PCA

MLT

Karthik Thiagarajan

Learning Representations

"Learning from data"

Learning Representations

"Learning from data"

Documentary

- Collect Data
  - Interview people
    * 50 people
    * 10 hourss

Learning Representations

Documentary

- Collect Data
  - Interview people
    * 50 people
    * 10 hourss
- Create documentary
  - 2 hours

Learning Representations

Documentary

- Collect Data
  - Interview people
    - 50 people
    - 10 hourss
- Create documentary
  - 2 hours
  - Editing

Learning Representations

Documentary

- Collect Data
  - Interview people
    - * 50 people
    - * 10 hourss

- Create documentary
  - 2 hours
  - Editing
    - * choose people
    - * choose key moments
    - * weave it into a narrative
    - * capture all important dimensions
    - * "essence"

Learning Representations

"Learning from data"

Documentary

- Collect Data
  - Interview people
    * 50 people
    * 10 hourss

- Create documentary
  - 2 hours
  - Editing
    * choose people
    * choose key moments
    * weave it into a narrative
    * capture all important dimensions
    * "essence"

Themes

- Compression
- Representation
- Reconstruction

Learning Representations

"Learning from data"

Documentary

- Collect Data
  - Interview people
    * 50 people
    * 10 hourss
- Create documentary
  - 2 hours
  - Editing
    * choose people
    * choose key moments
    * weave it into a narrative
    * capture all important dimensions
    * "essence"

$$\{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \xrightarrow[\text{editing}]{\text{PCA}} \{\mathbf{x}_1^{(r)}, \cdots, \mathbf{x}_n^{(r)}\}$$

Themes

- Compression
- Representation
- Reconstruction

Good Representations

$$\{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \rightarrow \{\mathbf{x}_1^{(r)}, \cdots, \mathbf{x}_n^{(r)}\}$$

Good Representations

$$\{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \rightarrow \{\mathbf{x}_1^{(r)}, \cdots, \mathbf{x}_n^{(r)}\}$$

$$\frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

Good Representations

$$\{\mathbf{x}_1, \; \cdots, \mathbf{x}_n\} \rightarrow \{\mathbf{x}_1^{(r)}, \; \cdots, \mathbf{x}_n^{(r)}\}$$

$$\min_{\left\{\mathbf{x}_i^{(r)}\right\}_{i=1}^{n}} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

$$\frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

Good Representations

$$\{\mathbf{x}_1,\ \cdots,\mathbf{x}_n\} \rightarrow \{\mathbf{x}_1^{(r)},\ \cdots,\mathbf{x}_n^{(r)}\}$$

$$\min_{\left\{\mathbf{x}_i^{(r)}\right\}_{i=1}^{n}} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

$$\frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

$$\mathbf{x}_i^{(r)} = \mathbf{x}_i$$

Good Representations

$$\{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \to \{\mathbf{x}_1^{(r)}, \cdots, \mathbf{x}_n^{(r)}\}$$

$$\min_{\left\{\mathbf{x}_i^{(r)}\right\}_{i=1}^{n}} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

$$\frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

$$\mathbf{x}_i^{(r)} = \mathbf{x}_i$$

Constraint: Compression

Good Representations

"Freedom without discipline is chaos"

$$\{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \to \{\mathbf{x}_1^{(r)}, \cdots, \mathbf{x}_n^{(r)}\}$$

$$\min_{\left\{\mathbf{x}_i^{(r)}\right\}_{i=1}^{n}} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

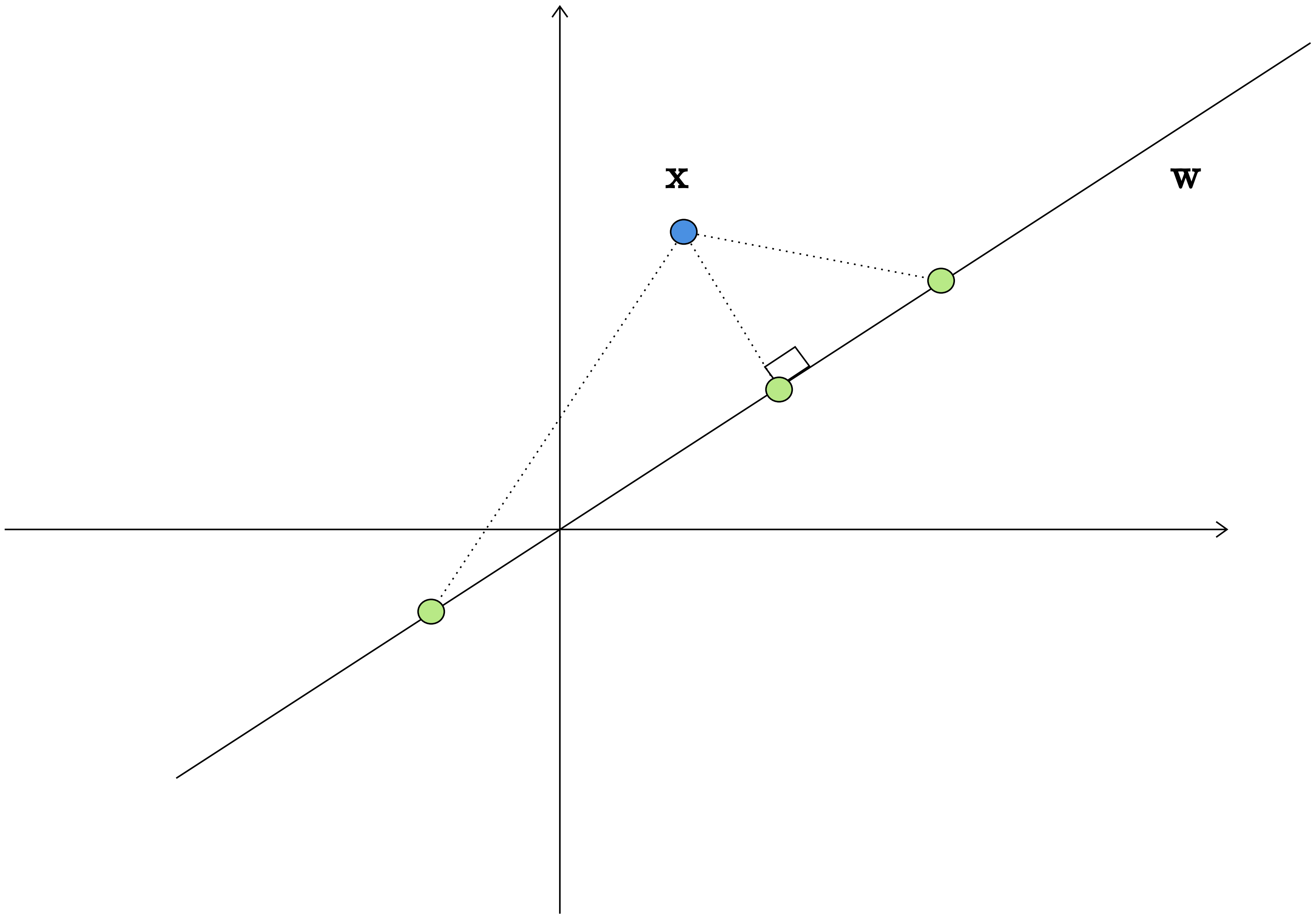
$$\frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{x}_i^{(r)}||^2$$

