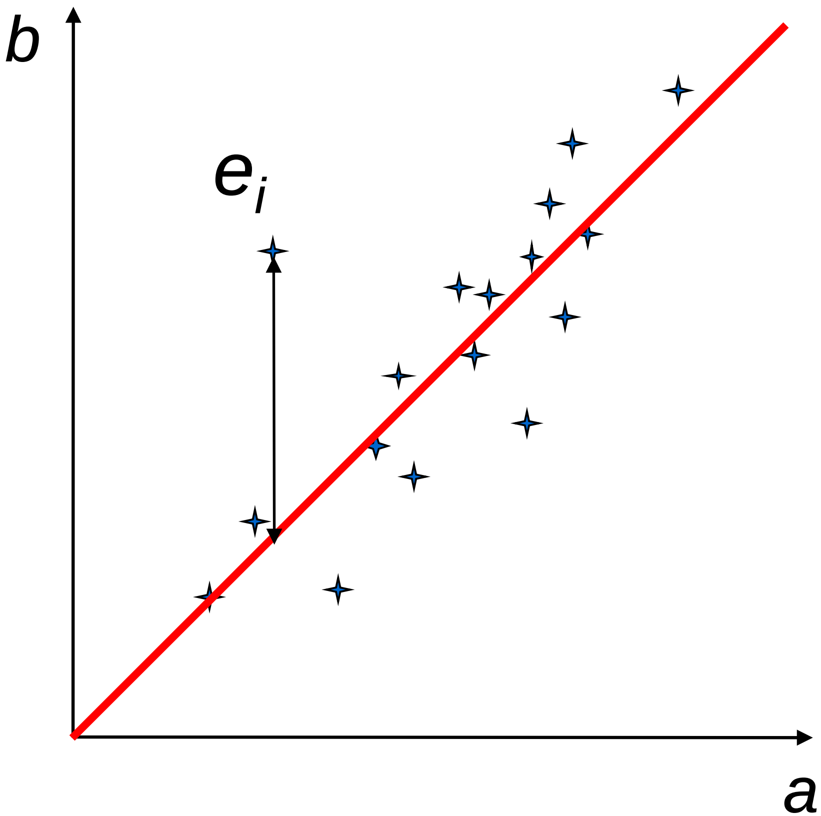
  If
  A1 : $\{a_{i,j}\}$ are not random variables,

  A2 : $E(e_i) = 0$, for all i,

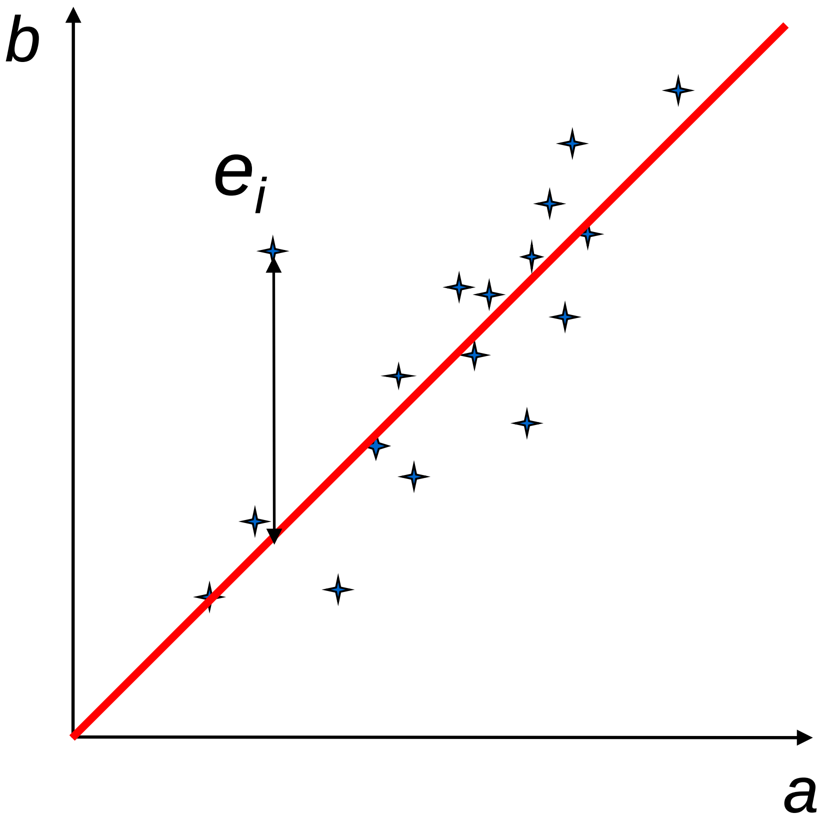
  A3 : $\mathrm{var}(e_i) = \sigma$, for all i,

