CSCI 347: Introduction to Data Mining

Lecture 2b - Linear Algebra

COMMON DATA FORMATS

- Data can often be represented by a *data matrix D*

$$D = \begin{array}{c|cccc} & X_1 & X_2 & X_3 & X_4 \\ x_1 & 0.2 & 23 & A & 5.7 \\ x_2 & 0.4 & 1 & B & 5.4 \\ x_3 & 1.8 & 0.5 & C & 5.2 \\ x_4 & 5.6 & 50 & A & 5.1 \\ x_5 & -0.5 & 34 & A & 5.3 \\ x_6 & 0.4 & 19 & B & 5.4 \\ x_7 & 1.1 & 11 & A & 5.5 \end{array}$$

COMMON DATA FORMATS

- Data can often be represented by a *data matrix* D

The columns commonly represent attributes/properties of the data

The rows commonly represent entities and their observed values for each attribute

$$D = \begin{array}{c|cccc} & X_1 & X_2 & X_3 & X_4 \\ \hline x_1 & 0.2 & 23 & A & 5.7 \\ x_2 & 0.4 & 1 & B & 5.4 \\ x_3 & 1.8 & 0 & C & 5.2 \\ x_4 & 5.6 & 50 & A & 5.1 \\ x_5 & -0.5 & 34 & A & 5.3 \\ x_6 & 0.4 & 19 & B & 5.4 \\ x_7 & 1.1 & 11 & A & 5.5 \end{array}$$

REVIEW STATS

- Estimated Mean $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$

- Estimated Variance $\hat{\sigma}_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \hat{\mu})^2$

- Estimated Std deviation $\hat{\sigma}_j = \sqrt{\hat{\sigma}_j^2}$

- Estimated covariance $\hat{\sigma}_{12} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i1} - \hat{\mu}_1)(x_{i2} - \hat{\mu}_2)$

- Covariance matrix $\Sigma = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \hat{\sigma}_{13} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 & \hat{\sigma}_{23} \\ \hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_3^2 \end{pmatrix}$

REVIEW DATA NORMALIZATION

- Range normalization $$x'_{ij} = \frac{x_{ij} - \min_i\{x_{ij}\}}{\max_i\{x_{ij}\} - \min_i\{x_{ij}\}}$$
- Mean centering $$x'_{ij} = x_{ij} - \hat{\mu}_j$$
- Z-Score normalization $$x'_{ij} = \frac{x_{ij} - \hat{\mu}_j}{\hat{\sigma}_j}$$

GEOMETRIC VIEW OF DATA

- Projection

$$D = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{array} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \\ 0.2 & 23 & A & 5.7 \\ 0.4 & 1 & B & 5.4 \\ 1.8 & 0.5 & C & 5.2 \\ 5.6 & 50 & A & 5.1 \\ -0.5 & 34 & A & 5.3 \\ 0.4 & 19 & B & 5.4 \\ 1.1 & 11 & A & 5.5 \end{array}$$

GEOMETRIC VIEW OF DATA

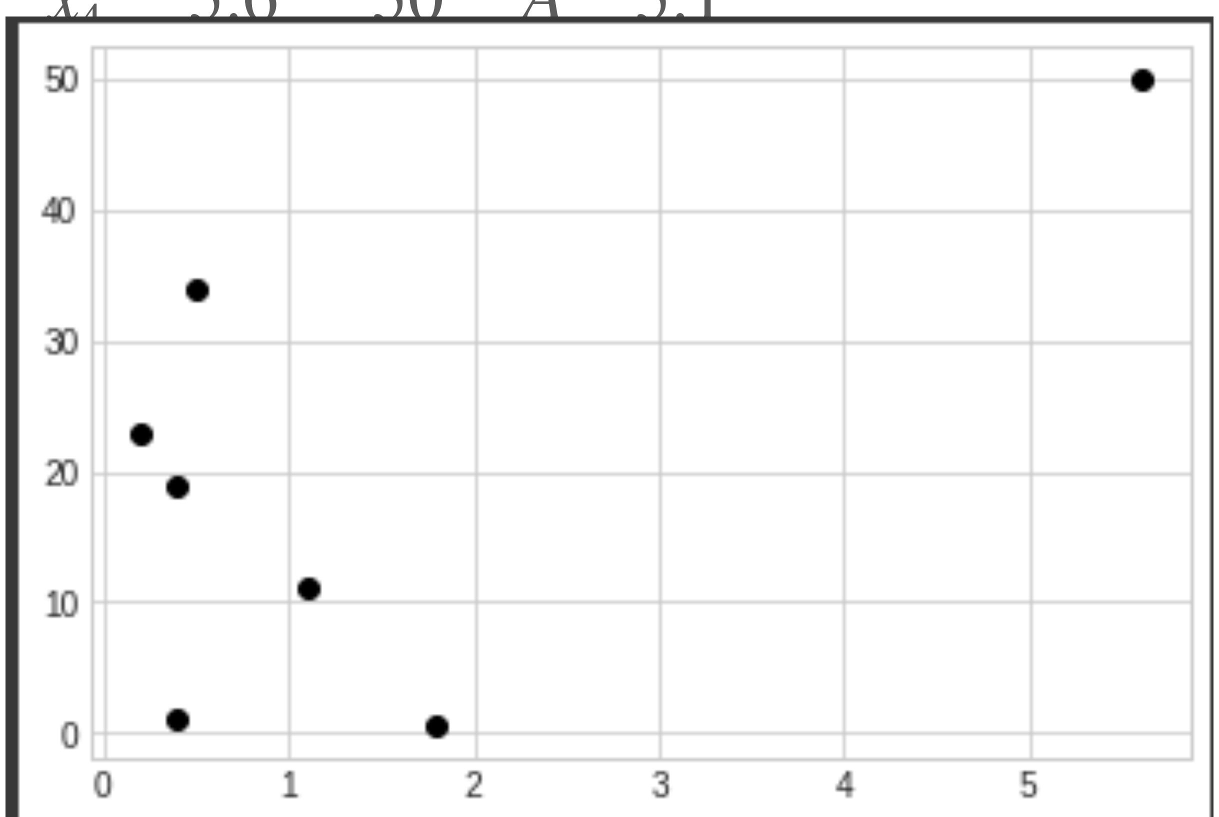
- Projection

$$D = \begin{array}{c|cccc} & X_1 & X_2 & X_3 & X_4 \\ x_1 & 0.2 & 23 & A & 5.7 \\ x_2 & 0.4 & 1 & B & 5.4 \\ x_3 & 1.8 & 0.5 & C & 5.2 \\ x_4 & 5.6 & 50 & A & 5.1 \\ x_5 & -0.5 & 34 & A & 5.3 \\ x_6 & 0.4 & 19 & B & 5.4 \\ x_7 & 1.1 & 11 & A & 5.5 \end{array} \xrightarrow{\pi_{12}} D' = \begin{array}{c|cc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

GEOMETRIC VIEW OF DATA

- Projection

$$D = \begin{array}{c|cccc} & X_1 & X_2 & X_3 & X_4 \\ x_1 & 0.2 & 23 & A & 5.7 \\ x_2 & 0.4 & 1 & B & 5.4 \\ x_3 & 1.8 & 0.5 & C & 5.2 \\ x_4 & 5.6 & 50 & A & 5.1 \end{array}$$



$$D' = \begin{array}{c|cc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

GEOMETRIC VIEW OF DATA

- Projection and re-lable (… technically reflect)

$$D = \begin{array}{c|cccc} & X_1 & X_2 & X_3 & X_4 \\ x_1 & 0.2 & 23 & A & 5.7 \\ x_2 & 0.4 & 1 & B & 5.4 \\ x_3 & 1.8 & 0.5 & C & 5.2 \\ x_4 & 5.6 & 50 & A & 5.1 \\ x_5 & -0.5 & 34 & A & 5.3 \\ x_6 & 0.4 & 19 & B & 5.4 \\ x_7 & 1.1 & 11 & A & 5.5 \end{array}$$

GEOMETRIC VIEW OF DATA

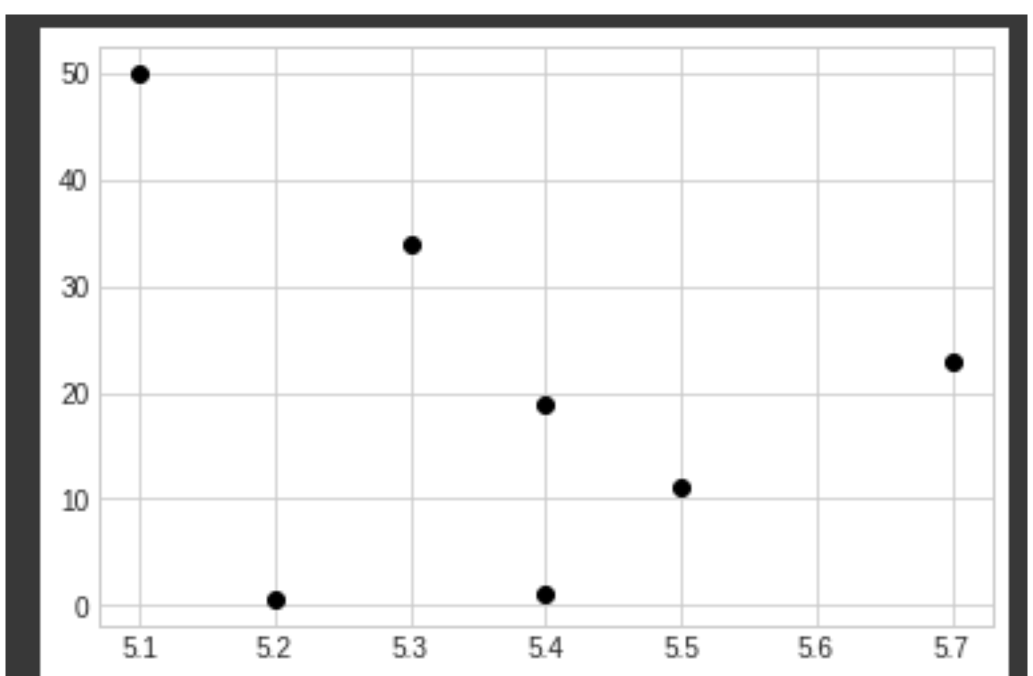
➤ Projection and re-lable (… technically reflect)

$$D = \begin{array}{c|cccc} & X_1 & X_2 & X_3 & X_4 \\ \hline x_1 & 0.2 & 23 & A & 5.7 \\ x_2 & 0.4 & 1 & B & 5.4 \\ x_3 & 1.8 & 0.5 & C & 5.2 \\ x_4 & 5.6 & 50 & A & 5.1 \\ x_5 & -0.5 & 34 & A & 5.3 \\ x_6 & 0.4 & 19 & B & 5.4 \\ x_7 & 1.1 & 11 & A & 5.5 \end{array} \xrightarrow{\pi_{42}} D = \begin{array}{c|cc} & X_1' & X_2' \\ \hline x_1 & 5.7 & 23 \\ x_2 & 5.4 & 1 \\ x_3 & 5.2 & 0.5 \\ x_4 & 5.1 & 50 \\ x_5 & 5.3 & 34 \\ x_6 & 5.4 & 19 \\ x_7 & 5.5 & 11 \end{array}$$

GEOMETRIC VIEW OF DATA

➤ Projection and re-lable (… technically reflect)

$$D = \begin{array}{c|cccc} & X_1 & X_2 & X_3 & X_4 \\ x_1 & 0.2 & 23 & A & 5.7 \\ x_2 & 0.4 & 1 & B & 5.4 \\ x_3 & 1.8 & 0.5 & C & 5.2 \\ x_4 & 5.6 & 50 & A & 5.1 \\ x_5 & -0.5 & 34 & A & 5.3 \end{array} \xrightarrow{\pi_{42}} D = \begin{array}{c|cc} & X_1' & X_2' \\ x_1 & 5.7 & 23 \\ x_2 & 5.4 & 1 \\ x_3 & 5.2 & 0.5 \\ x_4 & 5.1 & 50 \\ x_5 & 5.3 & 34 \\ x_6 & 5.4 & 19 \\ x_7 & 5.5 & 11 \end{array}$$

