# APPLIED MACHINE LEARNING

## *Principal Component Analysis (PCA)*

## *Part II – To find the right projections - Intuition*

# Formalism: Projection

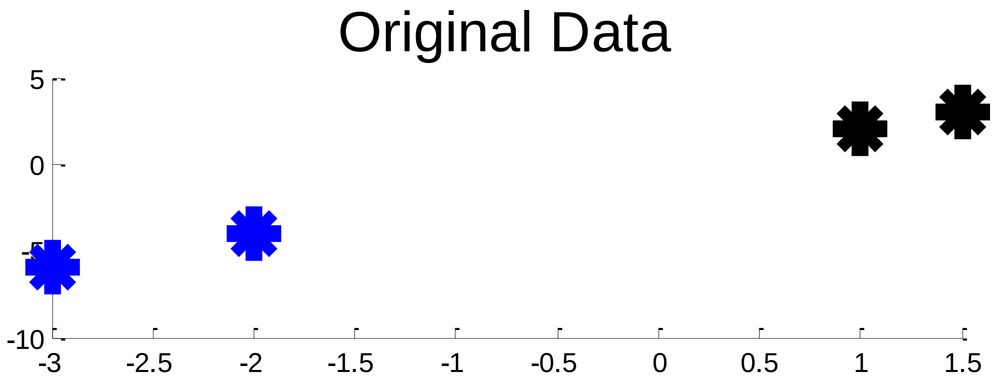
Let $X = [x^1 \ldots x^M]$ a set of $M$ $N$-dim. datapoints, $x^i \in \mathbb{R}^N$, $i = 1 \ldots M$.
A projection $Y$ of $X$ through a linear map $A: x \in \mathbb{R}^N \rightarrow y \in \mathbb{R}^p$, $p \leq N$
is given by: $Y = A \cdot X$.

$X: N \times M$

$A: p \times N$ and $p \leq N$

$Y: p \times M$

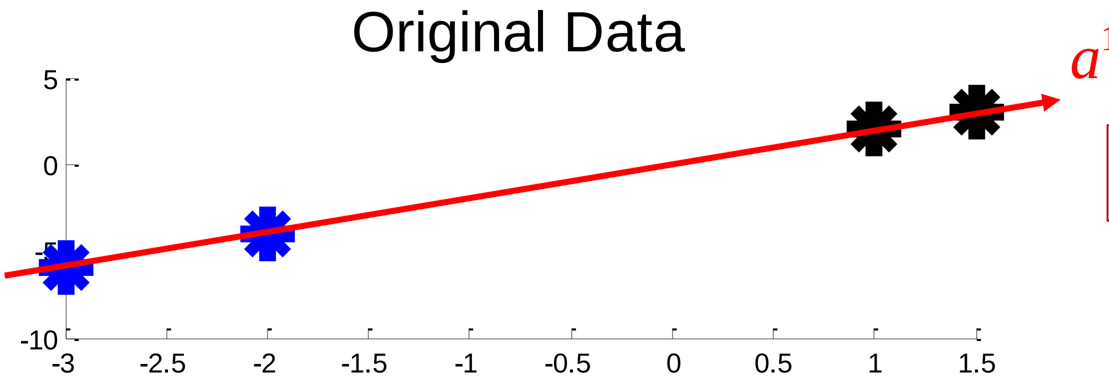
# *Exercise 1 Reducing dimensionality of dataset*




$$\text{Dataset } X = \begin{bmatrix} 1 & -2 & -3 & 1.5 \\ 2 & -4 & -6 & 3 \end{bmatrix}$$

Can you find a way to reduce the amount of information needed to store the coordinates of these 4 datapoints?