  A4 : $\mathrm{cov}(e_i, e_j) = 0$, for all i and j,

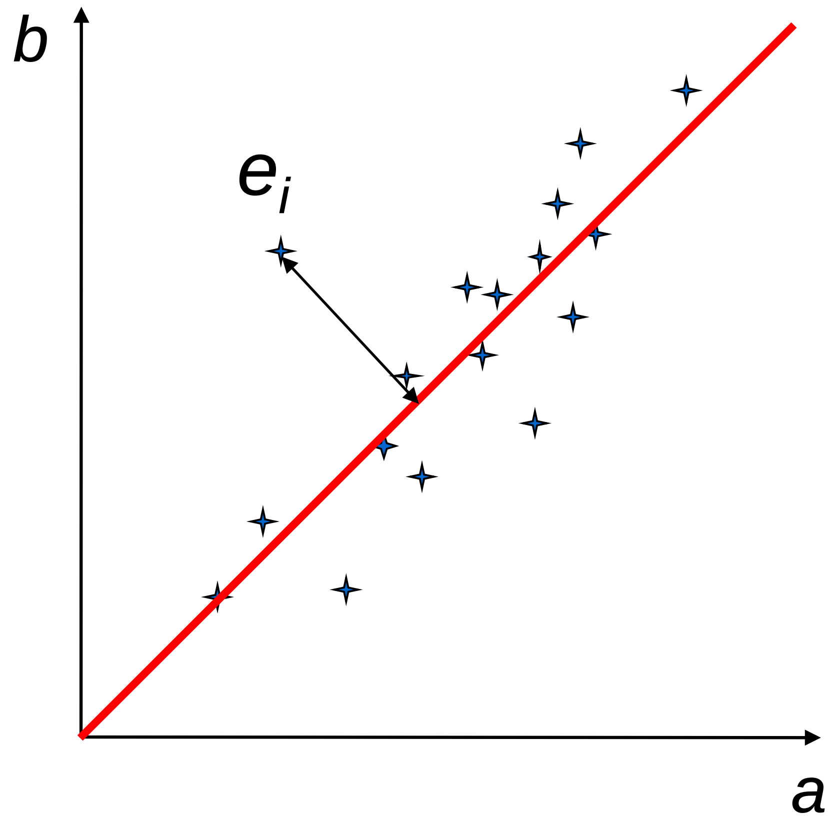
  Then $$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum e_i^2$$ is the best unbiased linear estimator