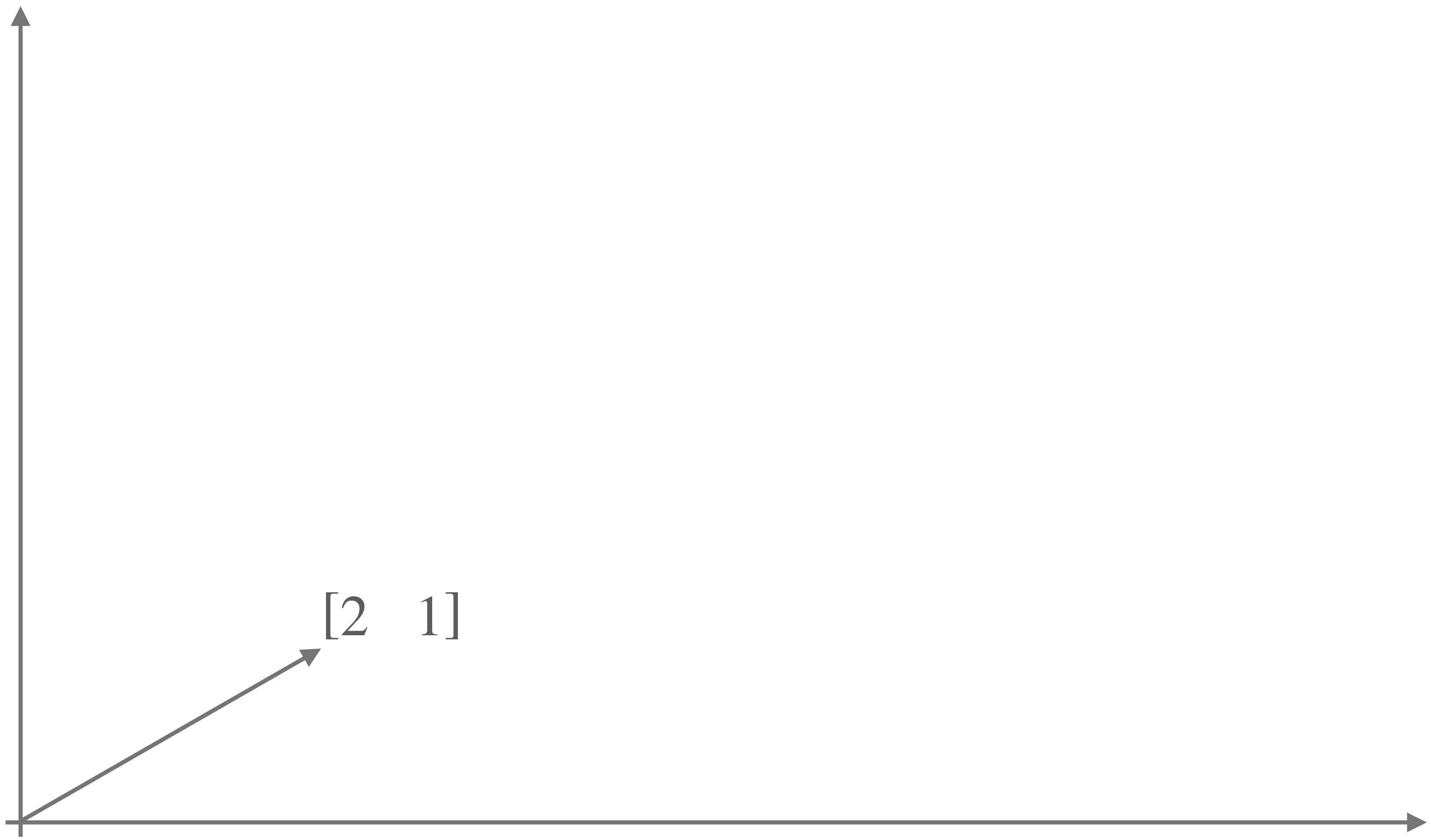


GEOMETRIC VIEW OF DATA

- Points and Vectors

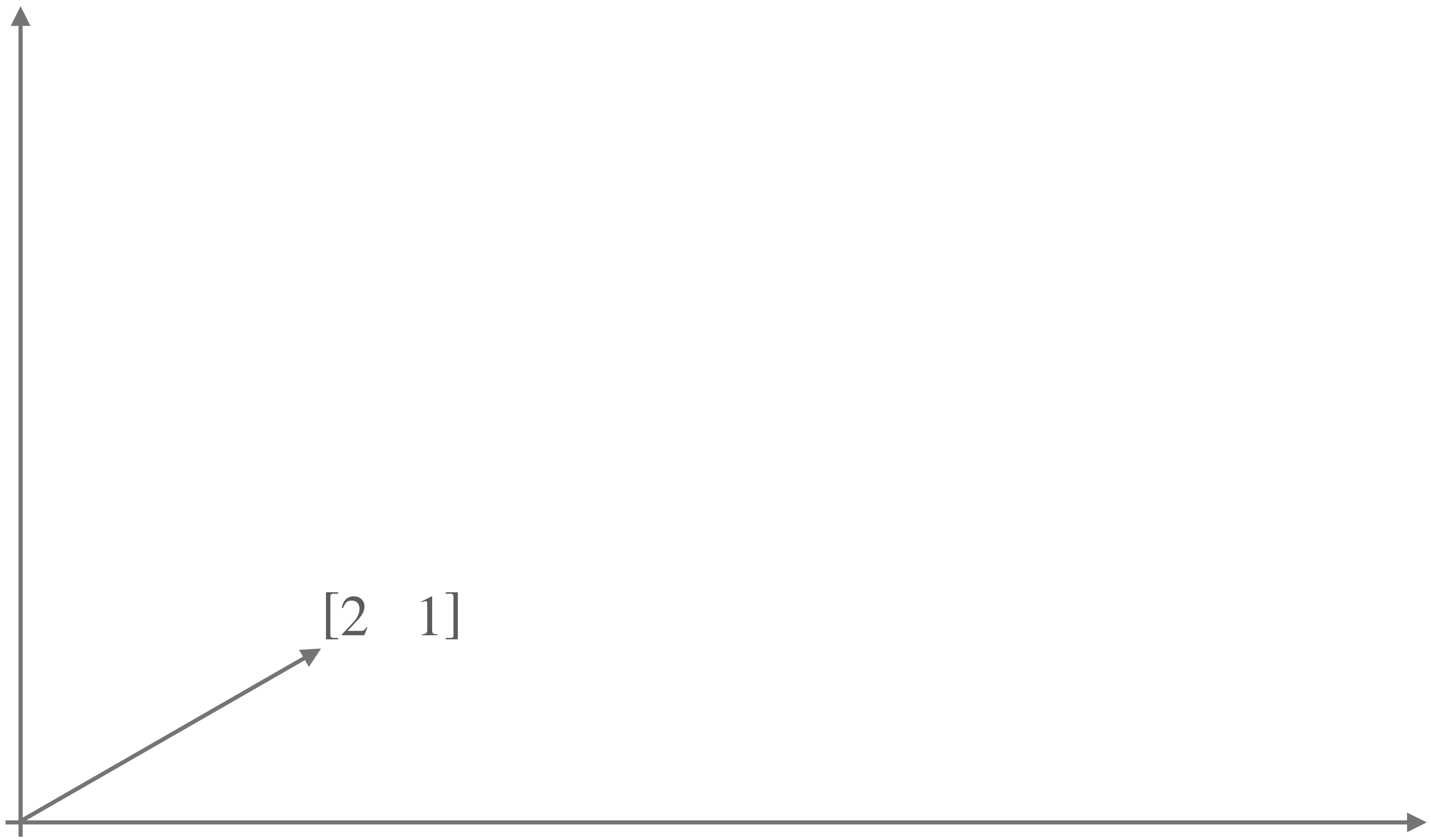
GEOMETRIC VIEW OF DATA

- Points and Vectors



GEOMETRIC VIEW OF DATA

- Points and Vectors (describe direction and magnitude)