# *Exercise 1 Reducing dimensionality of dataset*

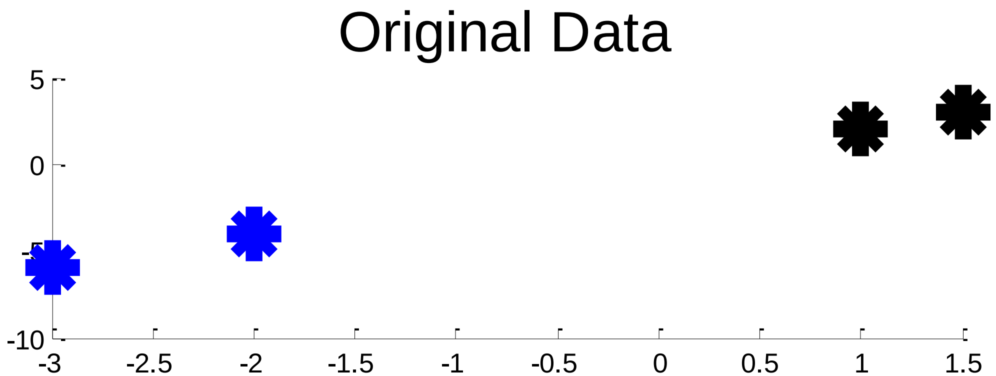



$$A = [a^1]$$

$$\text{Dataset } X = \begin{bmatrix} 1 & -2 & -3 & 1.5 \\ 2 & -4 & -6 & 3 \end{bmatrix}$$

The optimal projection is a line that entails the correlation across the first and second coordinates: $x_2 = 2x_1$

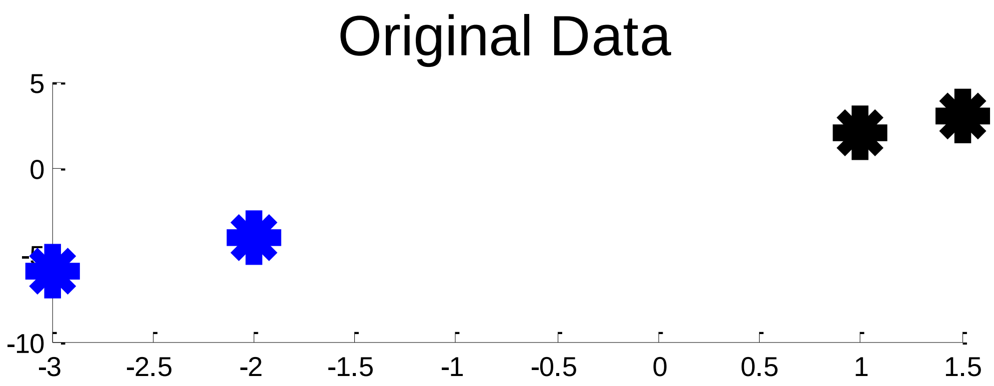
# *Exercise 2 Reducing dimensionality of dataset*




$$\text{Dataset } X = \begin{bmatrix} 1.1 & -2 & -2.9 & 1.5 \\ 2 & -4 & -6.2 & 3 \end{bmatrix}$$

If you use the same solution as before, how much error do you get?

# *Exercise 2 Reducing dimensionality of dataset*




$$\text{Dataset } X = \begin{bmatrix} 1.1 & -2 & -2.9 & 1.5 \\ 2 & -4 & -6.2 & 3 \end{bmatrix}$$

If you use the same solution as before, how much error do you get?

$$\text{Root MSQ: } \underbrace{\begin{bmatrix} 1.1 & -2 & -2.9 & 1.5 \\ 2 & -4 & -6.2 & 3 \end{bmatrix}}_{\text{Real measurement}} \text{ versus } \underbrace{\begin{bmatrix} 1.1 & -2 & -2.9 & 1.5 \\ 2.2 & -4 & -5.8 & 3 \end{bmatrix}}_{\text{Estimate}}$$

*Error* is 0.45.