$$\mathbf{x}_i^{(r)} = \mathbf{x}_i$$

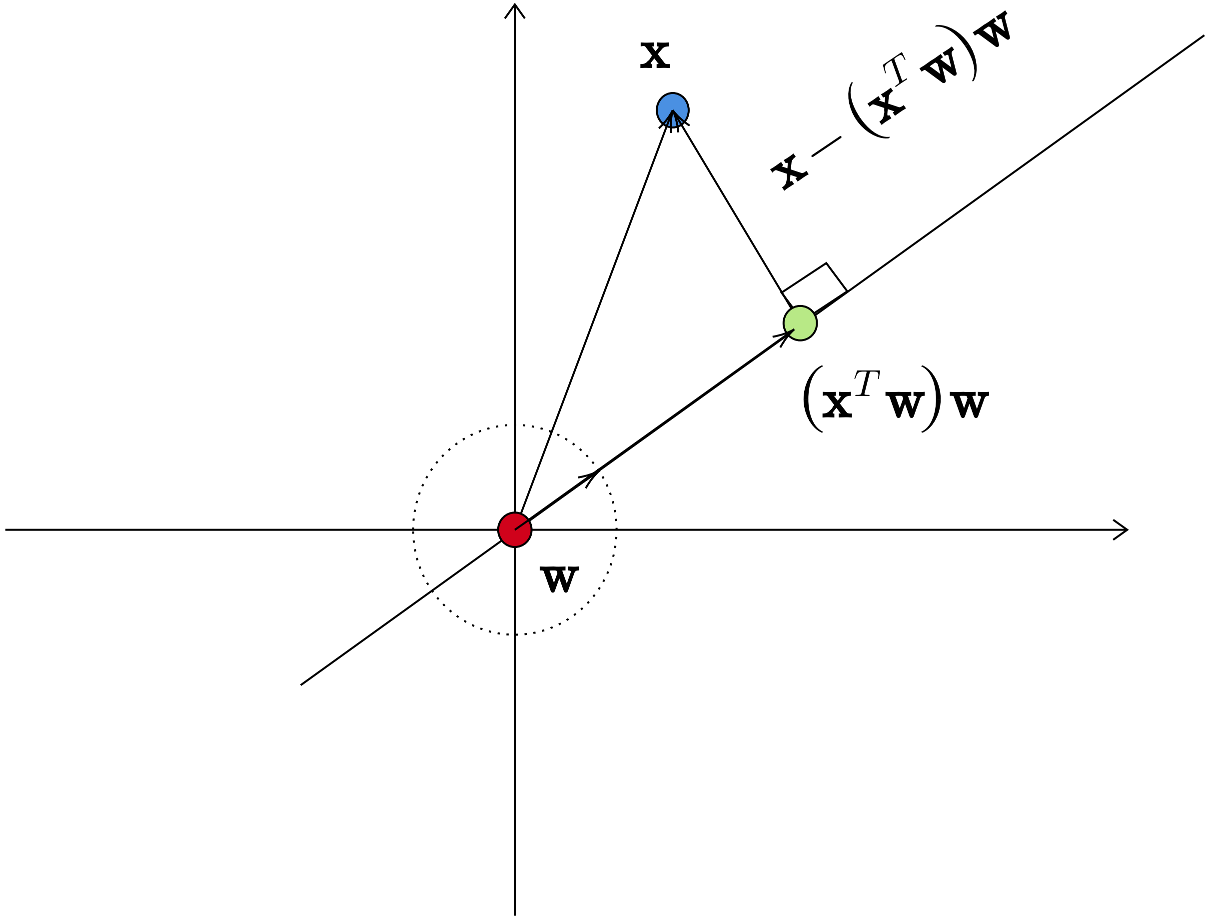
Constraint: Compression

$$\mathbf{x}_i^{(r)} = \alpha_i \mathbf{w}$$

Best Proxy



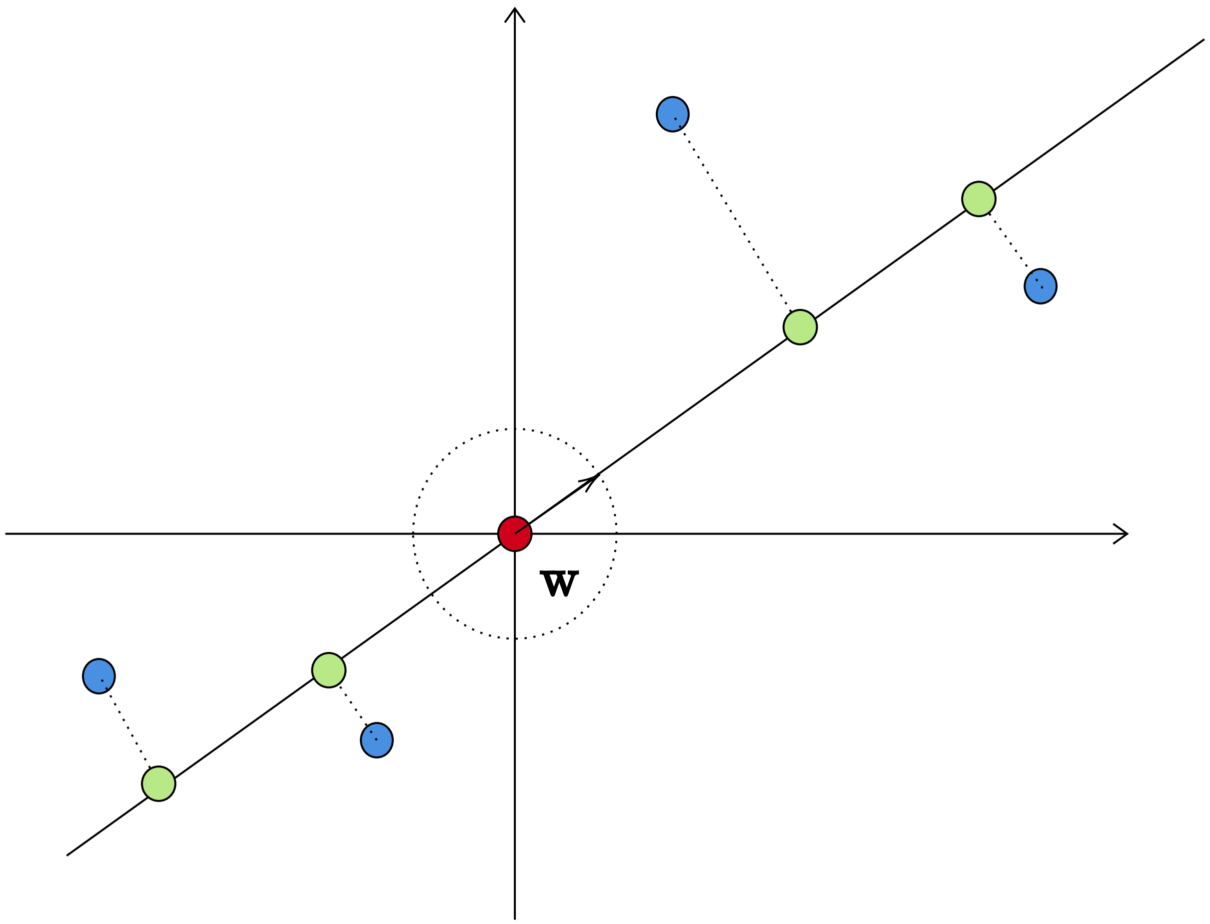
Best Proxy



$$||\mathbf{w}|| = 1$$

$$||\mathbf{x} - (\mathbf{x}^T\mathbf{w})\mathbf{w}||^2$$

Average Error

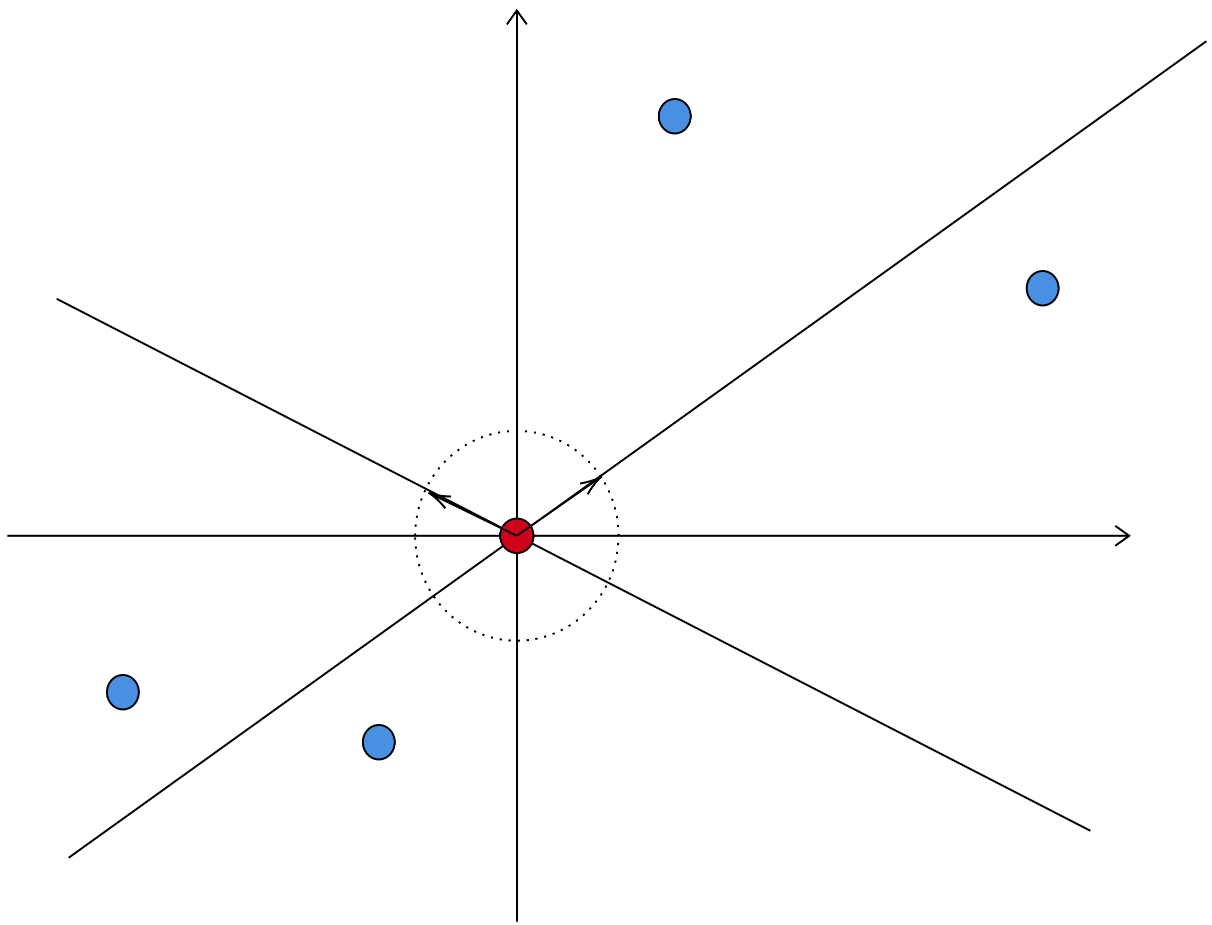


$$||\mathbf{w}|| = 1$$

Average Error

$$\frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

Error Minimization



$$||\mathbf{w}|| = 1$$

$$\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

Error Minimization (algebra)

$$\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

$$||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2 =$$

Error Minimization (algebra)

$$\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

$$||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2 = \Big(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\Big)^T \Big(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\Big)$$

$$||\mathbf{v}||^2 = \mathbf{v}^T \mathbf{v}$$

Error Minimization (algebra)

$$\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

$$||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2 = \Big(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\Big)^T \Big(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\Big)$$

$$||\mathbf{v}||^2 = \mathbf{v}^T \mathbf{v}$$

$$= \mathbf{x}_i^T \mathbf{x}_i - \Big(\mathbf{x}_i^T \mathbf{w}\Big)\Big(\mathbf{x}_i^T \mathbf{w}\Big) - \Big(\mathbf{x}_i^T \mathbf{w}\Big)\Big(\mathbf{w}^T \mathbf{x}_i\Big) + \Big(\mathbf{x}_i^T \mathbf{w}\Big)^2 \Big(\mathbf{w}^T \mathbf{w}\Big)$$

Error Minimization (algebra)

$$\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{w})\mathbf{w}||^2$$

$$||\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{w})\mathbf{w}||^2 = \left(\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{w})\mathbf{w}\right)^T \left(\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{w})\mathbf{w}\right)$$

$$= \mathbf{x}_i^T\mathbf{x}_i - \left(\mathbf{x}_i^T\mathbf{w}\right)\left(\mathbf{x}_i^T\mathbf{w}\right) - \left(\mathbf{x}_i^T\mathbf{w}\right)\left(\mathbf{w}^T\mathbf{x}_i\right) + \left(\mathbf{x}_i^T\mathbf{w}\right)^2\left(\mathbf{w}^T\mathbf{w}\right)$$

$$= \mathbf{x}_i^T\mathbf{x}_i - \left(\mathbf{x}_i^T\mathbf{w}\right)^2$$

$$||\mathbf{v}||^2 = \mathbf{v}^T\mathbf{v}$$
$$\mathbf{x}_i^T\mathbf{w} = \mathbf{w}^T\mathbf{x}_i$$
$$\mathbf{w}^T\mathbf{w} = 1$$

Error Minimization (algebra)

$$\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

$$||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2 = \Big(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\Big)^T \Big(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\Big)$$

$$||\mathbf{v}||^2 = \mathbf{v}^T \mathbf{v}$$
$$\mathbf{x}_i^T \mathbf{w} = \mathbf{w}^T \mathbf{x}_i$$
$$\mathbf{w}^T \mathbf{w} = 1$$

$$= \mathbf{x}_i^T \mathbf{x}_i - \Big(\mathbf{x}_i^T \mathbf{w}\Big)\Big(\mathbf{x}_i^T \mathbf{w}\Big) - \Big(\mathbf{x}_i^T \mathbf{w}\Big)\Big(\mathbf{w}^T \mathbf{x}_i\Big) + \Big(\mathbf{x}_i^T \mathbf{w}\Big)^2 \Big(\mathbf{w}^T \mathbf{w}\Big)$$

$$= \mathbf{x}_i^T \mathbf{x}_i - \Big(\mathbf{x}_i^T \mathbf{w}\Big)^2$$

$$\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

Error Minimization (algebra)

$$\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

$$||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2 = \Big(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\Big)^T \Big(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\Big)$$

$$= \mathbf{x}_i^T \mathbf{x}_i - \Big(\mathbf{x}_i^T \mathbf{w}\Big)\Big(\mathbf{x}_i^T \mathbf{w}\Big) - \Big(\mathbf{x}_i^T \mathbf{w}\Big)\Big(\mathbf{w}^T \mathbf{x}_i\Big) + \Big(\mathbf{x}_i^T \mathbf{w}\Big)^2 \Big(\mathbf{w}^T \mathbf{w}\Big)$$

$$= \mathbf{x}_i^T \mathbf{x}_i - \Big(\mathbf{x}_i^T \mathbf{w}\Big)^2$$

$$||\mathbf{v}||^2 = \mathbf{v}^T \mathbf{v}$$
$$\mathbf{x}_i^T \mathbf{w} = \mathbf{w}^T \mathbf{x}_i$$
$$\mathbf{w}^T \mathbf{w} = 1$$

$$\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2 \;\equiv\; \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \Big(\mathbf{x}_i^T \mathbf{w}\Big)^2$$