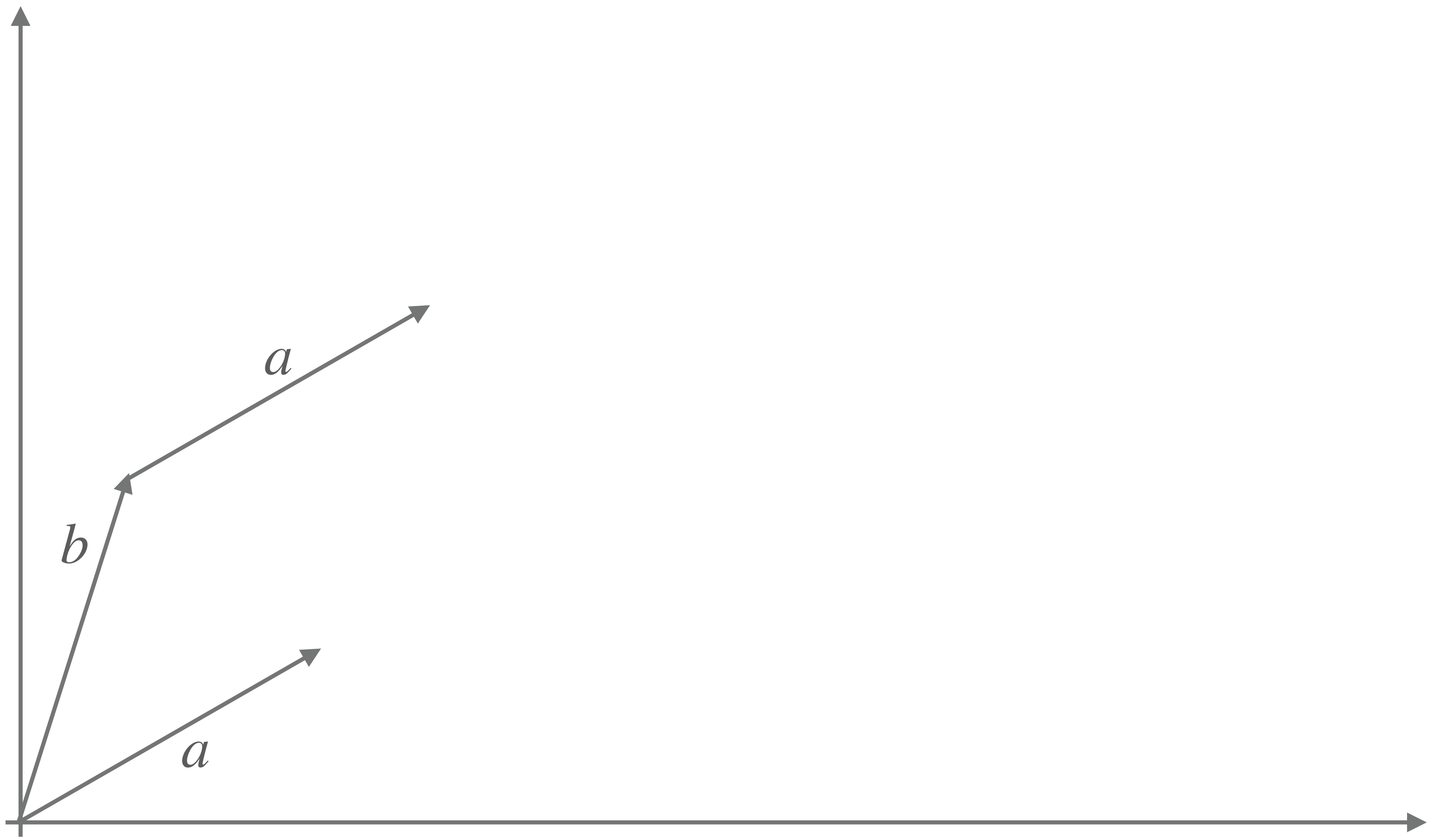


GEOMETRIC VIEW OF DATA

- Vector addition $a + b = \left[a_x + b_x \quad a_y + b_y \right]$

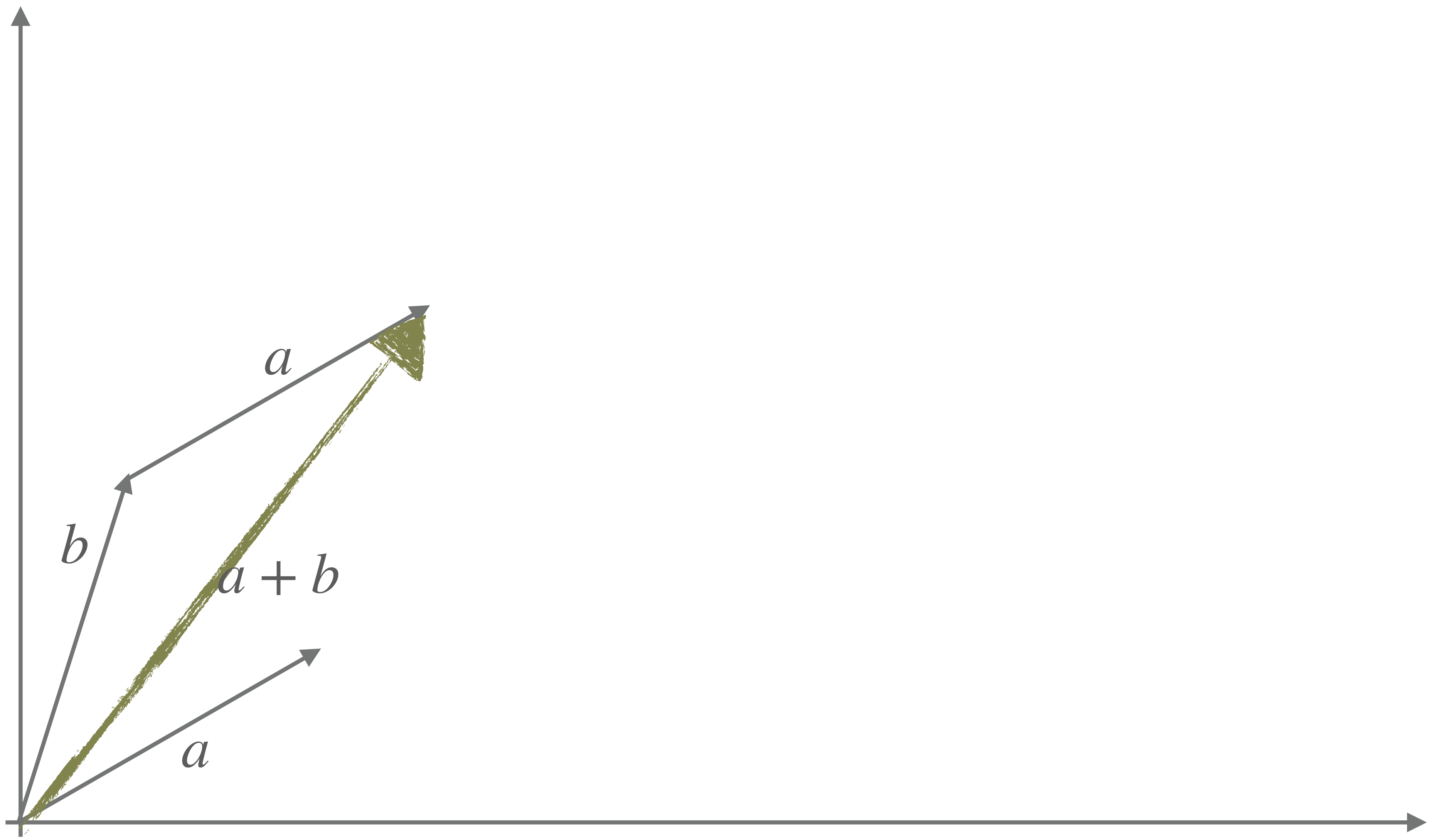
GEOMETRIC VIEW OF DATA

- Vector addition $a + b = \left[a_x + b_x \quad a_y + b_y\right]$



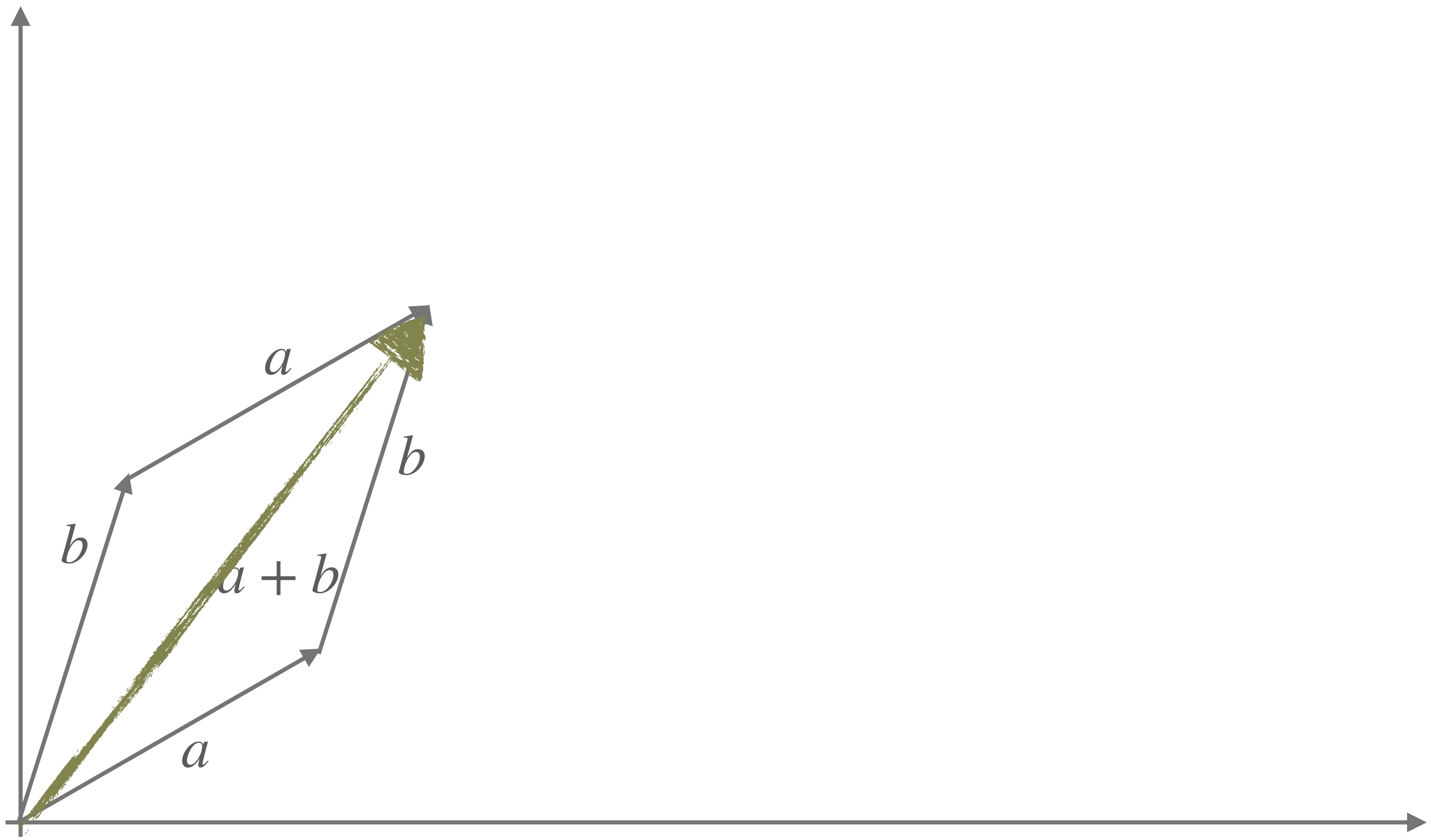
GEOMETRIC VIEW OF DATA

- Vector addition $a + b \;=\; \left[a_x + b_x \quad a_y + b_y \right]$



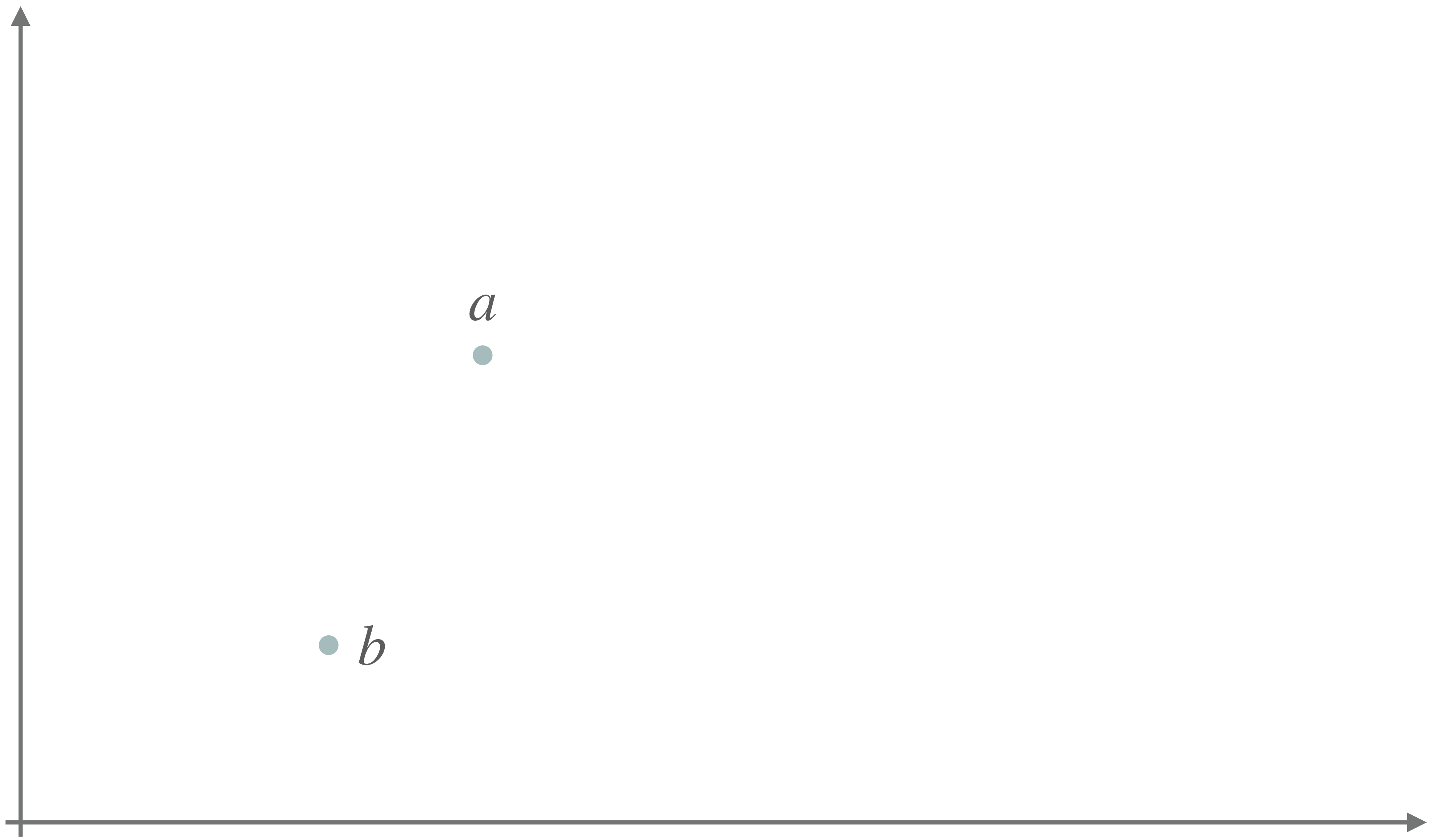
GEOMETRIC VIEW OF DATA

- Vector addition $a + b = \begin{bmatrix} a_x + b_x & a_y + b_y \end{bmatrix}$



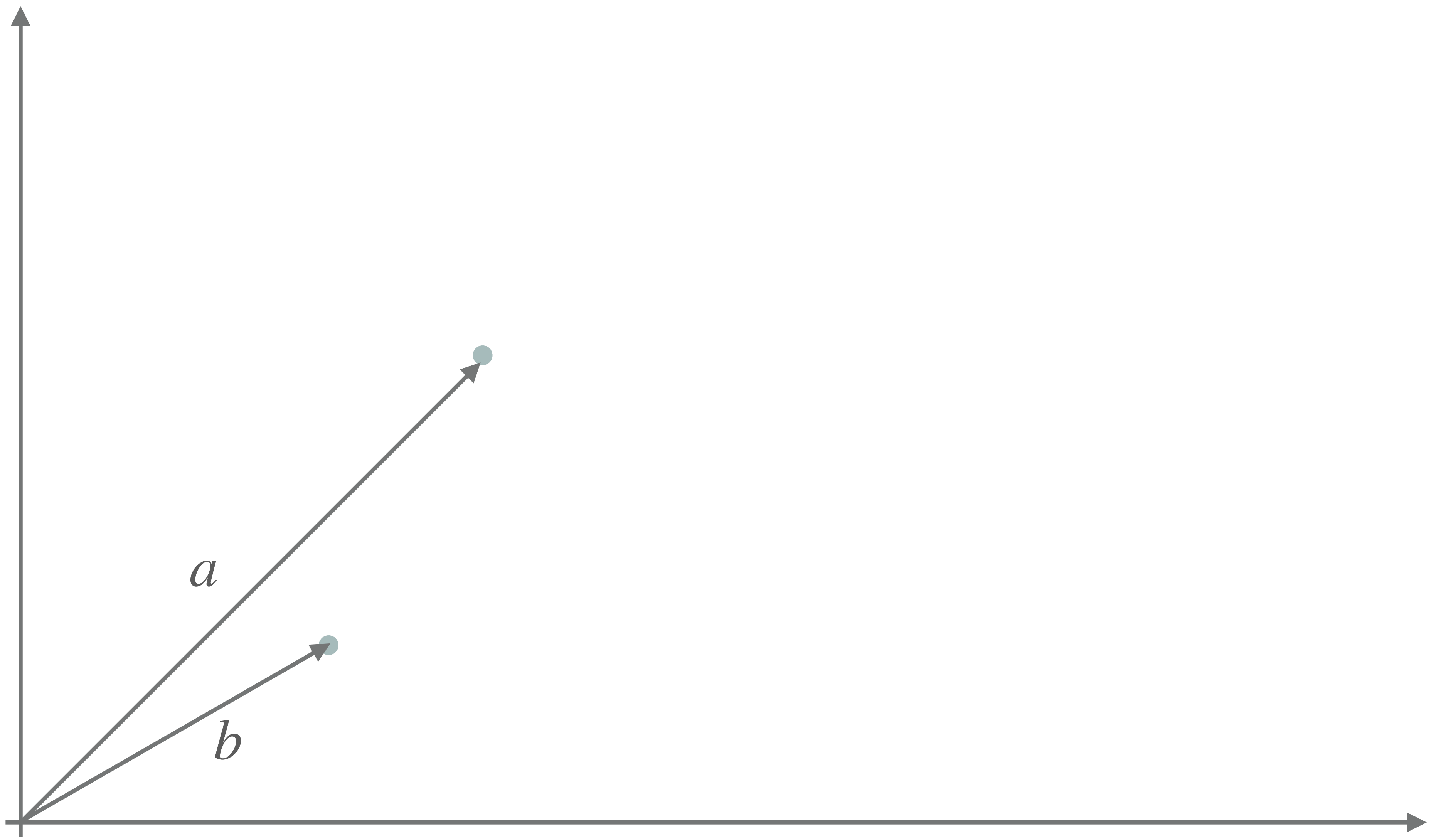
GEOMETRIC VIEW OF DATA