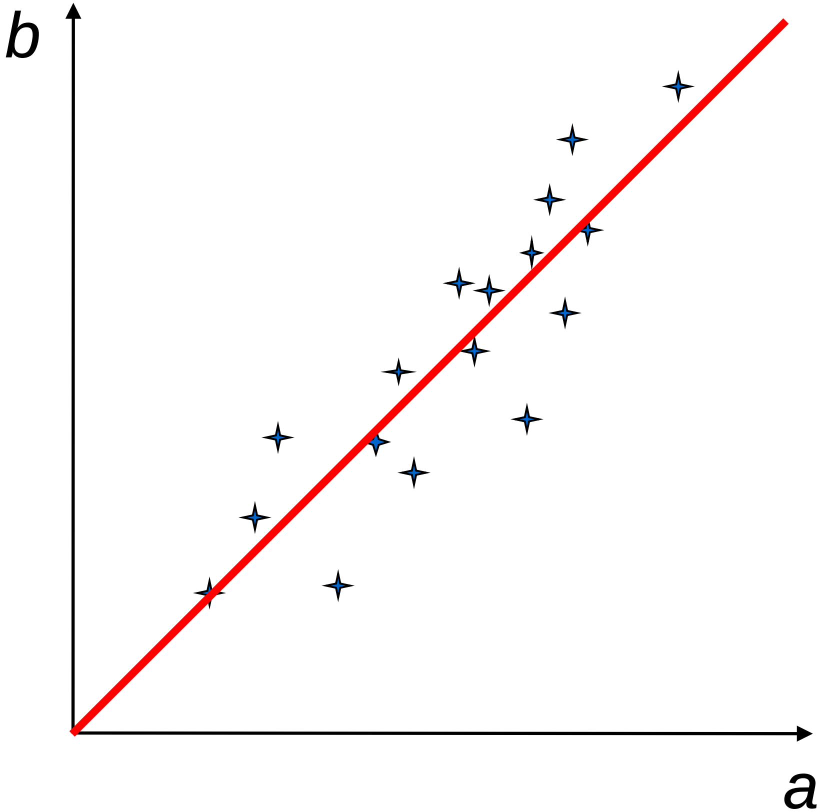
no errors in a_i



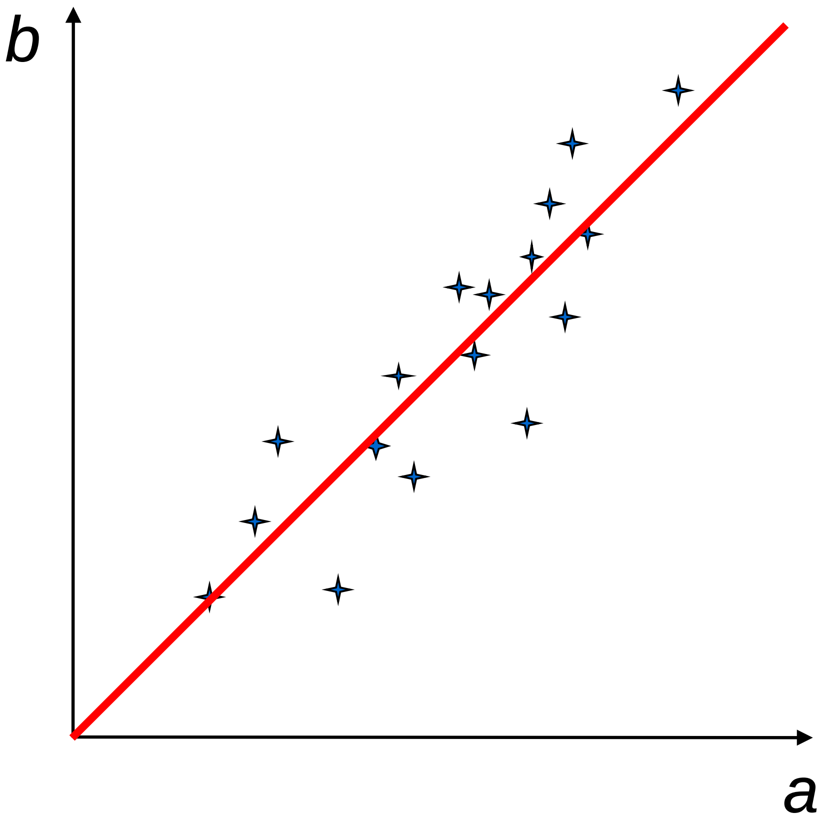
no errors in a_i