Error Minimization (algebra)

$$\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2$$

$$||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2 = \left(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\right)^T \left(\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\right)$$

$$||\mathbf{v}||^2 = \mathbf{v}^T \mathbf{v}$$
$$\mathbf{x}_i^T \mathbf{w} = \mathbf{w}^T \mathbf{x}_i$$
$$\mathbf{w}^T \mathbf{w} = 1$$

$$= \mathbf{x}_i^T \mathbf{x}_i - \left(\mathbf{x}_i^T \mathbf{w}\right)\left(\mathbf{x}_i^T \mathbf{w}\right) - \left(\mathbf{x}_i^T \mathbf{w}\right)\left(\mathbf{w}^T \mathbf{x}_i\right) + \left(\mathbf{x}_i^T \mathbf{w}\right)^2 \left(\mathbf{w}^T \mathbf{w}\right)$$

$$= \mathbf{x}_i^T \mathbf{x}_i - \left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

$$\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2 \equiv \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \left(\mathbf{x}_i^T \mathbf{w}\right)^2 \equiv \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} -\left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

Journey from $\min$ to $\max$

$$\boxed{\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2}$$

Journey from $\min$ to $\max$

$$\boxed{\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2} \equiv \min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

Journey from $\min$ to $\max$

$$\boxed{\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2} \equiv \min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

$$\equiv \min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} - \left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

Journey from $\min$ to $\max$

$$\boxed{\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}||\mathbf{x}_i-(\mathbf{x}_i^T\mathbf{w})\mathbf{w}||^2} \equiv \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i-\left(\mathbf{x}_i^T\mathbf{w}\right)^2$$

$$\equiv \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}-\left(\mathbf{x}_i^T\mathbf{w}\right)^2$$

$$\equiv \max_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{w}\right)^2$$

Journey from $\min$ to $\max$

$$\boxed{\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}||\mathbf{x}_i-(\mathbf{x}_i^T\mathbf{w})\mathbf{w}||^2} \equiv \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i-\left(\mathbf{x}_i^T\mathbf{w}\right)^2$$

$$\equiv \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}-\left(\mathbf{x}_i^T\mathbf{w}\right)^2$$

$$\equiv \max_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{w}\right)^2$$

$$\equiv \max_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n}\cdot\sum_{i=1}^{n}\left(\mathbf{w}^T\mathbf{x}_i\right)\left(\mathbf{x}_i^T\mathbf{w}\right)$$

Journey from $\min$ to $\max$

$$
\begin{aligned}
\boxed{\min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2} &\equiv \min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \left(\mathbf{x}_i^T \mathbf{w}\right)^2 \\
&\equiv \min_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} -\left(\mathbf{x}_i^T \mathbf{w}\right)^2 \\
&\equiv \max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{w}\right)^2 \\
&\equiv \max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathbf{w}^T \mathbf{x}_i\right)\left(\mathbf{x}_i^T \mathbf{w}\right) \\
&\equiv \max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \left[\frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right] \mathbf{w}
\end{aligned}
$$

Journey from $\min$ to $\max$

$$\boxed{\min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}||^2} \equiv \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

$$\equiv \min_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} -\left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

$$\equiv \max_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

Covariance Matrix
(Centered Dataset)

$$\equiv \max_{\mathbf{w},||\mathbf{w}||=1} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathbf{w}^T \mathbf{x}_i\right)\left(\mathbf{x}_i^T \mathbf{w}\right)$$

$$\mathbf{C} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$$

$$\equiv \max_{\mathbf{w},||\mathbf{w}||=1} \quad \mathbf{w}^T \left[\frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right] \mathbf{w}$$

$$\equiv \boxed{\max_{\mathbf{w},||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w}}$$

Optimization Problem

$$\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w}$$

$$\arg\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w}$$

Optimization Problem

$$\max_{\mathbf{w}, ||\mathbf{w}||=1} \mathbf{w}^T \mathbf{C} \mathbf{w} = \lambda_1 \qquad \arg\max_{\mathbf{w}, ||\mathbf{w}||=1} \mathbf{w}^T \mathbf{C} \mathbf{w} = \mathbf{w}_1$$

Optimization Problem

$$\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w} = \lambda_1$$

$$\arg\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w} = \mathbf{w}_1$$

$$\mathbf{C} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$$

$$\lambda_1 \geq \quad \cdots \quad \geq \lambda_d \geq 0$$

$$\mathbf{w}_1, \ \cdots, \mathbf{w}_d$$

Optimization Problem

$$\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w} = \lambda_1$$

$$\arg\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w} = \mathbf{w}_1$$

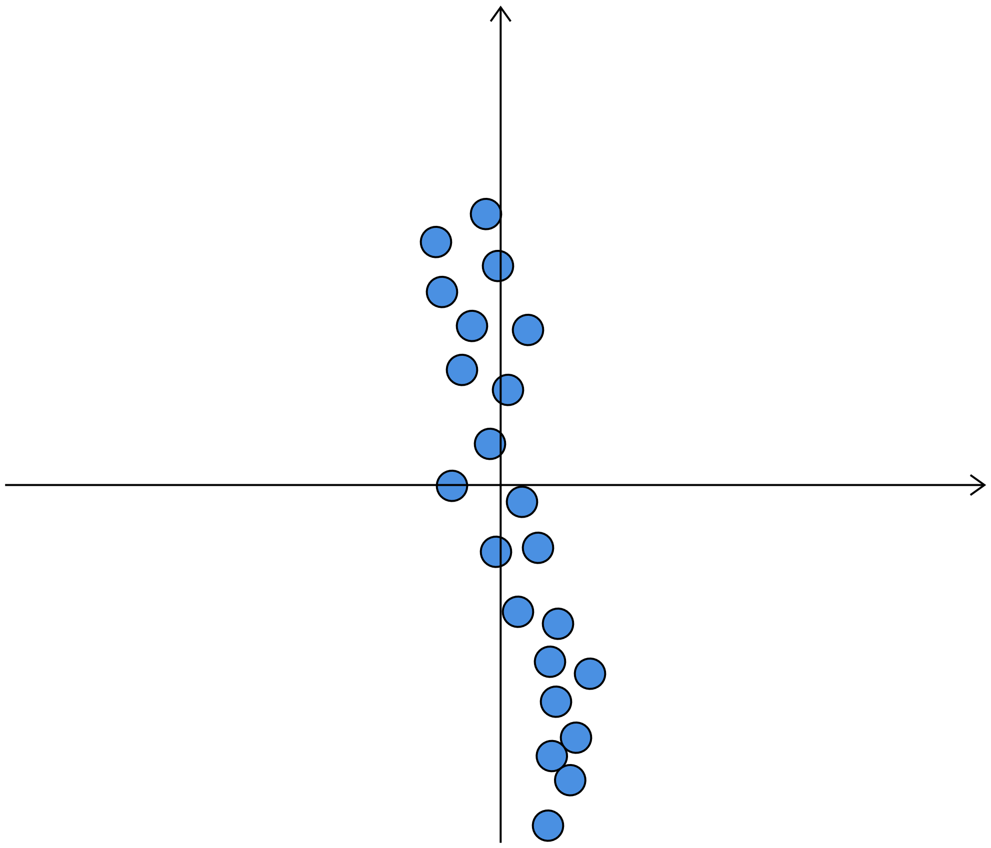
$$\mathbf{C} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$$

$$\lambda_1 \geq \quad \cdots \quad \geq \lambda_d \geq 0$$

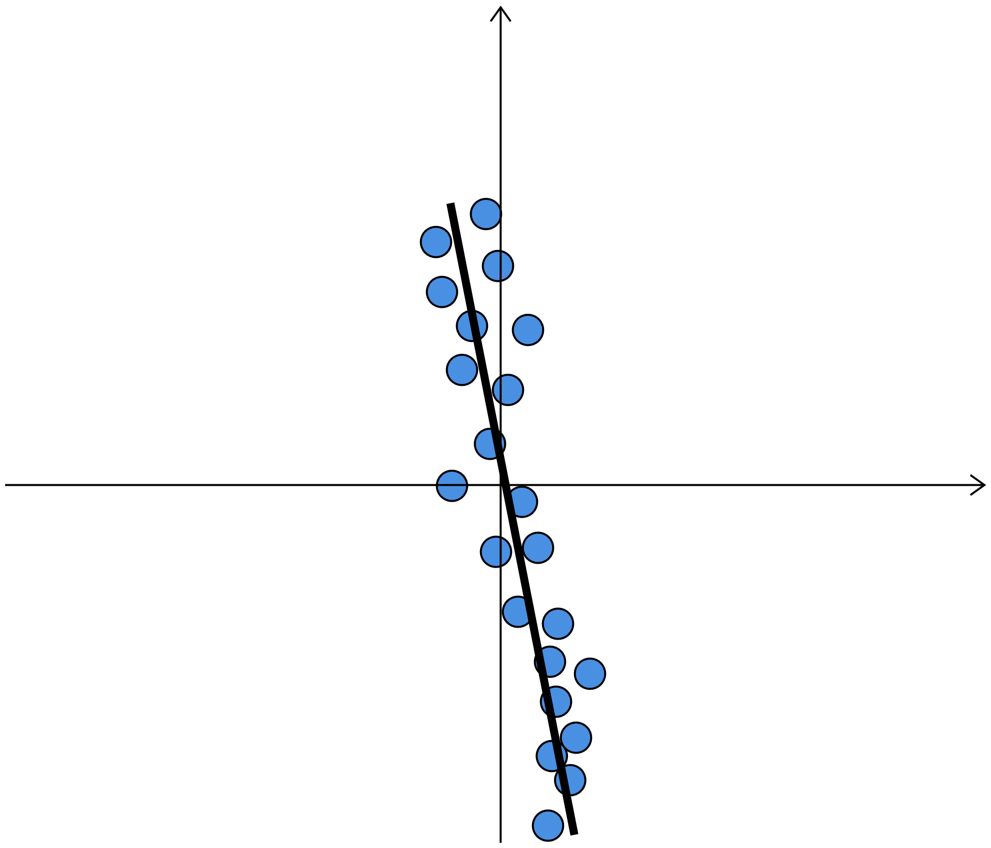
$$\mathbf{w}_1, \ \cdots, \mathbf{w}_d$$

The direction that minimizes the reconstruction error is the eigenvector corresponding to the largest eigenvalue of the covariance matrix.