➤ Subtraction $a - b = \begin{bmatrix} a_x - b_x & a_y - b_y \end{bmatrix}$



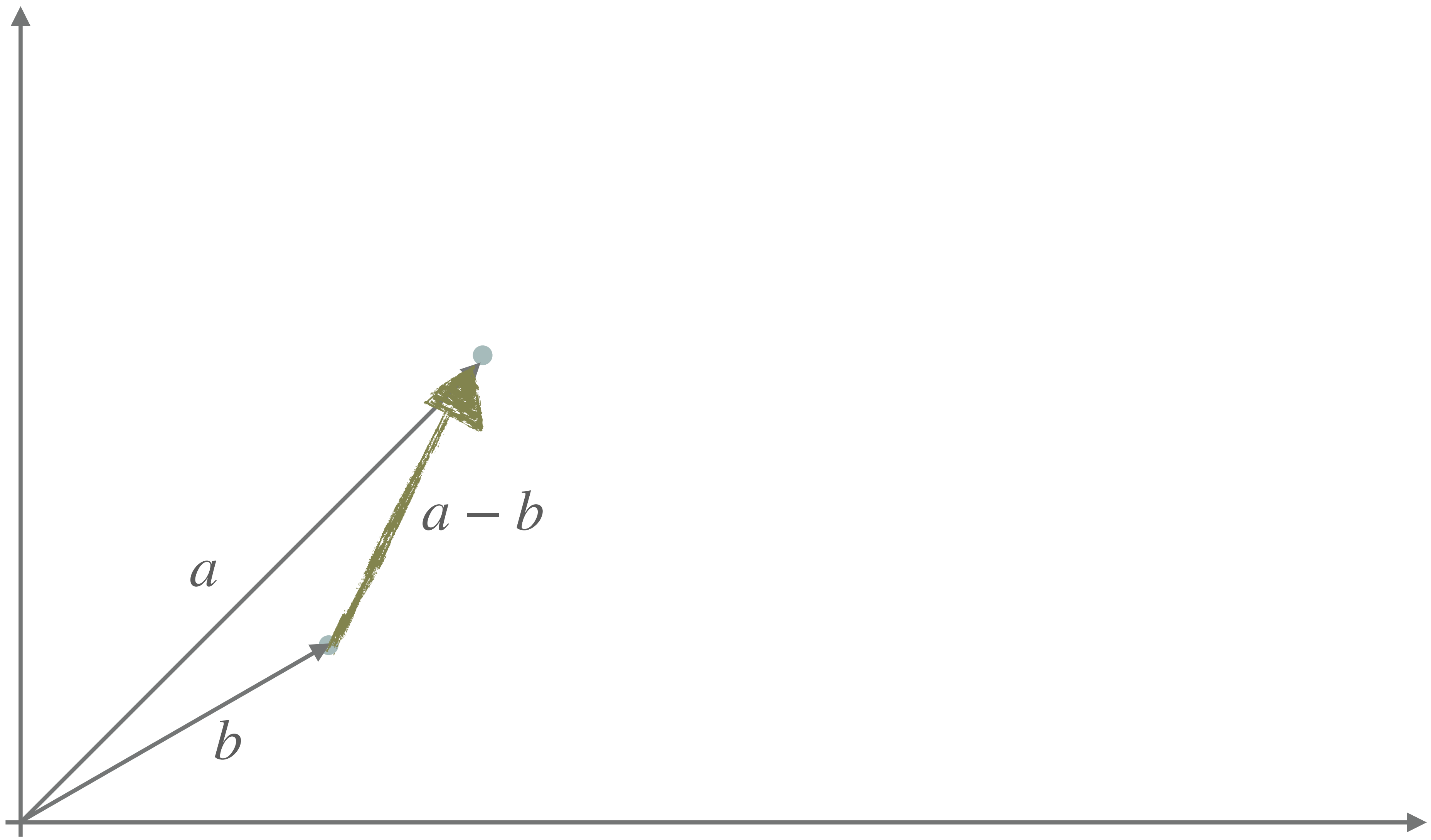
GEOMETRIC VIEW OF DATA

- Subtraction $a - b = \begin{bmatrix} a_x - b_x & a_y - b_y \end{bmatrix}$



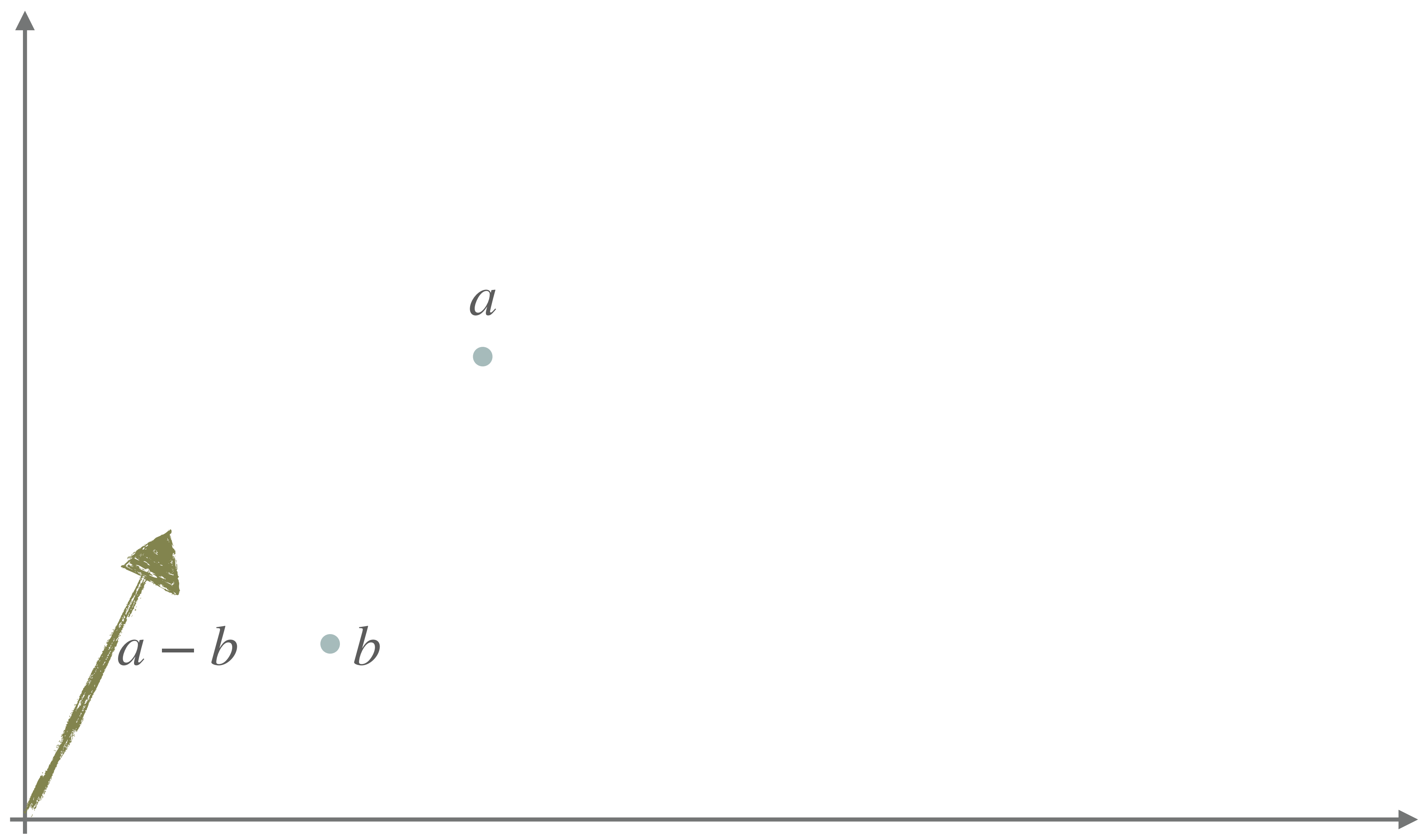
GEOMETRIC VIEW OF DATA

- Subtraction $a - b = \begin{bmatrix} a_x - b_x & a_y - b_y \end{bmatrix}$



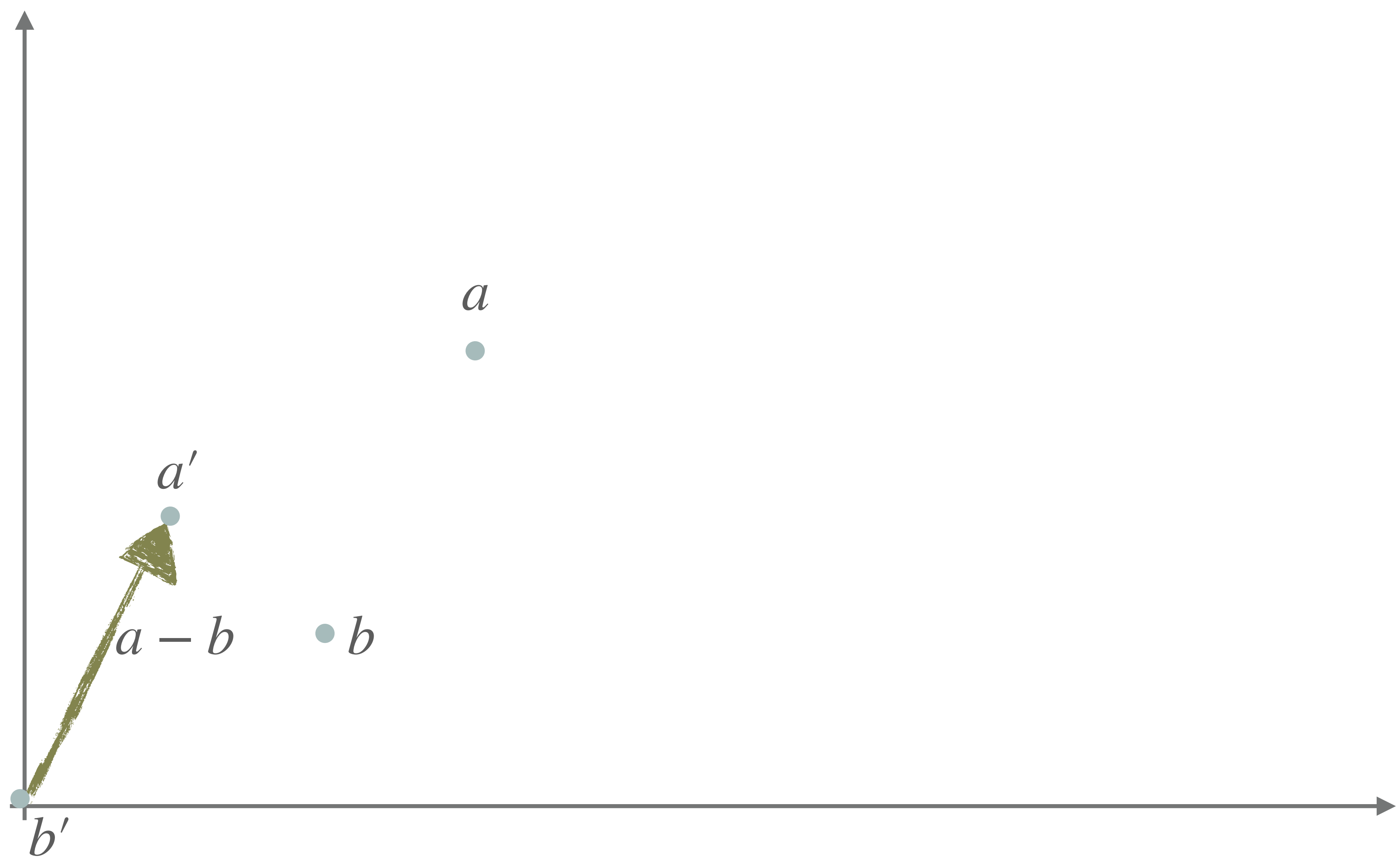
GEOMETRIC VIEW OF DATA

- Subtraction $a - b = \begin{bmatrix} a_x - b_x & a_y - b_y \end{bmatrix}$



GEOMETRIC VIEW OF DATA

- Subtraction $a - b = \begin{bmatrix} a_x - b_x & a_y - b_y \end{bmatrix}$

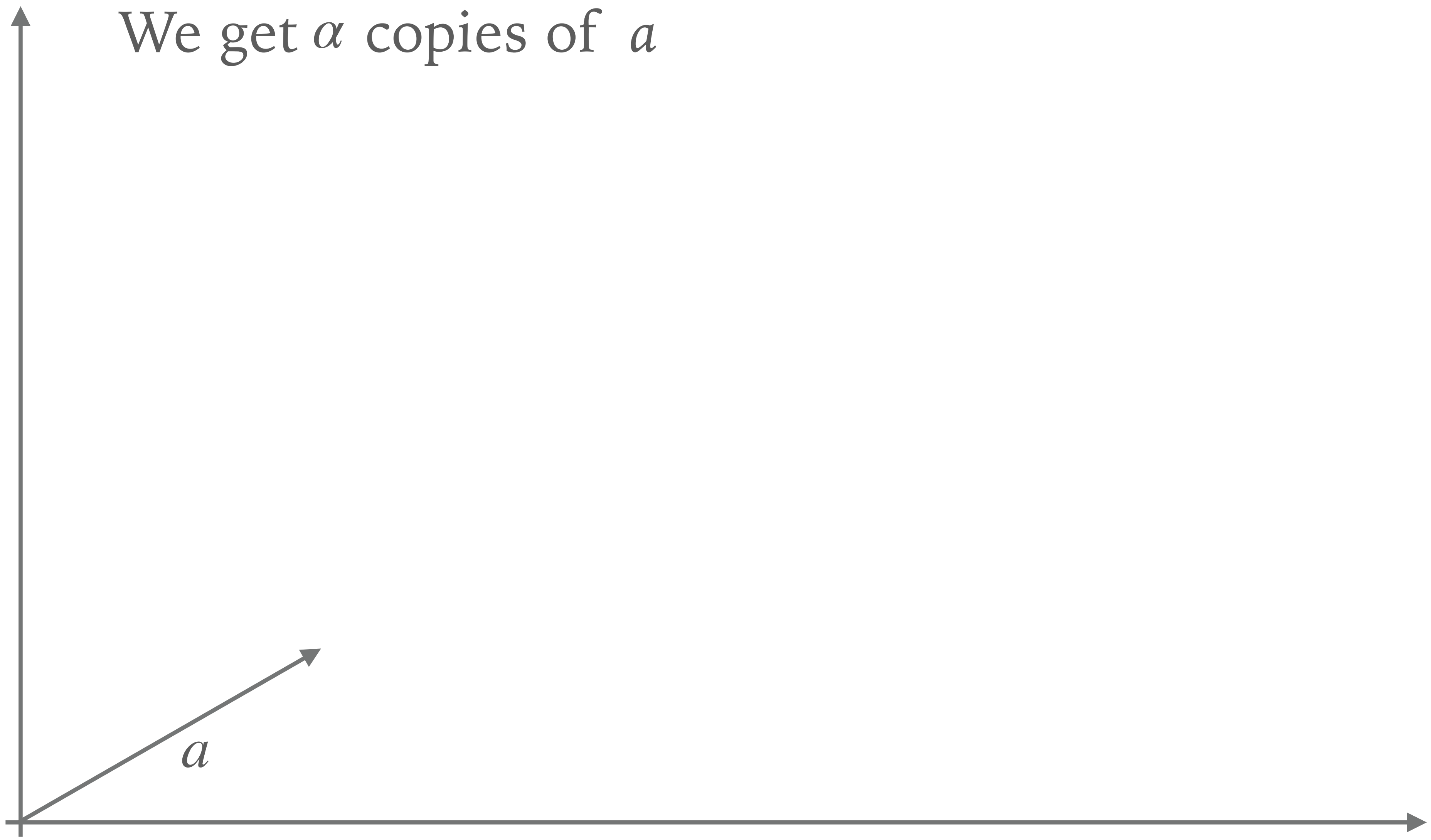