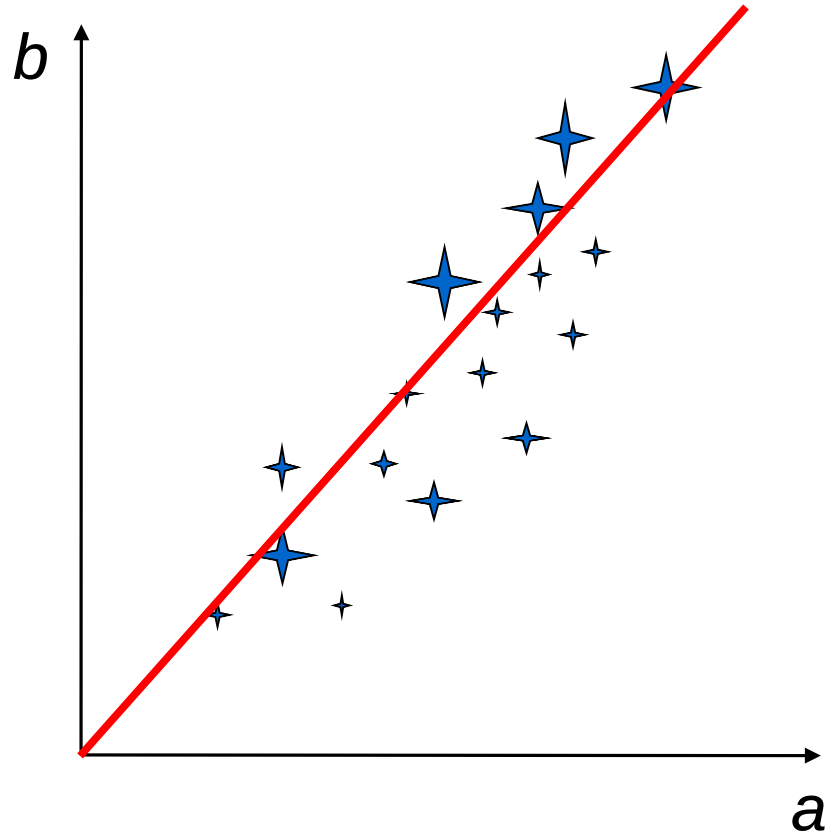
errors in a_i



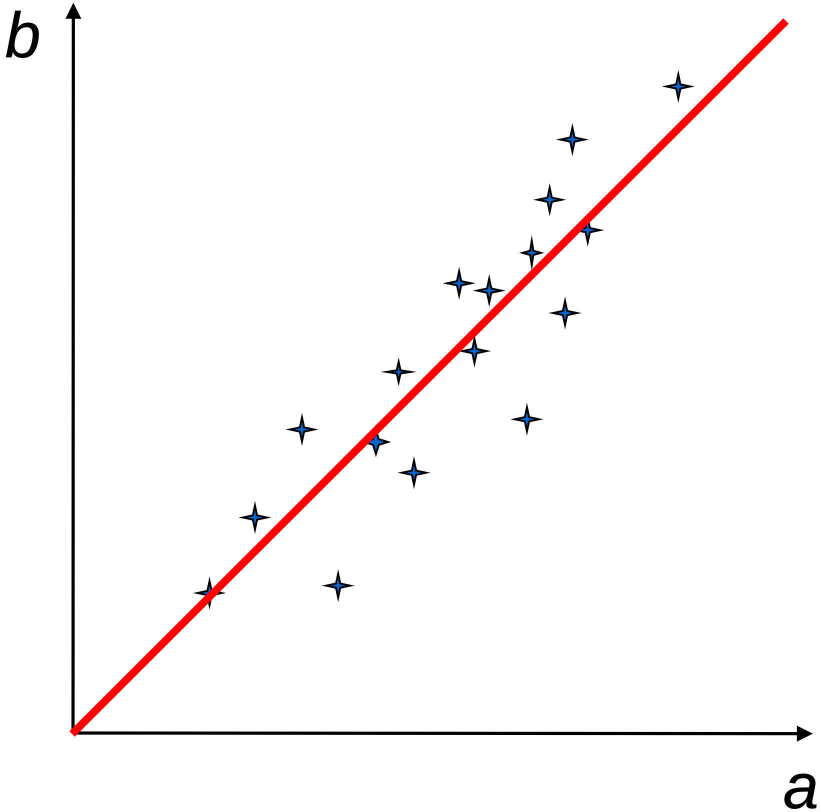
homogeneous errors