First Principal Component



First Principal Component



First Principal Component

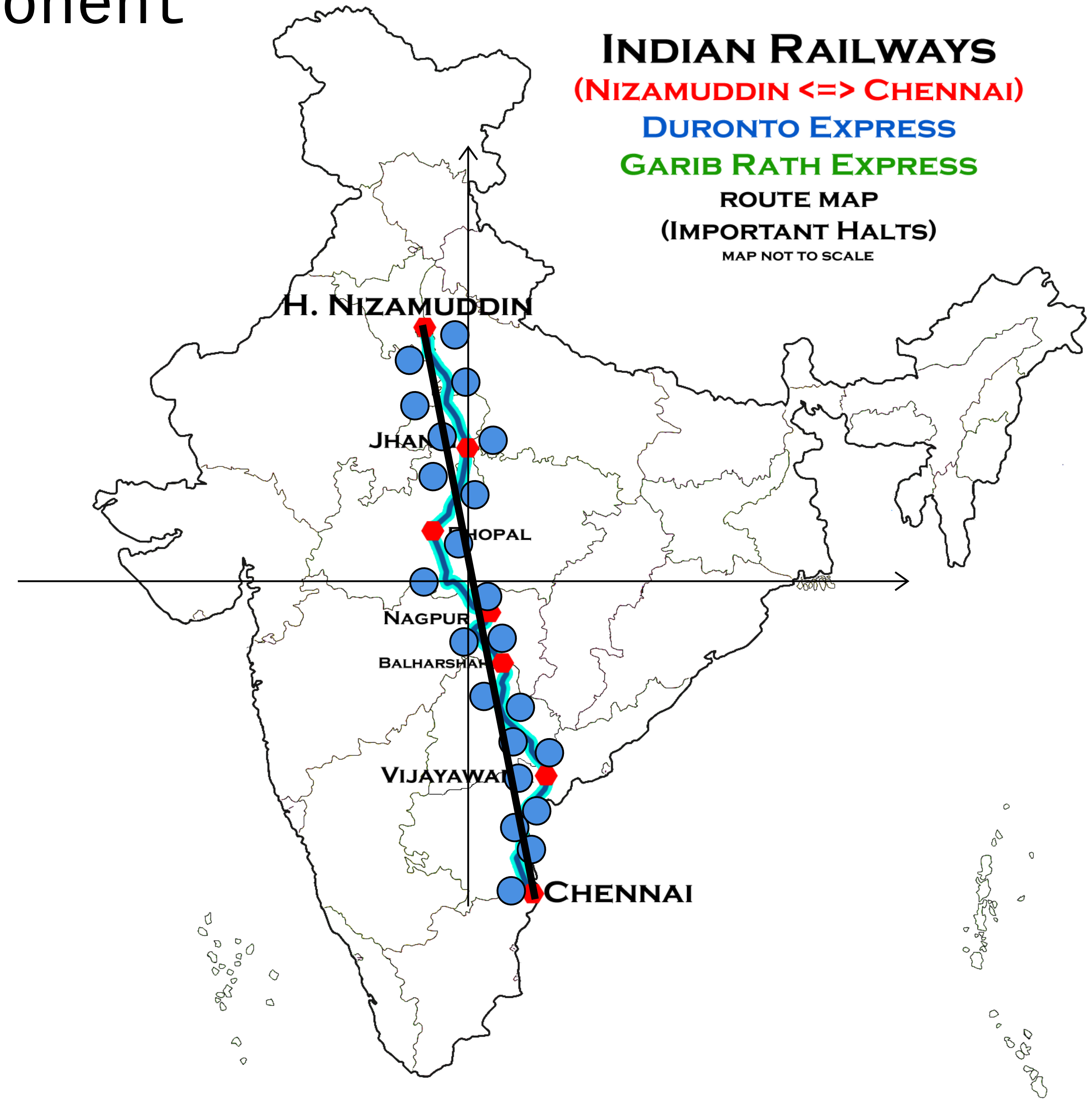


Image Credits

First Principal Component

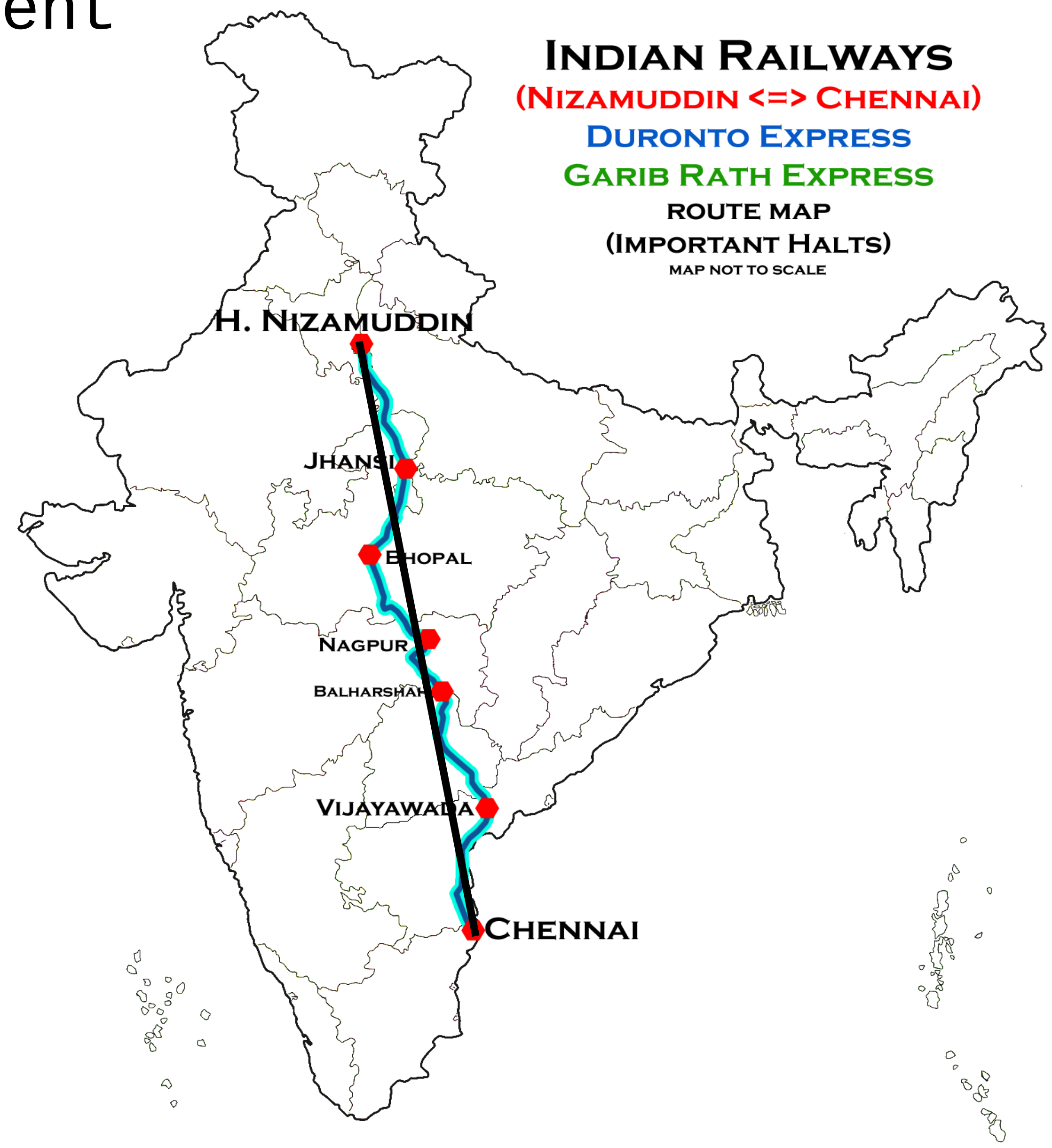


Image Credits

First Principal Component

Chennai-Delhi Garib Rath

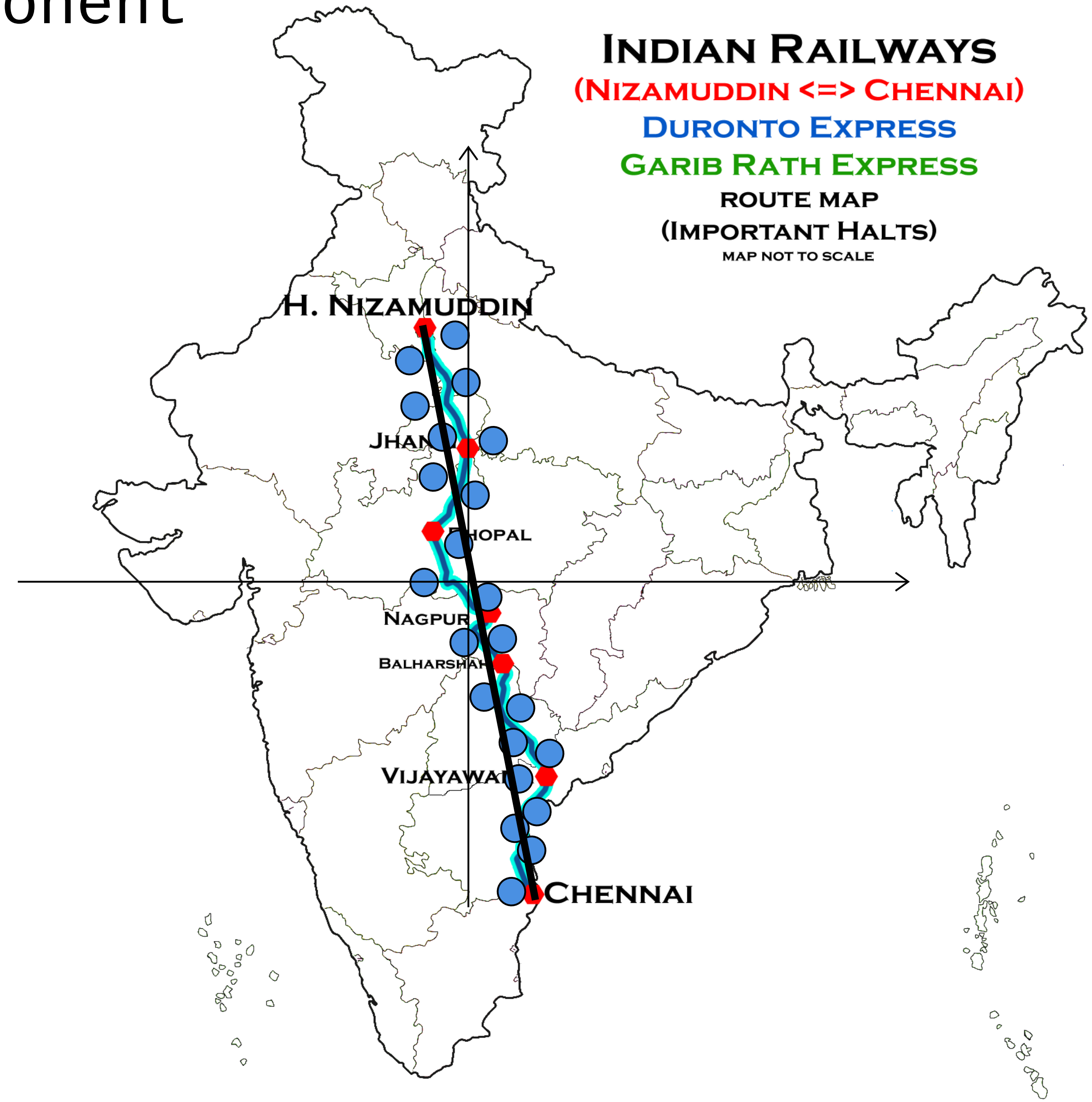


Image Credits

First Principal Component

Chennai-Delhi Garib Rath

- Data
  - matrix
  - abstract

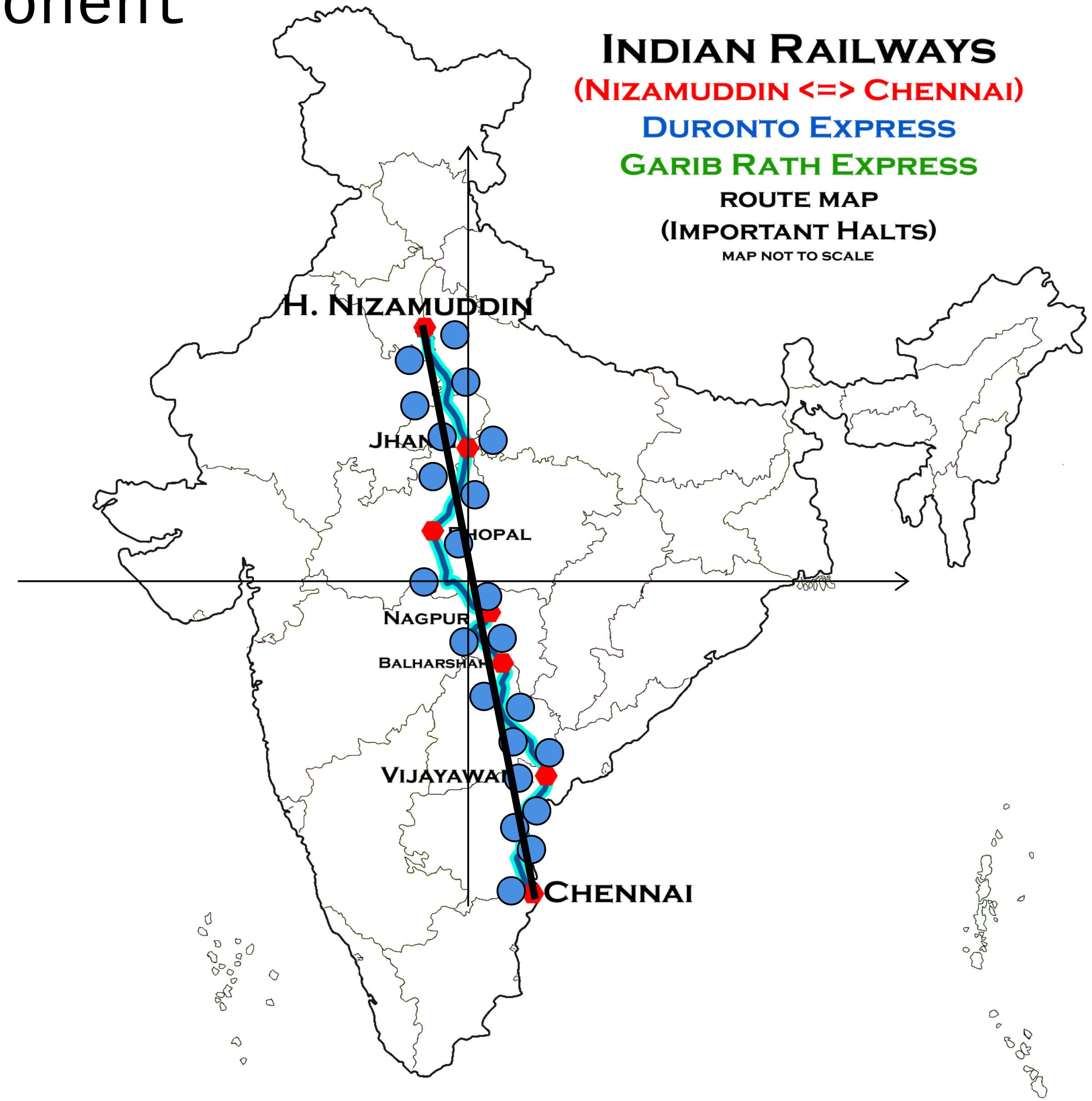