GEOMETRIC VIEW OF DATA

- Scaling: For $\alpha \in \mathbb{R}, \alpha a = \begin{bmatrix} \alpha a_x & \alpha a_y \end{bmatrix}$

GEOMETRIC VIEW OF DATA

- Scaling: For $\alpha \in \mathbb{R}, \alpha a = \begin{bmatrix} \alpha a_x & \alpha a_y \end{bmatrix}$

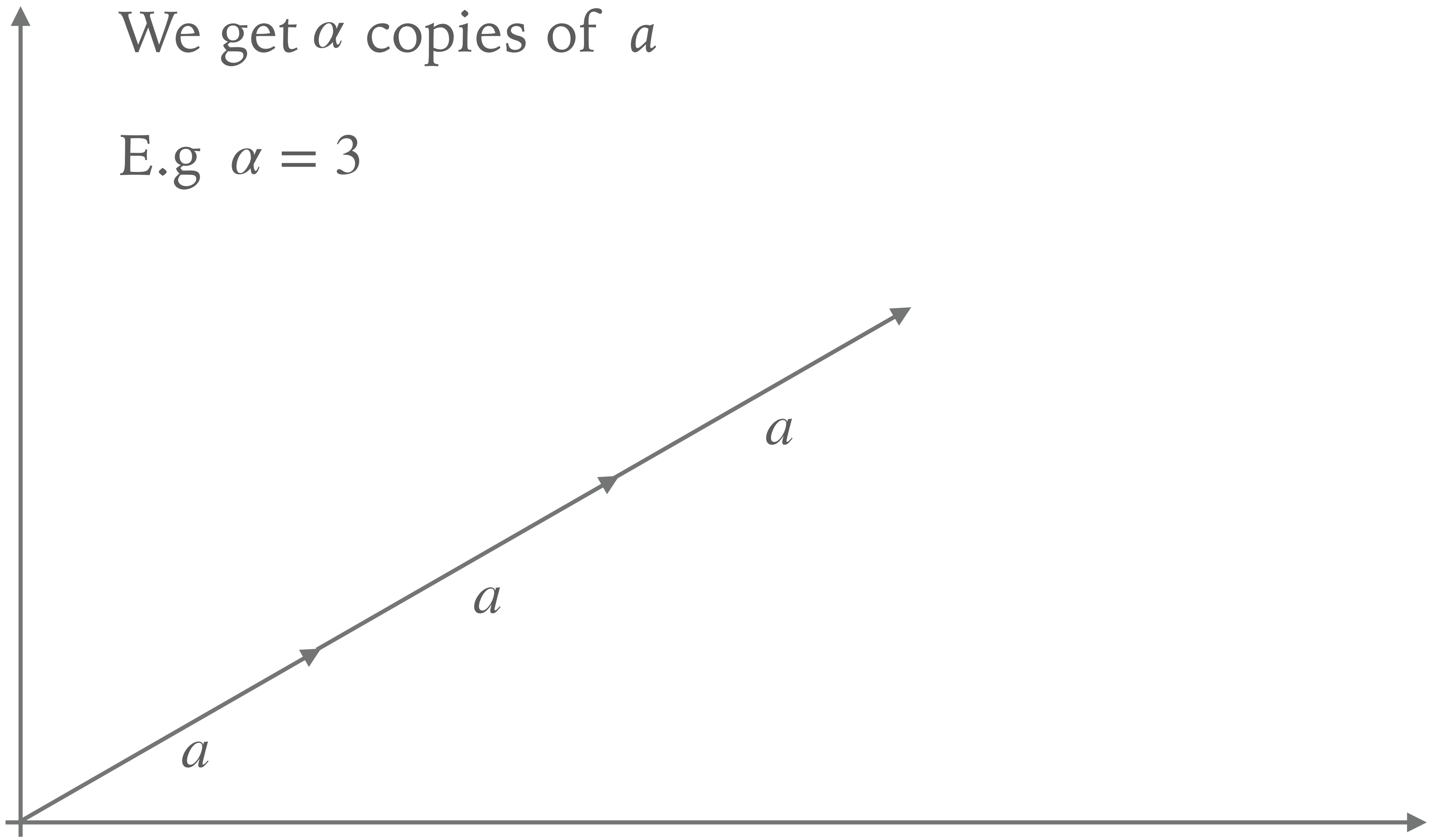


GEOMETRIC VIEW OF DATA

- Scaling: For $\alpha \in \mathbb{R}, \alpha a = \begin{bmatrix} \alpha a_x & \alpha a_y \end{bmatrix}$

We get α copies of a

E.g $\alpha = 3$

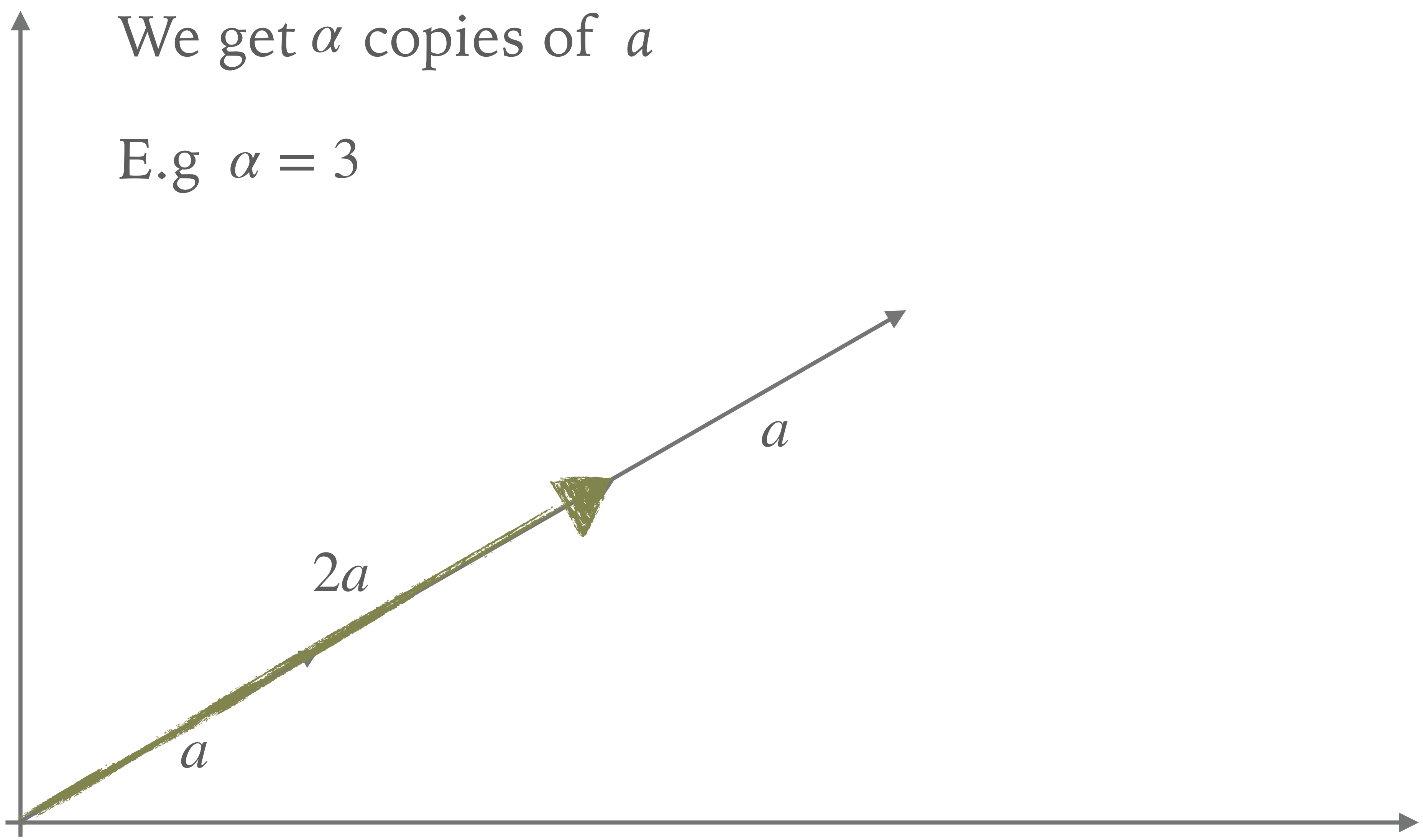


GEOMETRIC VIEW OF DATA

- Scaling: For $\alpha \in \mathbb{R}, \alpha a = \begin{bmatrix} \alpha a_x & \alpha a_y \end{bmatrix}$