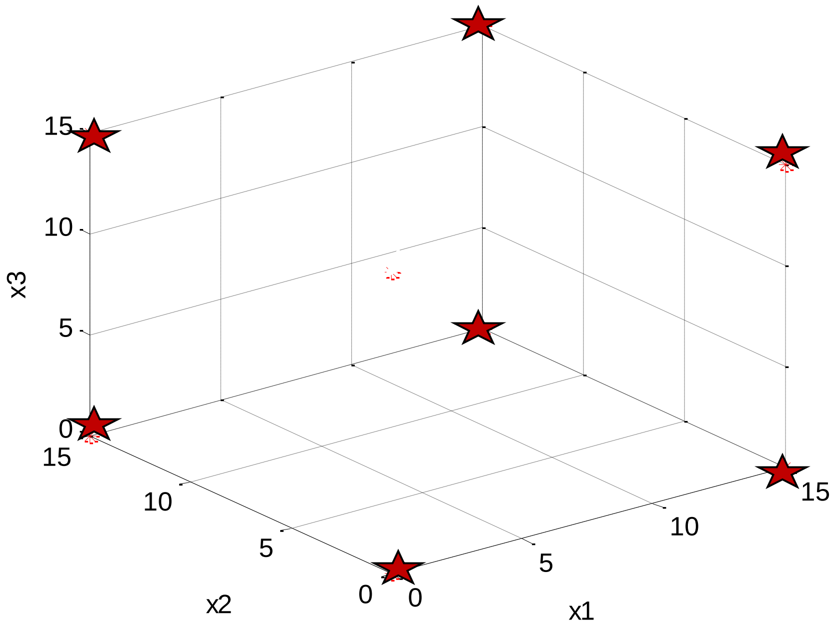
# Constructing a projection: Exercise 3

Example: 2-dimensional projection through a matrix A

$$X = \begin{bmatrix} 0 & 15 & 0 & 15 & 0 & 15 & 0 & 15 \\ 0 & 0 & 15 & 15 & 0 & 0 & 15 & 15 \\ 0 & 0 & 0 & 0 & 15 & 15 & 15 & 15 \end{bmatrix}$$

**Find a matrix A which groups the points into 4 groups.**

**Original data X**

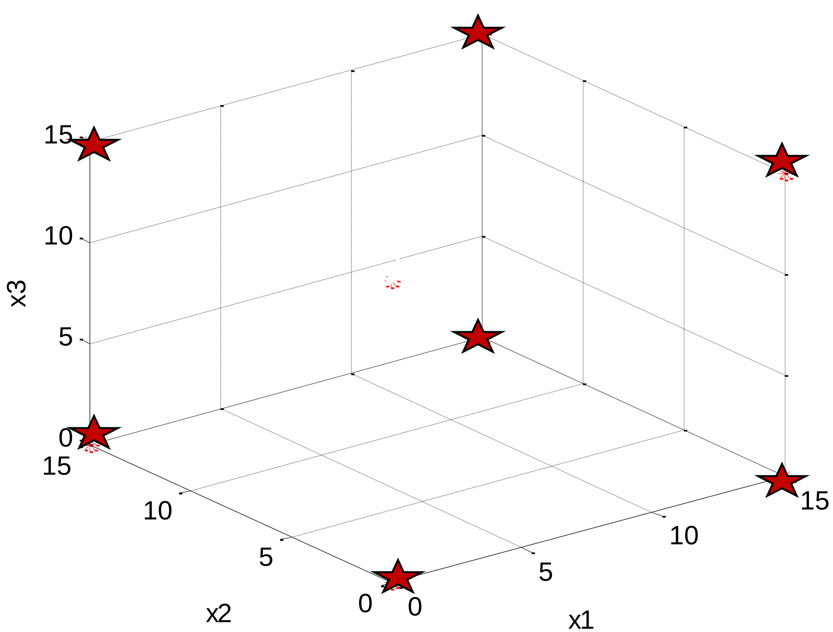


# Constructing a projection: Exercise 3

Example: 2-dimensional projection through a matrix A

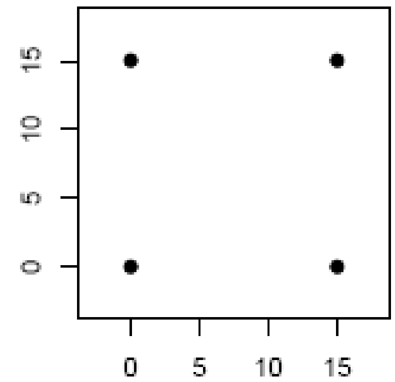
$$X = \begin{bmatrix} 0 & 15 & 0 & 15 & 0 & 15 & 0 & 15 \\ 0 & 0 & 15 & 15 & 0 & 0 & 15 & 15 \\ 0 & 0 & 0 & 0 & 15 & 15 & 15 & 15 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