Image Credits

First Principal Component

Chennai-Delhi Garib Rath

- Data
  - matrix
  - abstract
- **Algorithm**
  - general
  - powerful

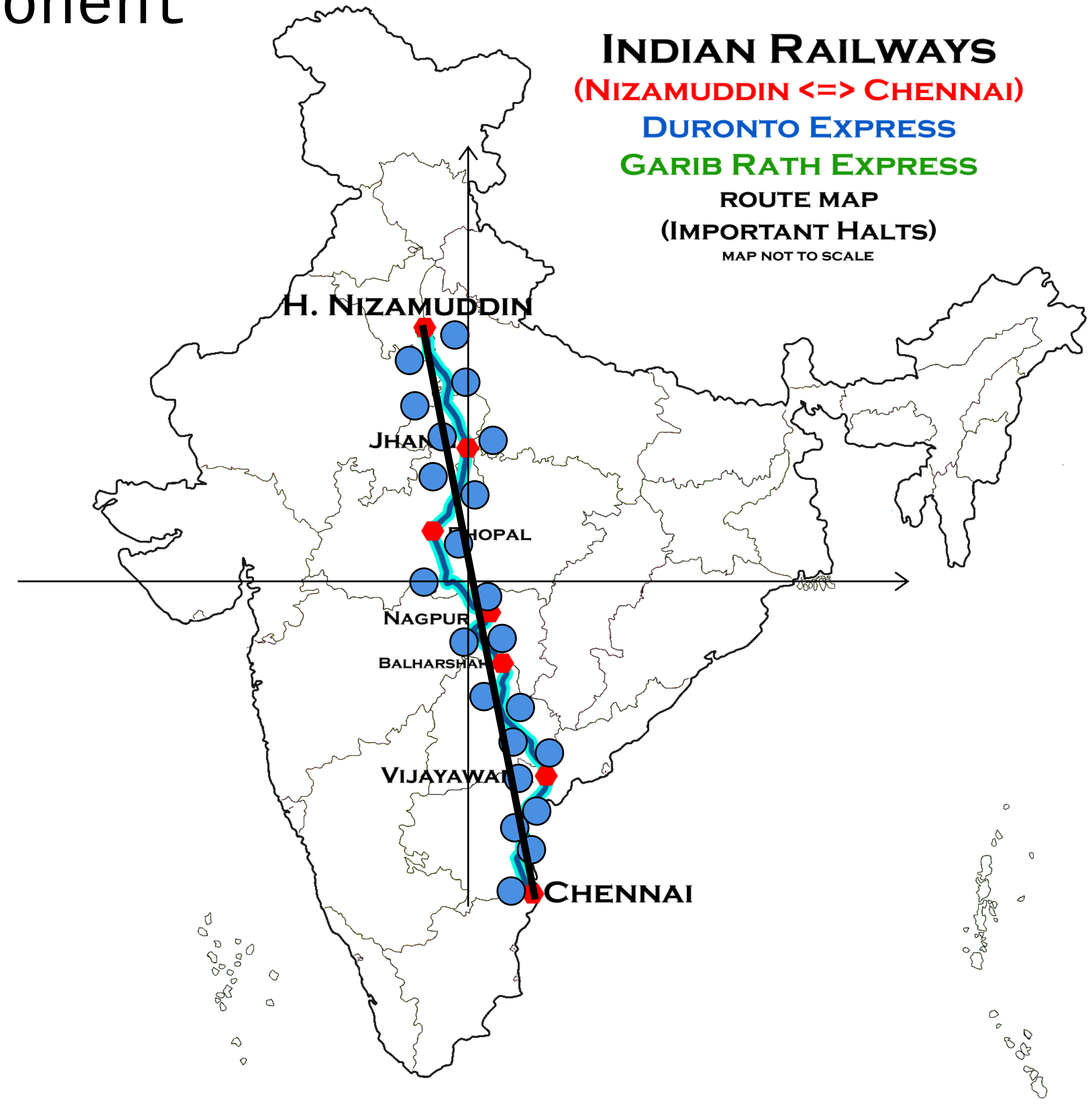


Image Credits

First Principal Component

Chennai-Delhi Garib Rath

- Data
  - matrix
  - abstract
- **Algorithm**
  - general
  - powerful
- Domain knowledge

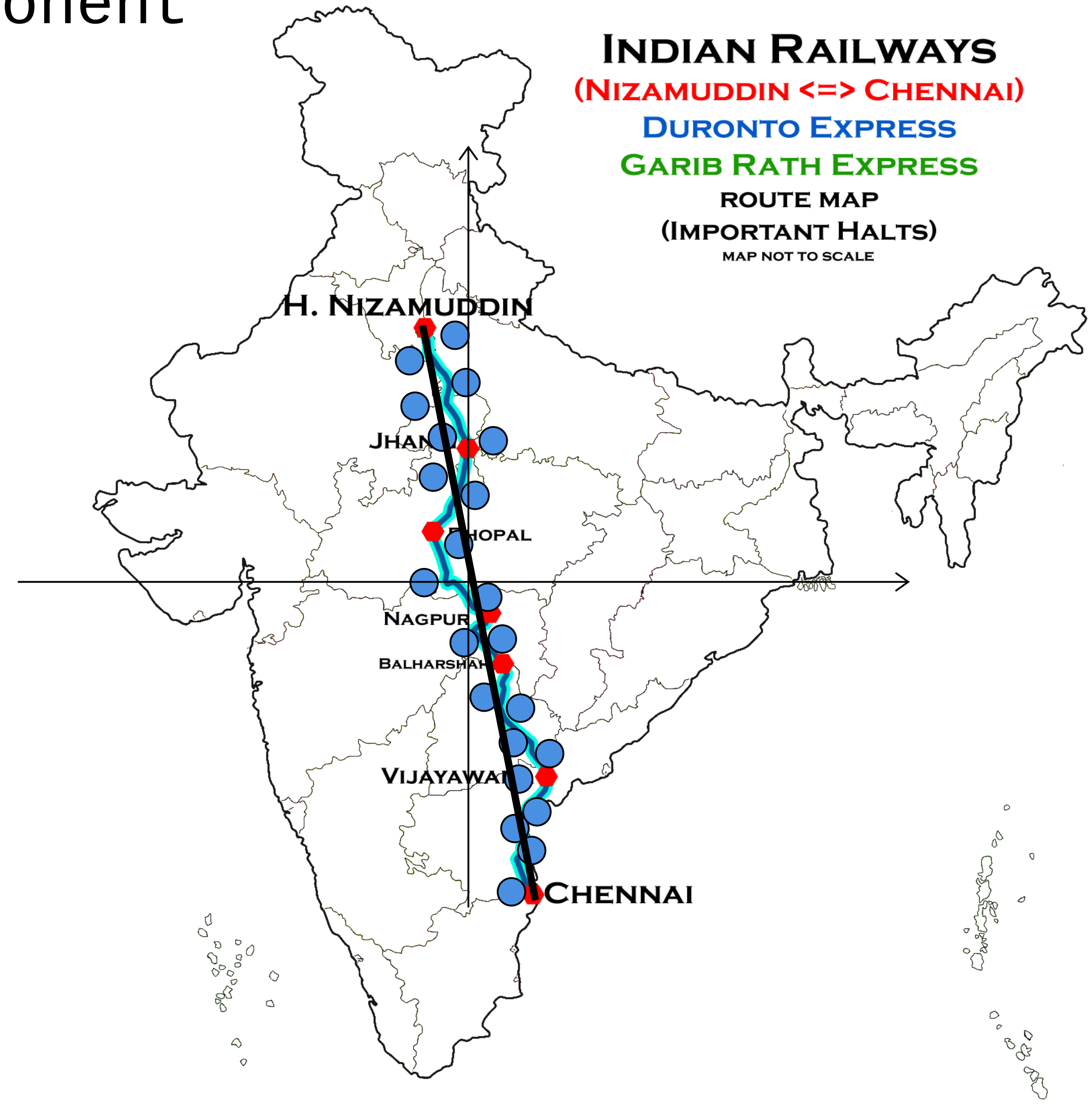
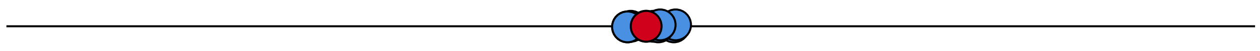