We get α copies of a

E.g $\alpha = 3$

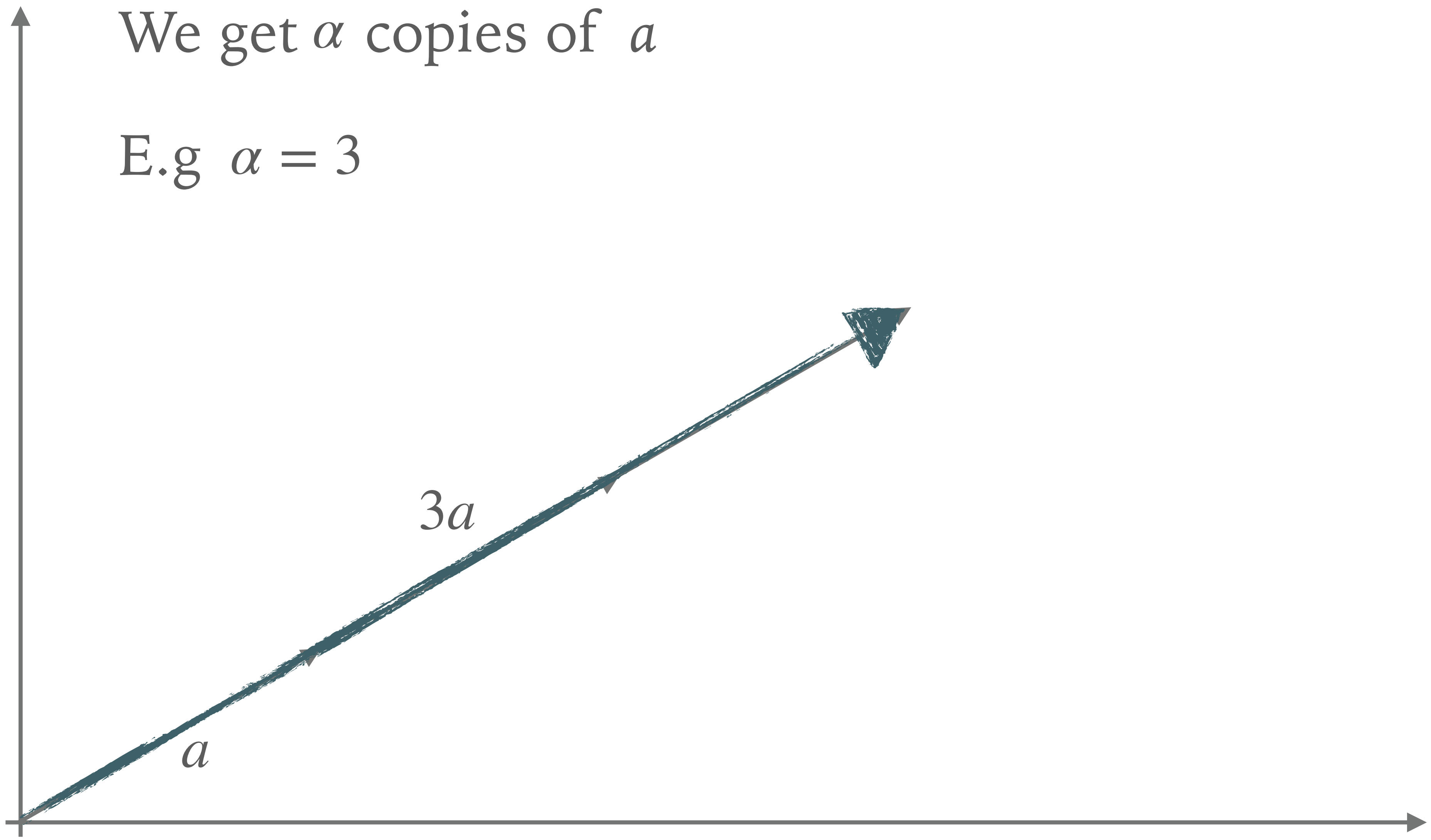


GEOMETRIC VIEW OF DATA

- Scaling: For $\alpha \in \mathbb{R}, \alpha a = \begin{bmatrix} \alpha a_x & \alpha a_y \end{bmatrix}$

We get α copies of a

E.g $\alpha = 3$



DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

First, some notation:

L_2 norm of a vector x_i with m dimensions (columns/attributes):

$$\|x_i\|_2 = \sqrt{\sum_{k=1}^{m} x_{ik}^2}$$

$$D = \begin{array}{cccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

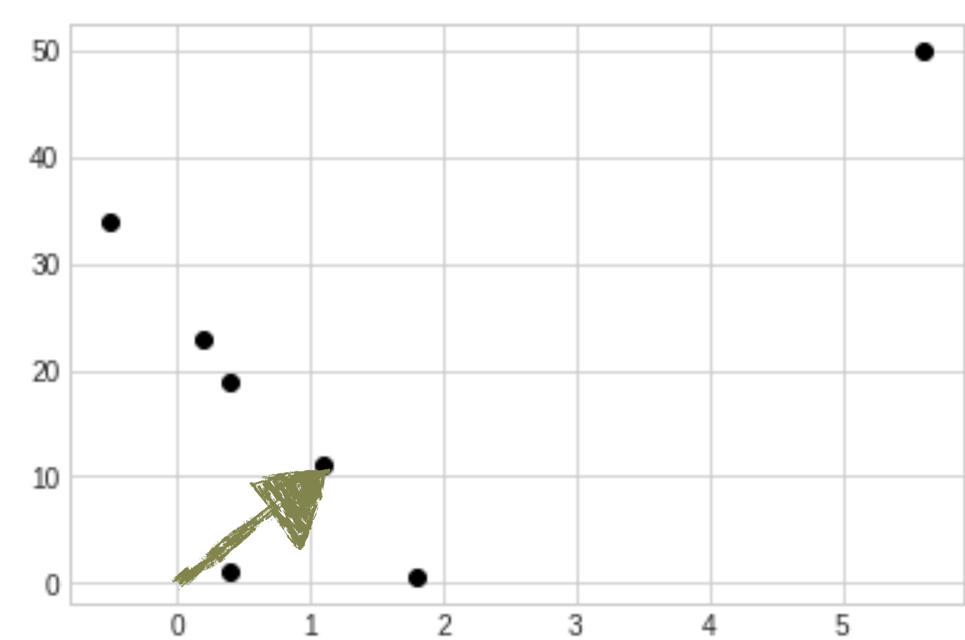
First, some notation:

L_2 norm of a vector x_i with m dimensions:

$$\|x_i\|_2 = \sqrt{\sum_{k=1}^{m} x_{ik}^2}$$

$$D = \begin{array}{c|cc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

$$\|x_7\|_2 = \sqrt{\sum_{k=1}^{2} x_{7k}^2} = \sqrt{(x_{71}^2 + x_{72}^2)} = \sqrt{(1.1^2 + 11^2)} = 11.05$$



DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

L_2 **norm:**

$$\|x_i - x_j\|_2 = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2}$$ **where x_i and x_j are vectors, and there are m dimensions**

$$D = \begin{array}{cccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

L_2 **norm:**

$$\|x_i - x_j\|_2 = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2}$$ **where x_i and x_j are vectors, and there are m dimensions**

$$D = \begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

$$\|x_1 - x_2\|_2 = \sqrt{\sum_{k=1}^{3} (x_{1k} - x_{2k})^2}$$

$$= \sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2 + (x_{13} - x_{23})^2}$$

$$= \sqrt{(0.2 - 0.4)^2 + (23 - 1)^2 + (5.7 - 5.4)^2}$$

$$= \sqrt{(-0.2)^2 + (22)^2 + (0.3)^2}$$

$$= 22.0$$

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

L_1 norm:

$$\|x_i - x_j\|_1 = \sum_{k=1}^{m} |x_{ik} - x_{jk}|$$ **where x_i and x_j are vectors, and there are m dimensions**

$$D = \begin{array}{cccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

$$\|x_1 - x_2\|_1 = \sum_{k=1}^{3} |x_{1k} - x_{2k}|$$

$$= |x_{11} - x_{21}| + |x_{12} - x_{22}| + |x_{13} - x_{23}|$$

$$= |0.2 - 0.4| + |23 - 1| + |5.7 - 5.4|$$

$$= |-0.2| + |22| + |0.3|$$

$$= 22.5$$

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

L_p **norm:**

$$\|x_i - x_j\|_p = \sqrt[p]{\sum_{k=1}^{m} |x_{ik} - x_{jk}|^p}$$ **where x_i and x_j are vectors, and there are m dimensions**

$$D = \begin{array}{cccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

$$\|x_1 - x_2\|_4 = \sqrt{4} \sum_{k=1}^{3} |x_{1k} - x_{2k}|^4$$

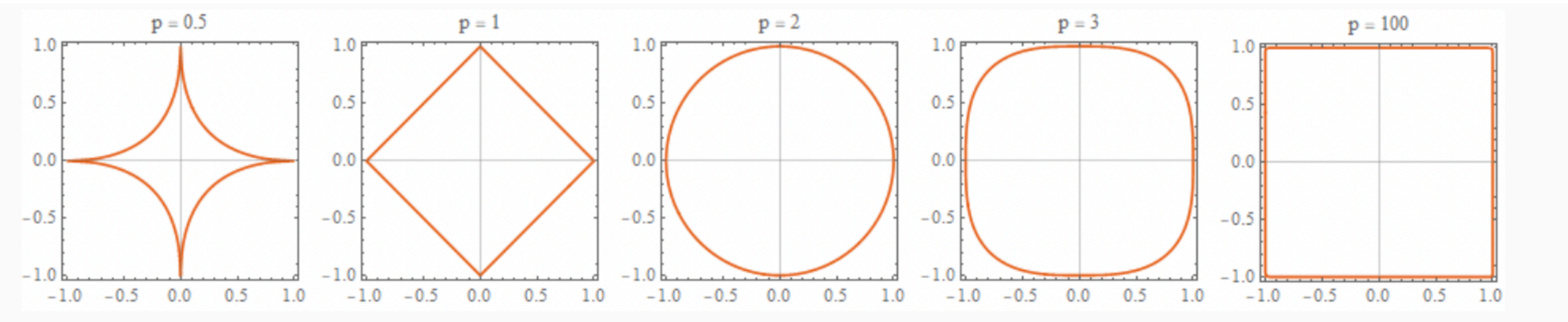
$$= \sqrt[4]{|x_{11} - x_{21}|^4 + |x_{12} - x_{22}|^4 + |x_{13} - x_{23}|^4}$$