**Original data X**




$$Y = AX = \begin{bmatrix} 0 & 0 & 15 & 15 & 0 & 0 & 15 & 15 \\ 0 & 15 & 0 & 15 & 0 & 15 & 0 & 15 \end{bmatrix}$$

**Projected data Y**




The data are well grouped into 4 tiny clusters

# Constructing a projection

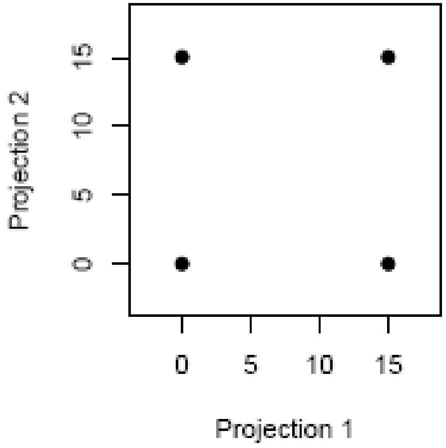
Example: 2-dimensional projection through a matrix A

$$X = \begin{bmatrix} 0 & 15 & 0 & 15 & 0 & 15 & 0 & 15 \\ 0 & 0 & 15 & 15 & 0 & 0 & 15 & 15 \\ 0 & 0 & 0 & 0 & 15 & 15 & 15 & 15 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{matrix} a^1 \\ a^2 \end{matrix}$$

The rows of $A$ are composed of two orthonormal vectors: $(a^1, a^2)$.

The product of each $a^j$ with each datapoint $x^i$ corresponds to the coordinate of the image $y^i$ of the point $x^i$ in the projected space.



# Constructing a projection

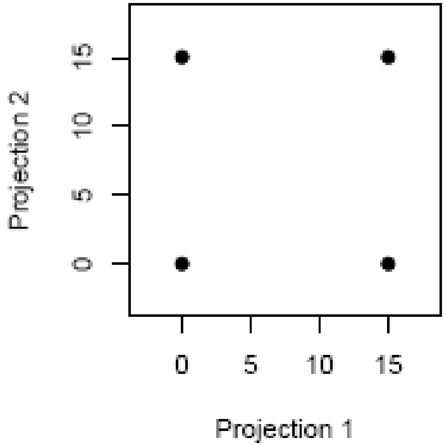
Example: 2-dimensional projection through a matrix A

$$X = \begin{bmatrix} 0 & 15 & 0 & 15 & 0 & 15 & 0 & 15 \\ 0 & 0 & 15 & 15 & 0 & 0 & 15 & 15 \\ 0 & 0 & 0 & 0 & 15 & 15 & 15 & 15 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Projections 1 & 2

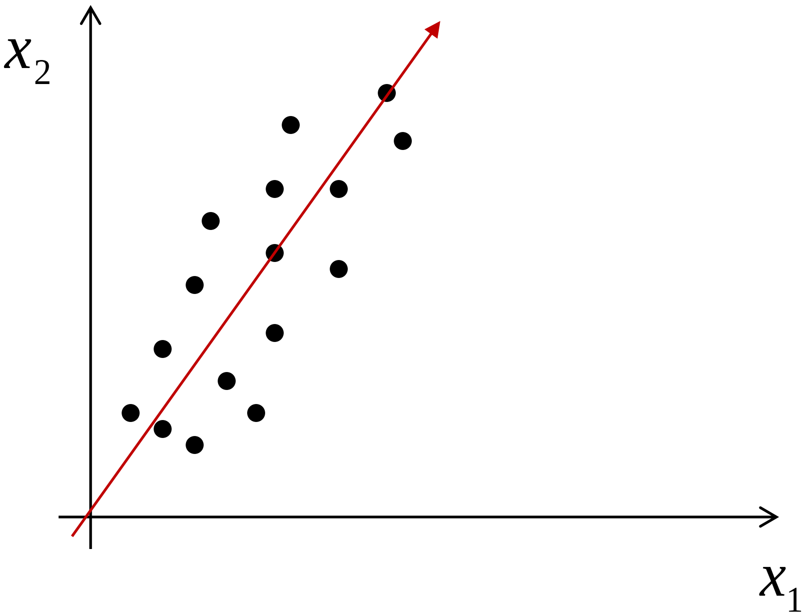
The columns of $A$ represent the images (in the projected space) of the axes of the original space ($\mathbb{R}^3$); in the example, the first two columns form a basis of $\mathbb{R}^2$.