Image Credits

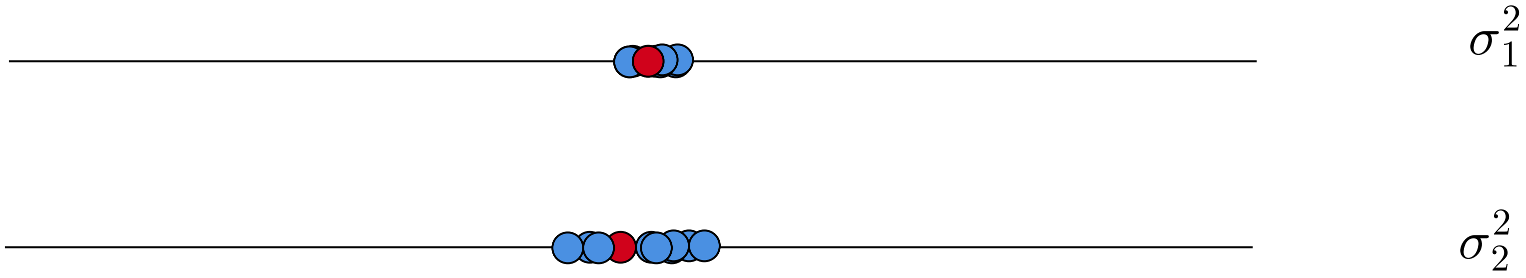
Variance

Variance

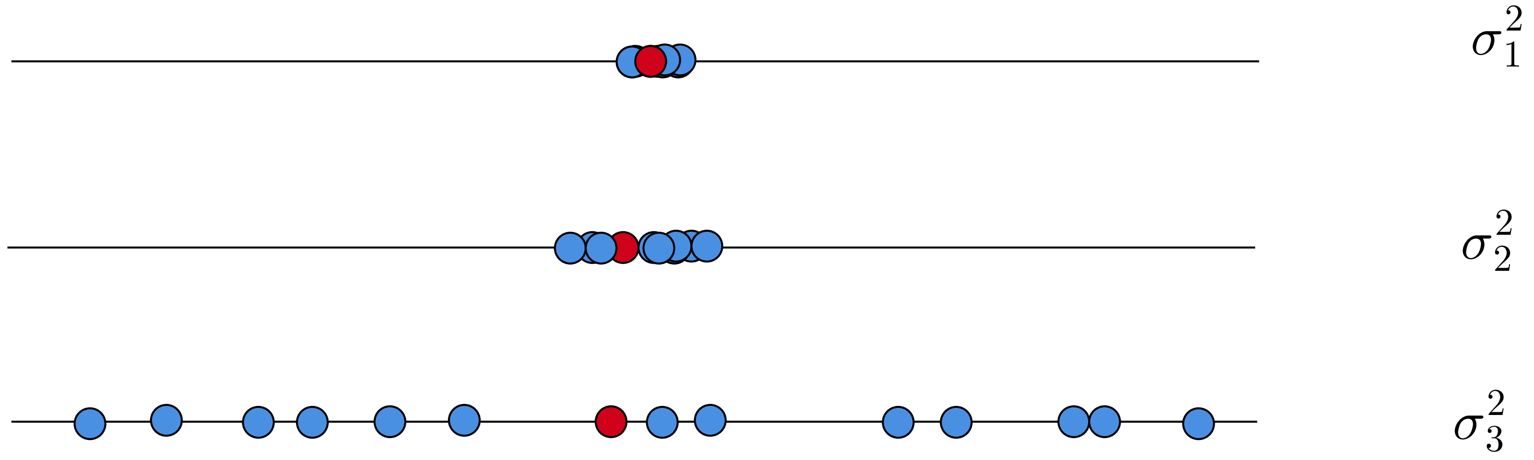


σ_1^2

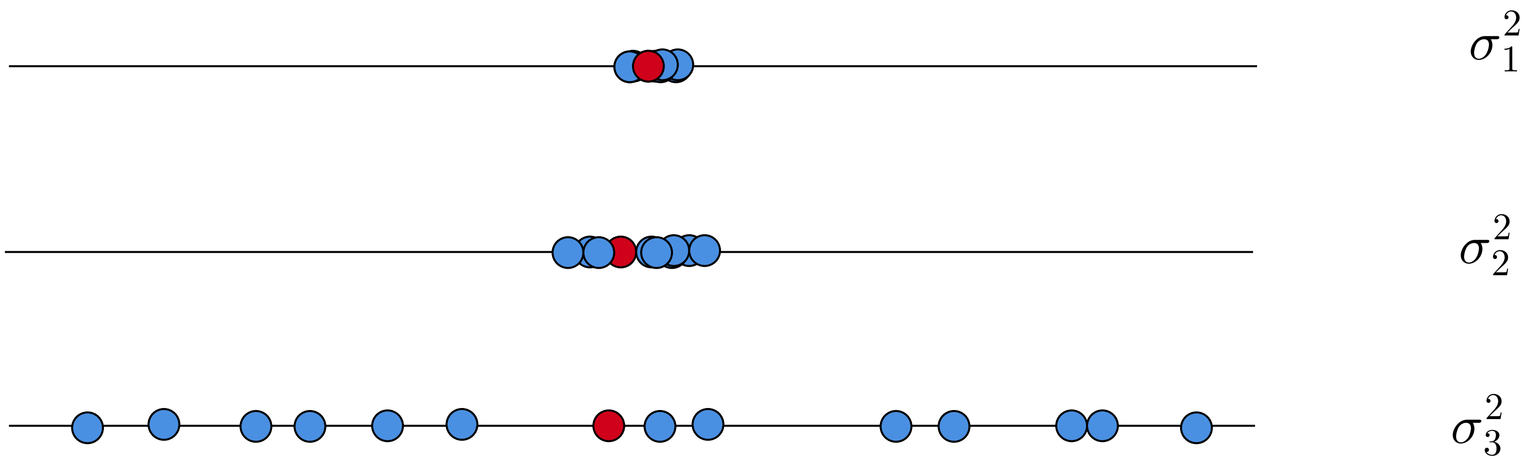
Variance



Variance

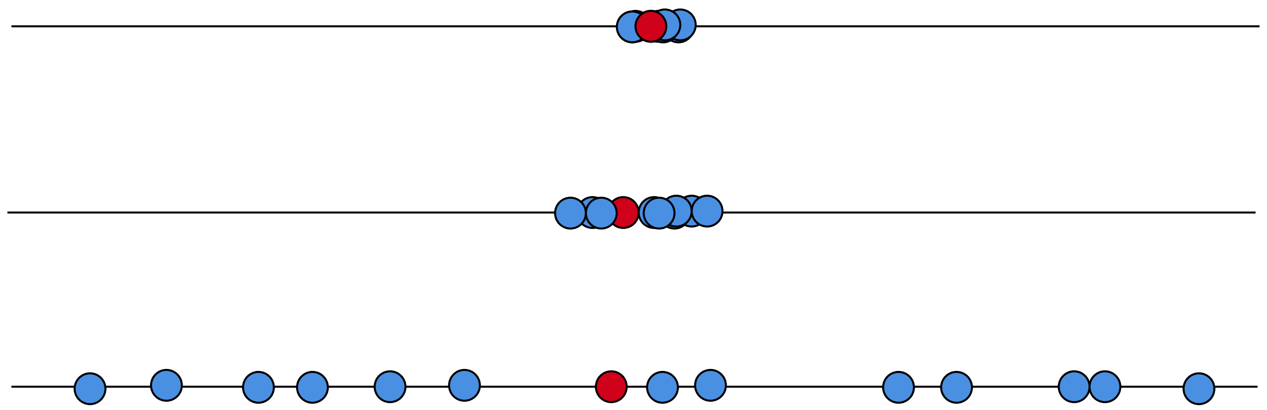


Variance



$$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$$

Variance



σ_1^2

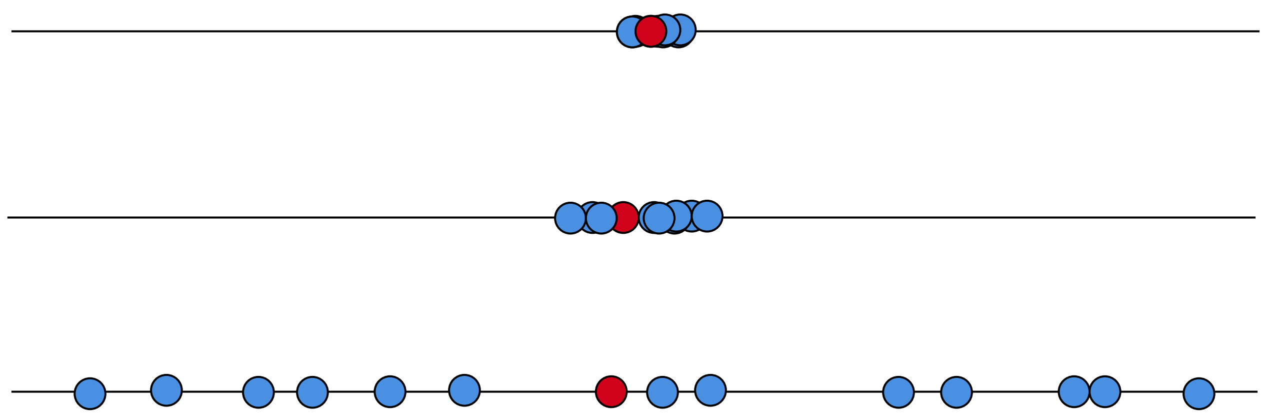
σ_2^2

σ_3^2

$$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$$