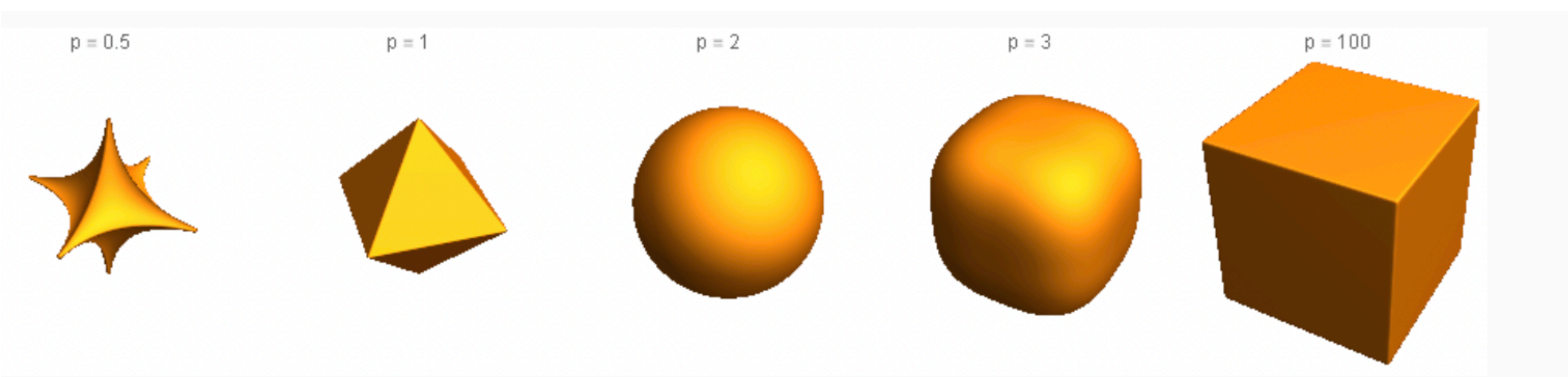
$$= \sqrt[4]{|0.2 - 0.4|^4 + |23 - 1|^4 + |5.7 - 5.4|^4}$$

$$\approxeq 22$$

DISTANCE BETWEEN VECTORS

- Set of points in which p-norm is 1 in 2d and 3d





Images from https://ekamperi.github.io/machine%20learning/2019/10/19/norms-in-machine-learning.html

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

Dot product:

$$a \cdot b = \sum_{k=1}^{m} a_k b_k$$

where a and b are vectors,

and there are m dimensions

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

Dot product:

$$a \cdot b = \sum_{k=1}^{m} a_k b_k$$ **where a and b are vectors, and there are m dimensions**

$$D = \begin{array}{cccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

$$x_3 \cdot x_4 = \sum_{k=1}^{3} x_{3k} x_{4k}$$

$$= x_{31}x_{41} + x_{32}x_{42} + x_{33}x_{43}$$

$$= (1.8)(5.6) + (0.5)(50) + (5.2)(5.1)$$

$$= 61.6$$

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

Cosine of the angle between two vectors x_i and x_j:

$$cos(\theta) = \frac{x_i \cdot x_j}{\|x_i\|_2 \|x_j\|_2}$$ **where x_i and x_j are vectors and $x_i \cdot x_j$ is their dot product**

$$D = \begin{array}{cccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

Cosine of the angle between two vectors x_i and x_j:

$$cos(\theta) = \frac{x_i \cdot x_j}{\|x_i\|_2 \|x_j\|_2}$$ **where x_i and x_j are vectors and $x_i \cdot x_j$ is their dot product**

$$D = \begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

cosine of the angle between x_2 and x_3 is:

$$\frac{x_2 \cdot x_3}{\|x_2\|_2 \|x_3\|_2}$$

DISTANCE BETWEEN VECTORS

We are often interested in some measure of distance between vectors representing separate entities.

Cosine of the angle between two vectors x_i and x_j:

$$cos(\theta) = \frac{x_i \cdot x_j}{\|x_i\|_2 \|x_j\|_2}$$ **where x_i and x_j are vectors and $x_i \cdot x_j$ is their dot product**

$$D = \begin{array}{cccc} & X_1 & X_2 & X_3 \\ x_1 & 0.2 & 23 & 5.7 \\ x_2 & 0.4 & 1 & 5.4 \\ x_3 & 1.8 & 0.5 & 5.2 \\ x_4 & 5.6 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.3 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

cosine of the angle between x_2 and x_3 is:

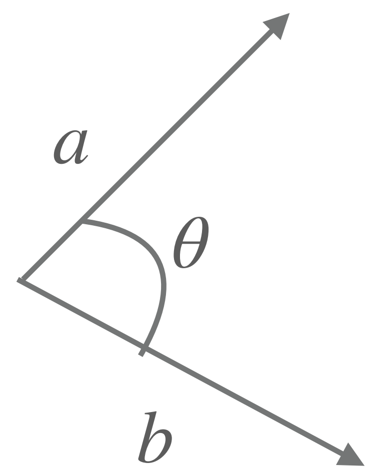
$$\frac{x_2 \cdot x_3}{\|x_2\|_2 \|x_3\|_2}$$

$$= \frac{(0.4 \quad 1 \quad 5.4) \cdot (1.8 \quad 0.5 \quad 5.2)}{\sqrt{(0.4^2 + 1^2 + 5.4^2)}\sqrt{(1.8^2 + 0.5^2 + 5.2^2)}}$$

$$= \frac{(0.4)(1.8) + (1)(0.5) + (5.4)(5.2))}{\sqrt{(0.4^2 + 1^2 + 5.4^2)}\sqrt{(1.8^2 + 0.5^2 + 5.2^2)}}$$

$$= 0.96$$

DISTANCE BETWEEN VECTORS



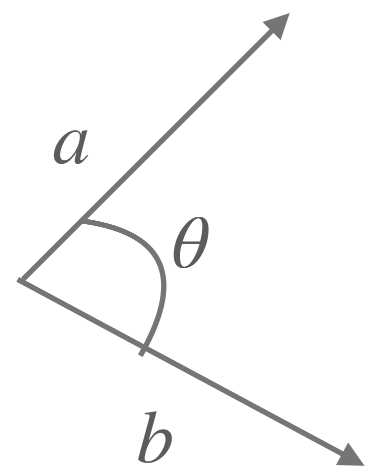
$$cos(\theta) = \frac{a \cdot b}{\|a\| \, \|b\|}$$

$$cos(\theta_1) \approx 1$$

$$cos(\theta_2) \approx 0$$

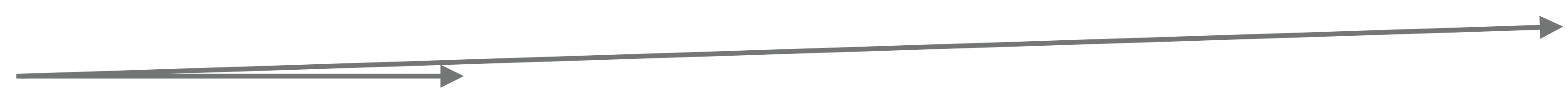
$$cos(\theta_3) \approx -1$$

DISTANCE BETWEEN VECTORS



$$cos(\theta) = \frac{a \cdot b}{\|a\| \, \|b\|}$$

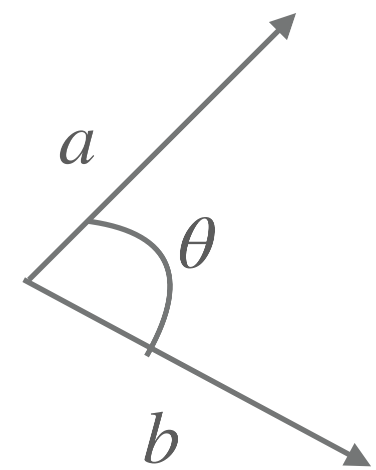
$cos(\theta_1) \approx 1$



$cos(\theta_2) \approx 0$

$cos(\theta_3) \approx -1$

DISTANCE BETWEEN VECTORS



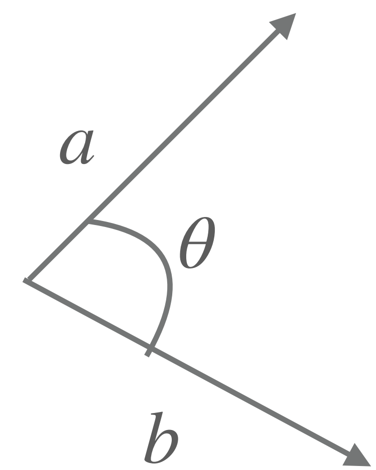
$$cos(\theta) = \frac{a \cdot b}{\|a\| \|b\|}$$

$cos(\theta_1) \approx 1$

$cos(\theta_2) \approx 0$

$cos(\theta_3) \approx -1$

DISTANCE BETWEEN VECTORS



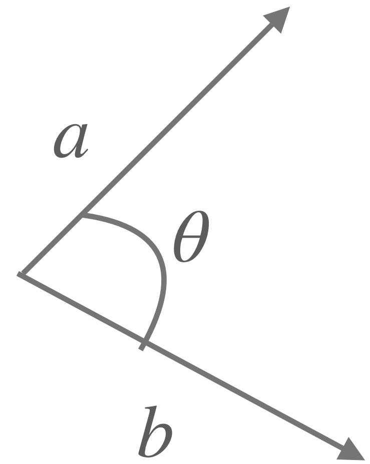
$$cos(\theta) = \frac{a \cdot b}{\|a\| \, \|b\|}$$

$$cos(\theta_1) \approx 1$$

$$cos(\theta_2) \approx 0$$

$$cos(\theta_3) \approx -1$$

DISTANCE BETWEEN VECTORS



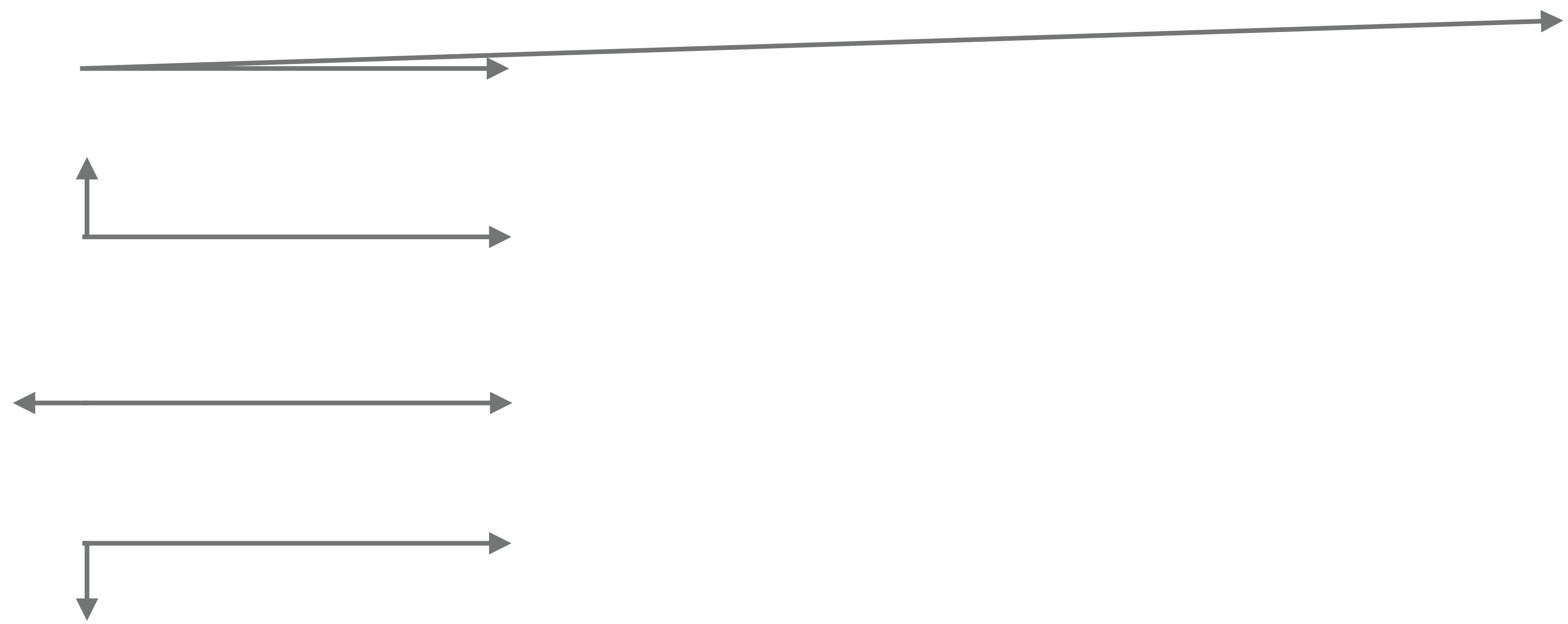
$$cos(\theta) = \frac{a \cdot b}{\|a\| \, \|b\|}$$