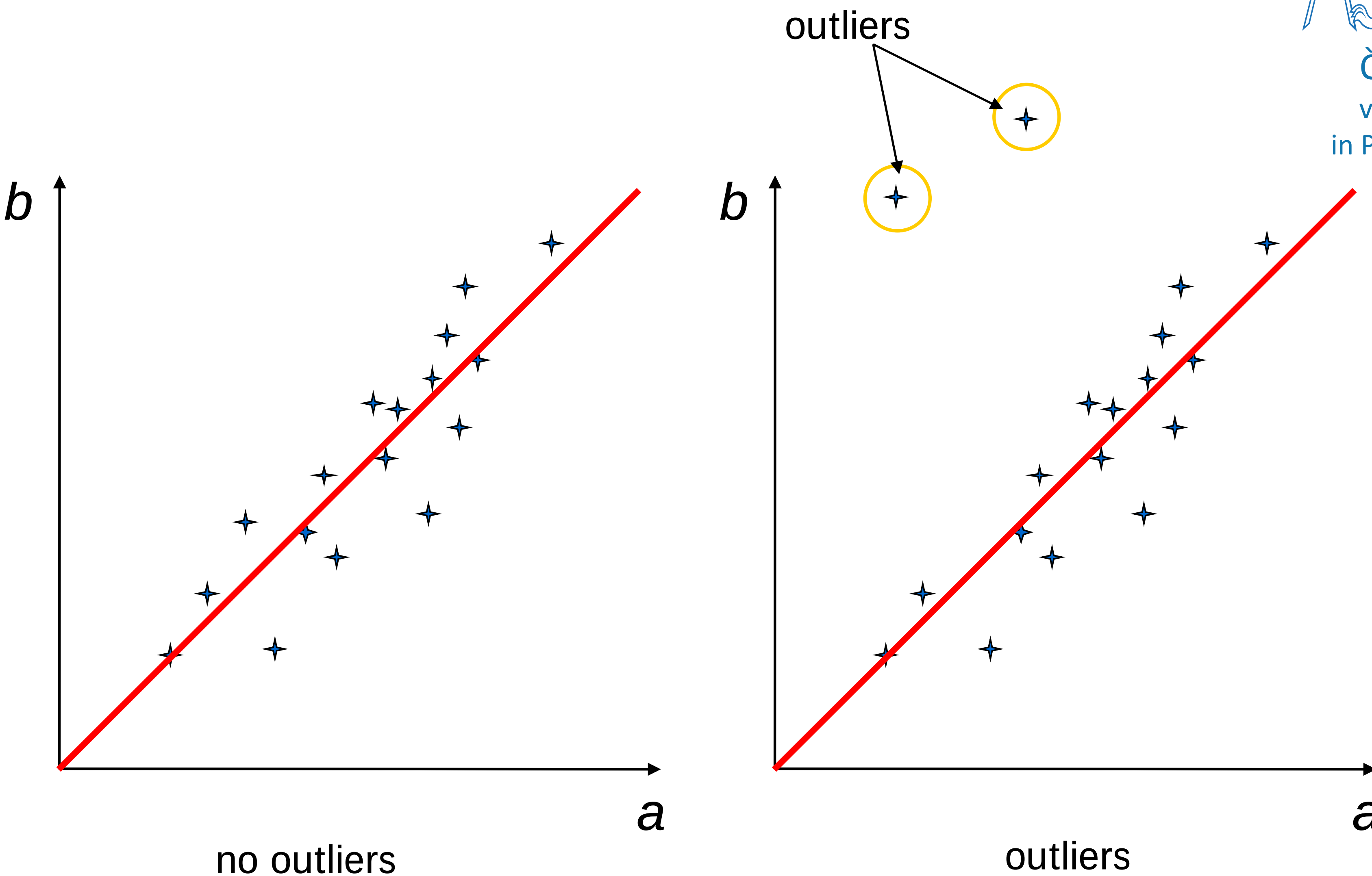
homogeneous errors



non-homogeneous errors



no outliers


outliers
b
a
no outliers
b
a
outliers