- Temperature
- Weight

Variance



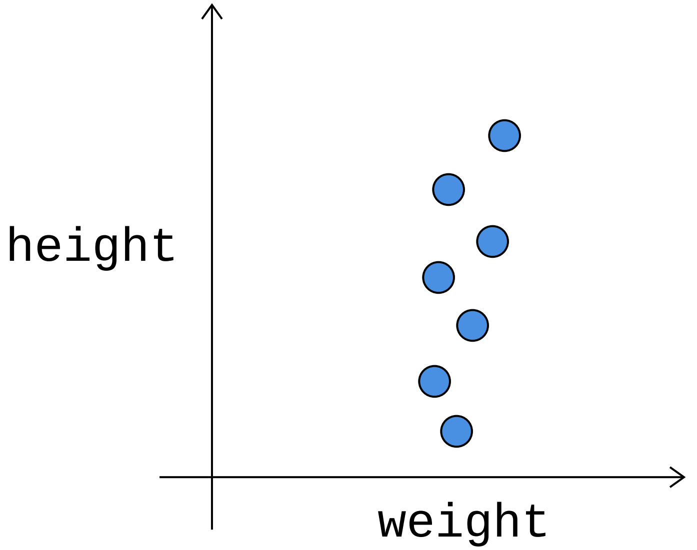
σ_1^2

σ_2^2

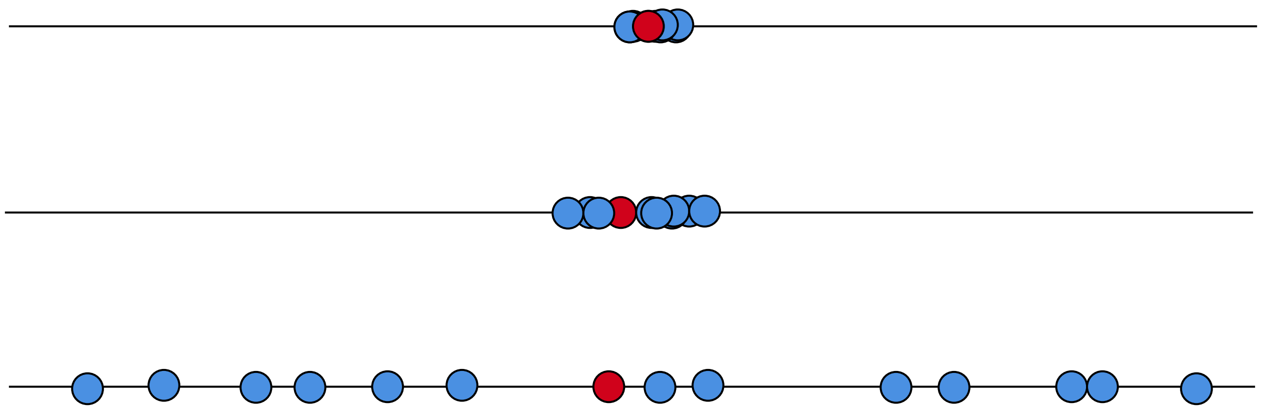
σ_3^2

$$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$$

- Temperature
- Weight



Variance



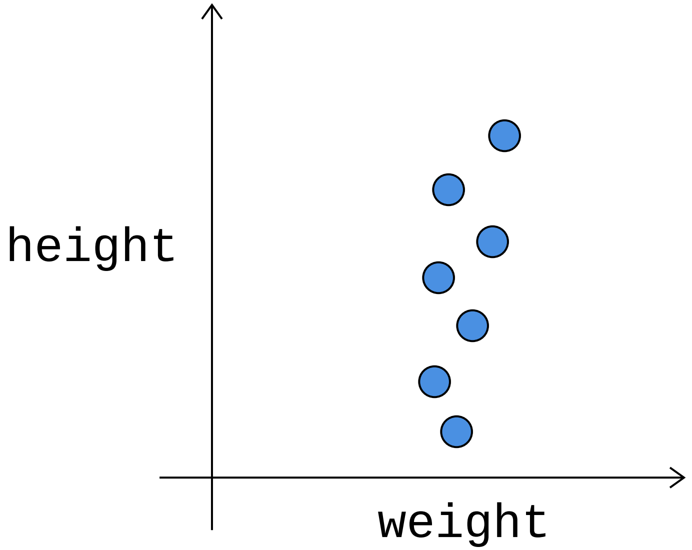
σ_1^2

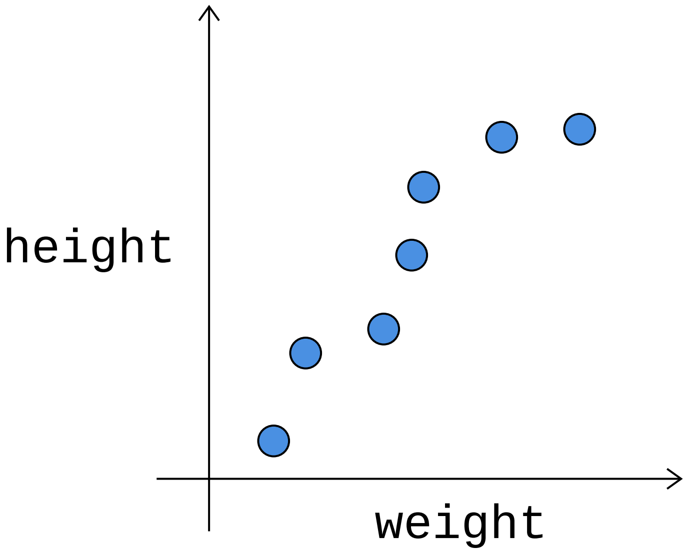
σ_2^2

σ_3^2

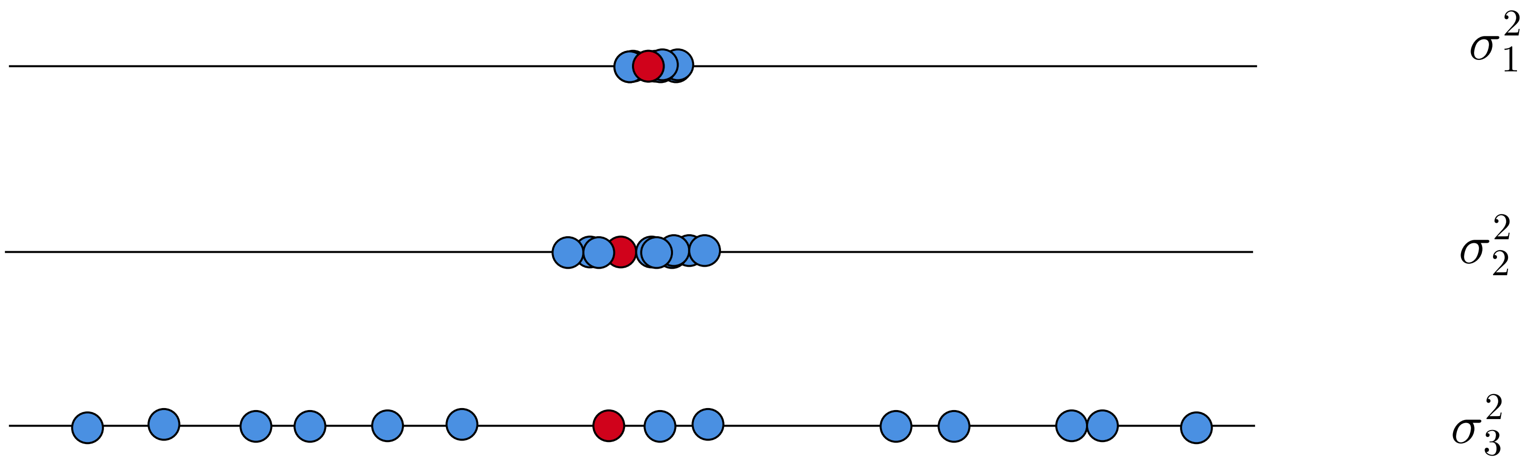
$$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$$