$$cos(\theta_1) \approx 1$$

$$cos(\theta_2) \approx 0$$

$$cos(\theta_3) \approx -1$$

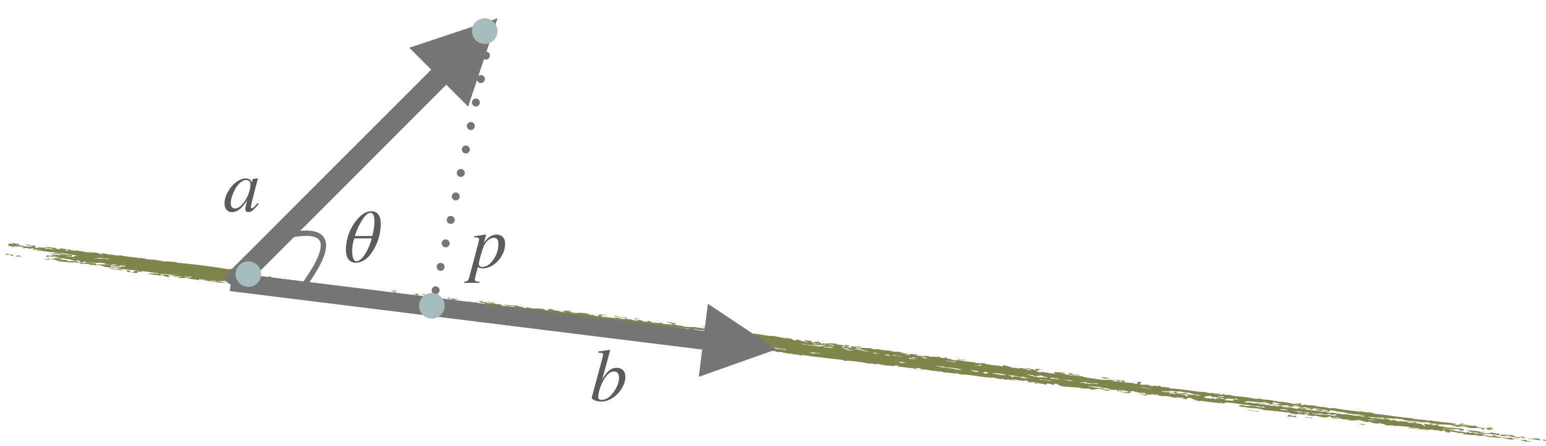
$$cos(\theta_4) \approx 0$$



PUTTING IT TOGETHER

- Application: How far is *a* from the line though *b*?

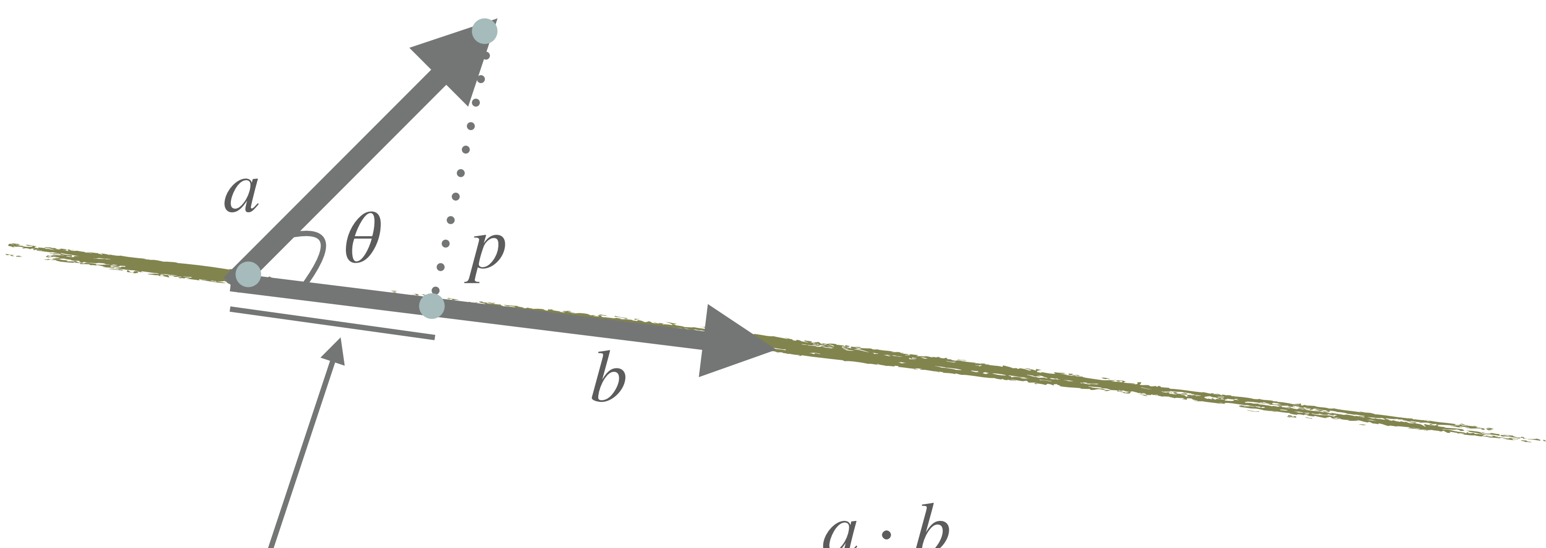
  …How long is the dotted line?



DISTANCE BETWEEN VECTORS

- Application: How far is *a* from the line though *b*?

…How long is the dotted line?



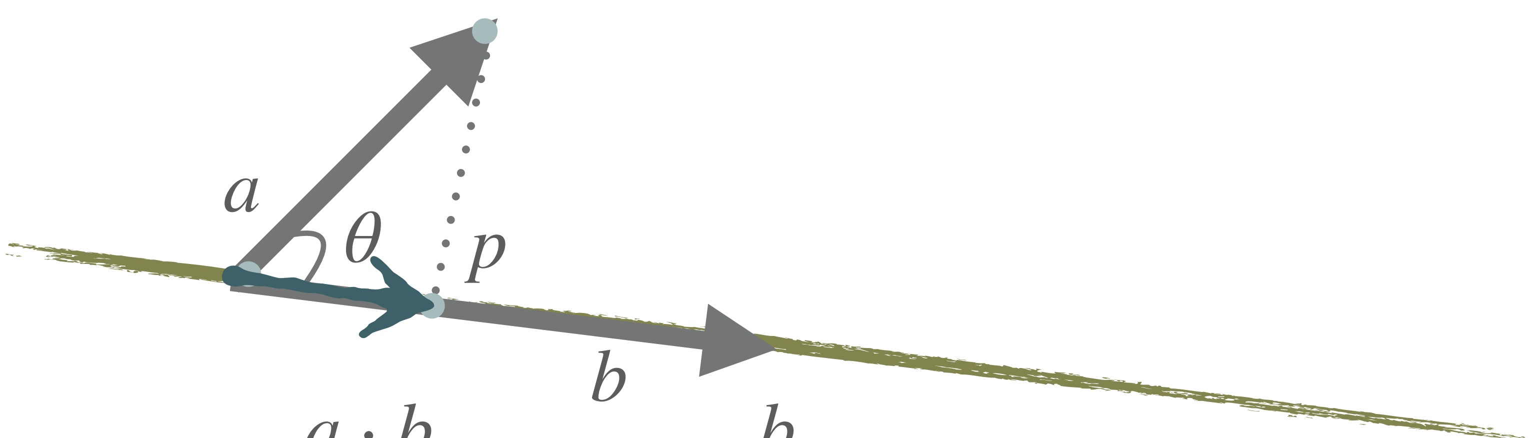
$$a_b = \|a\| \cos(\theta) \;\; = \frac{a \cdot b}{\|b\|}$$

Scalar projection of *a* in direction *b*

DISTANCE BETWEEN VECTORS

- Application: How far is *a* from the line though *b*?

...How long is the dotted line?

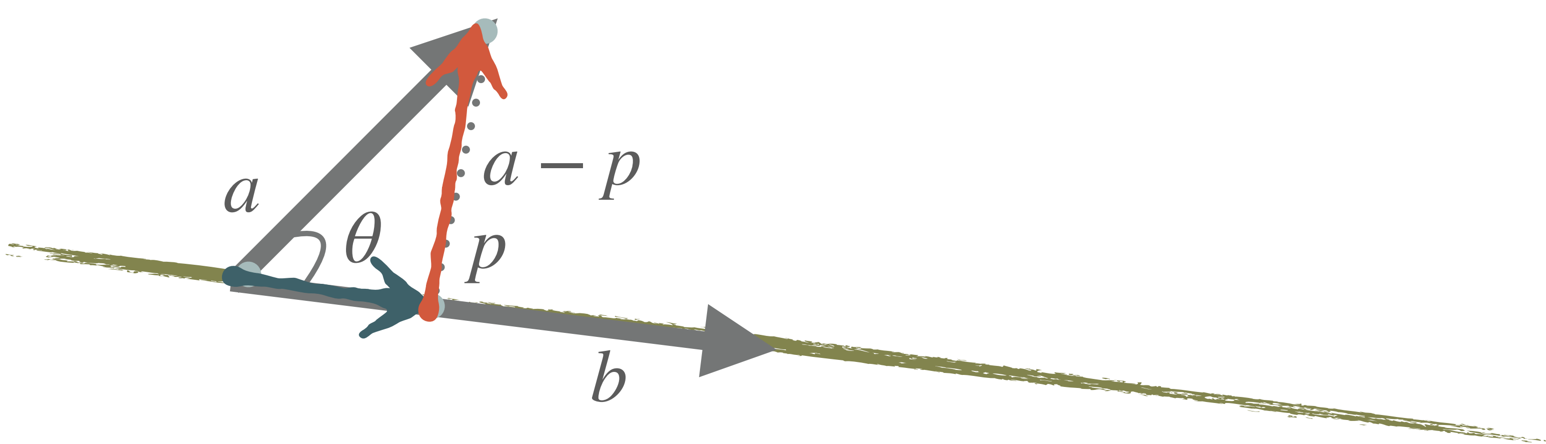


$$p = \frac{a \cdot b}{||b||} b = a \cdot b \frac{b}{||b||}$$

DISTANCE BETWEEN VECTORS

- Application: How far is *a* from the line though *b*?

…How long is the dotted line?

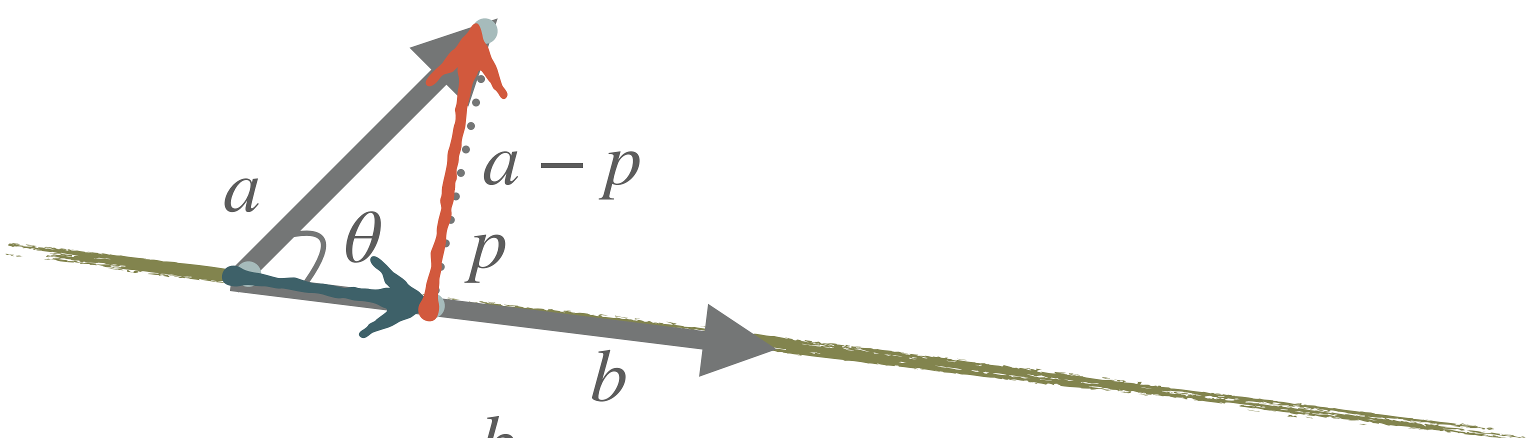


DISTANCE BETWEEN VECTORS

- Application: How far is a from the line though b?

…How long is the dotted line?

Answer: $\|a - p\|_2$



$$p = a \cdot b \frac{b}{\|b\|}$$

Next Week

Data sets, tools, data wrangling

If you have a laptop, you may want to bring it