# *PCA: Reduction of dimensionality*

Infinite number of choices for projecting the data
→ need criteria to reduce the choice

1: minimum information loss (minimal reconstruction error)

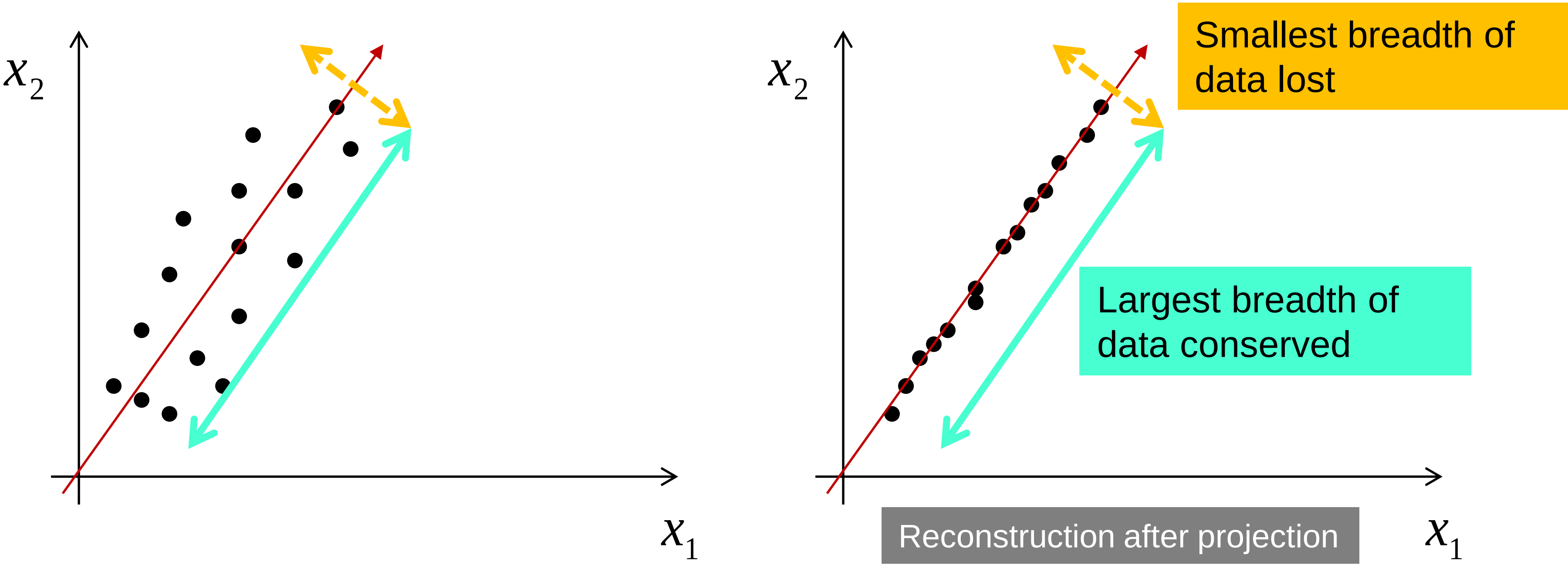



What is the 2D to 1D projection that minimizes the reconstruction error?

# *PCA: Reduction of dimensionality*

Infinite number of choices for projecting the data
→ need criteria to reduce the choice

1: minimum information loss(minimal reconstruction error)
2: equivalent to finding the direction with maximum variance




What is the 2D to 1D projection that minimizes the reconstruction error?