- Temperature
- Weight





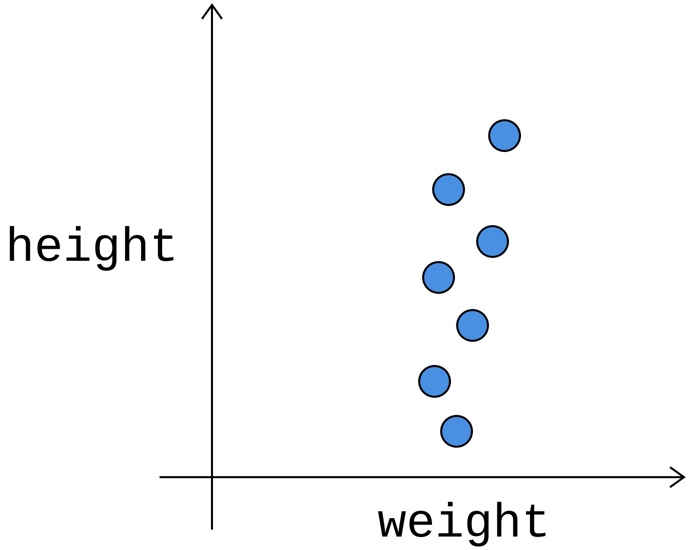
Variance

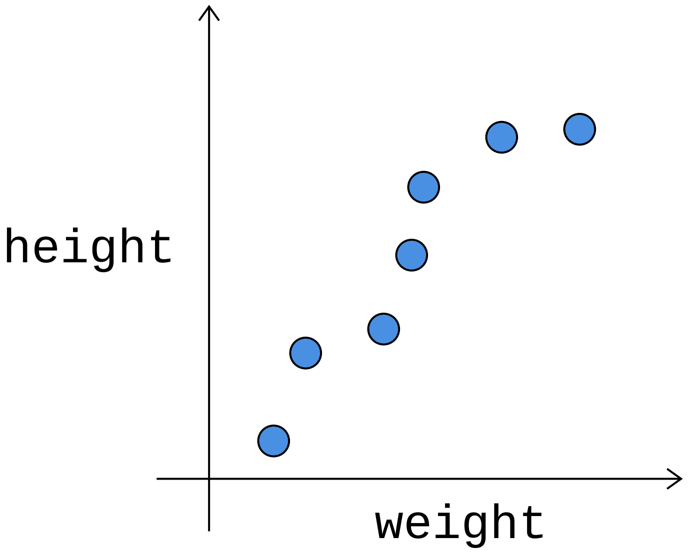


$$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$$

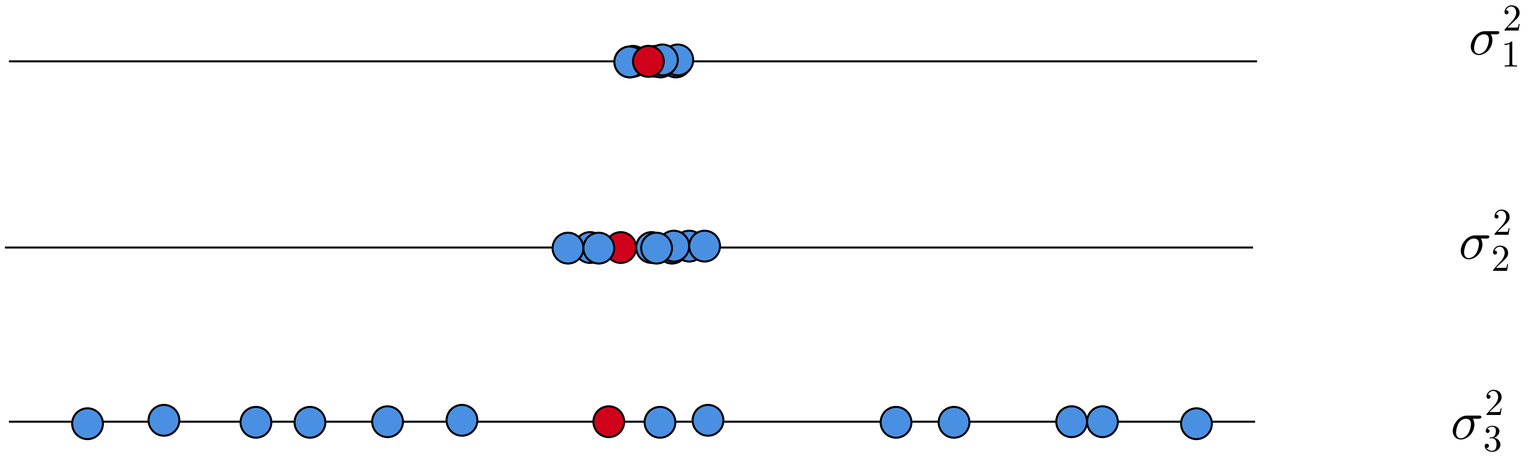
- Temperature
- Weight

- Variance
- Information
- Explanatory power





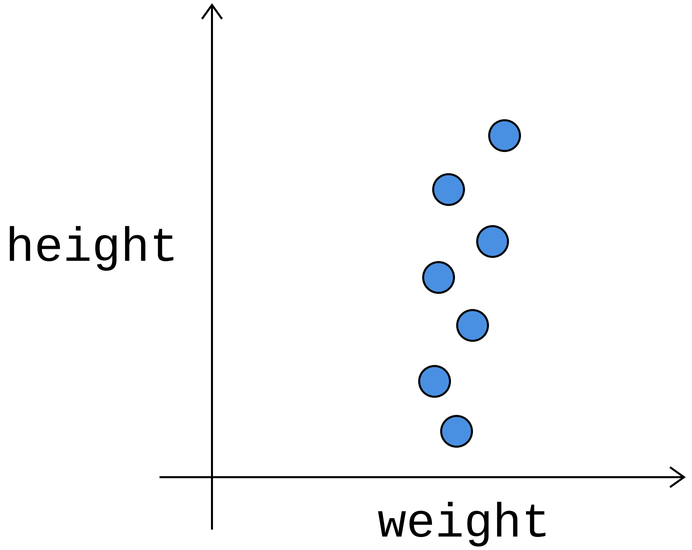
Variance

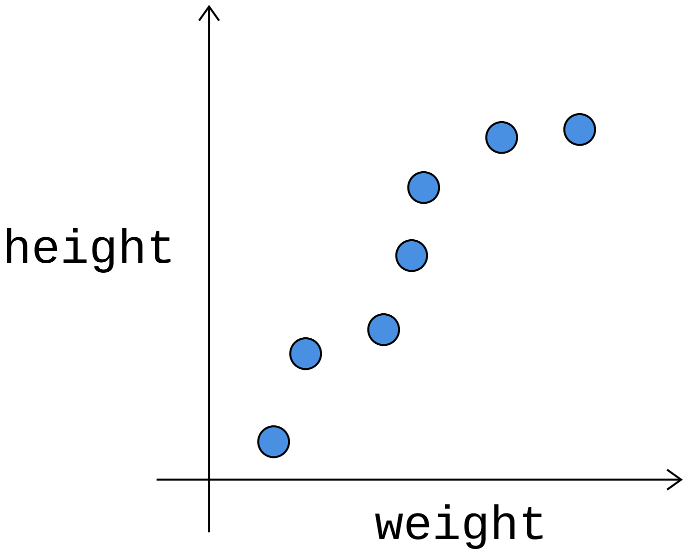


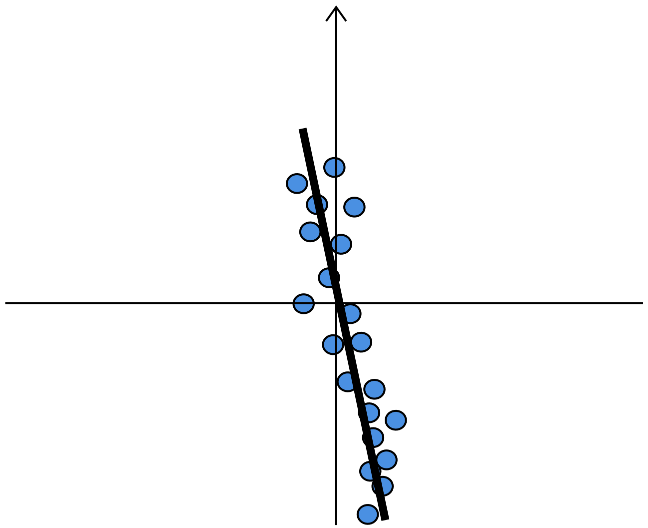
$$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$$

- Temperature
- Weight

- Variance
- Information
- Explanatory power





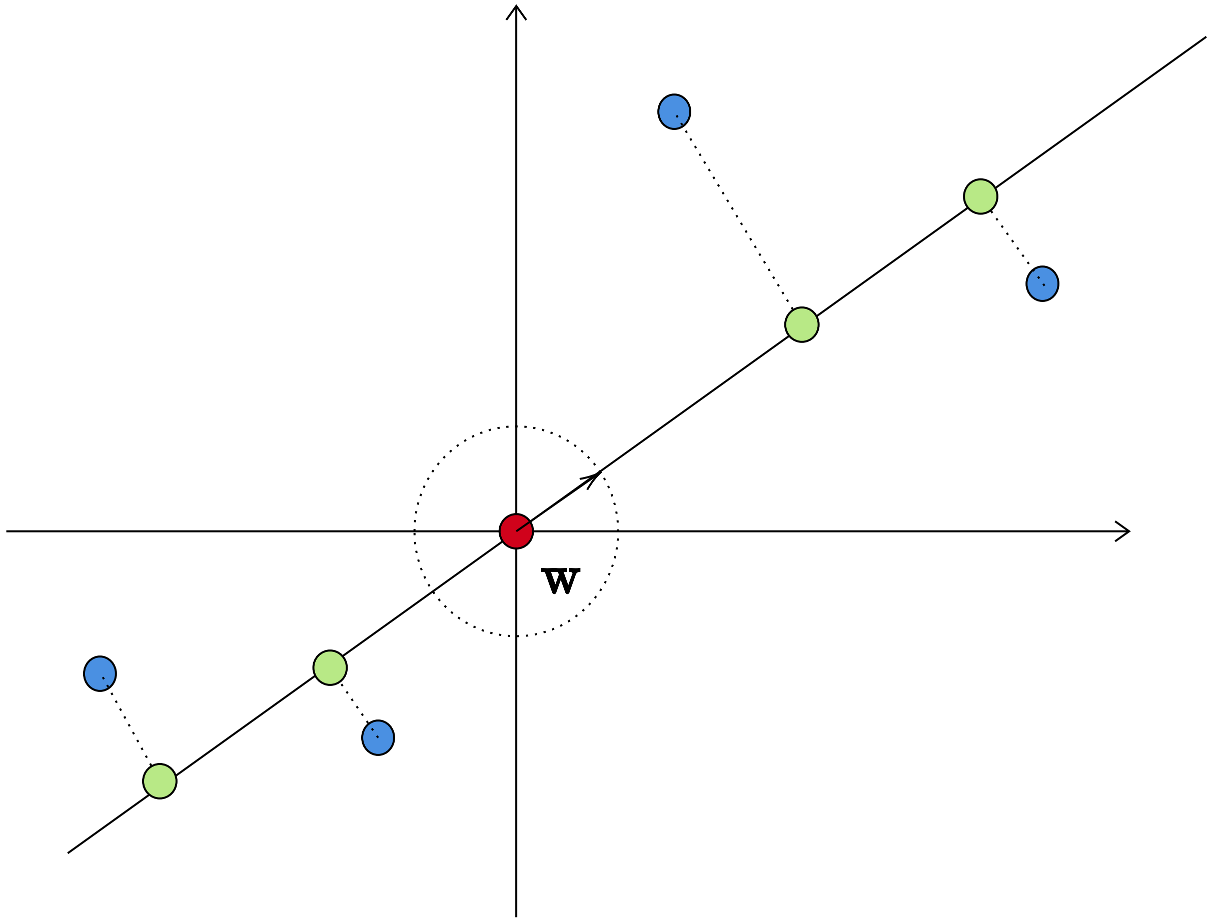


Variance Maximization

$$D = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$$

$$\boldsymbol{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

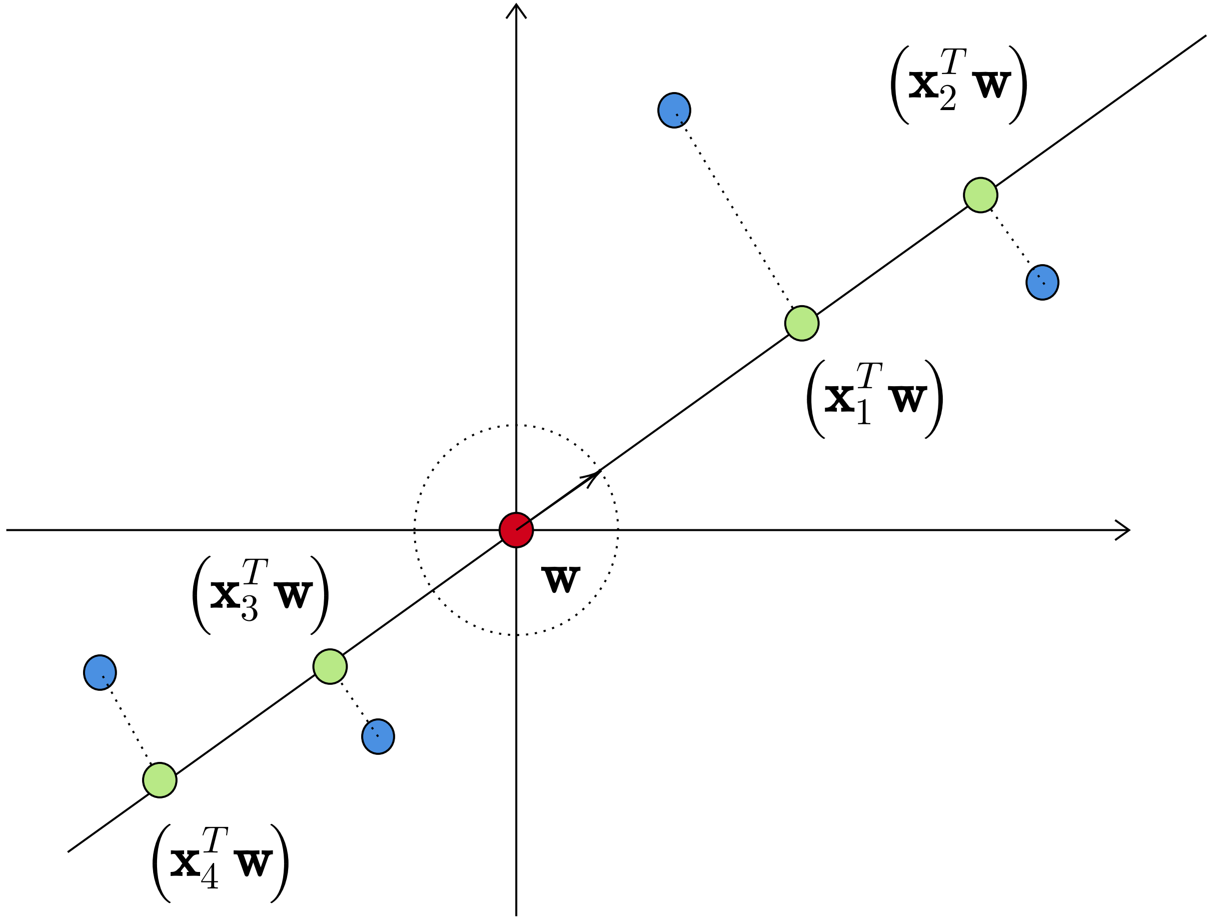
Variance Maximization



$$D = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$$

$$\boldsymbol{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

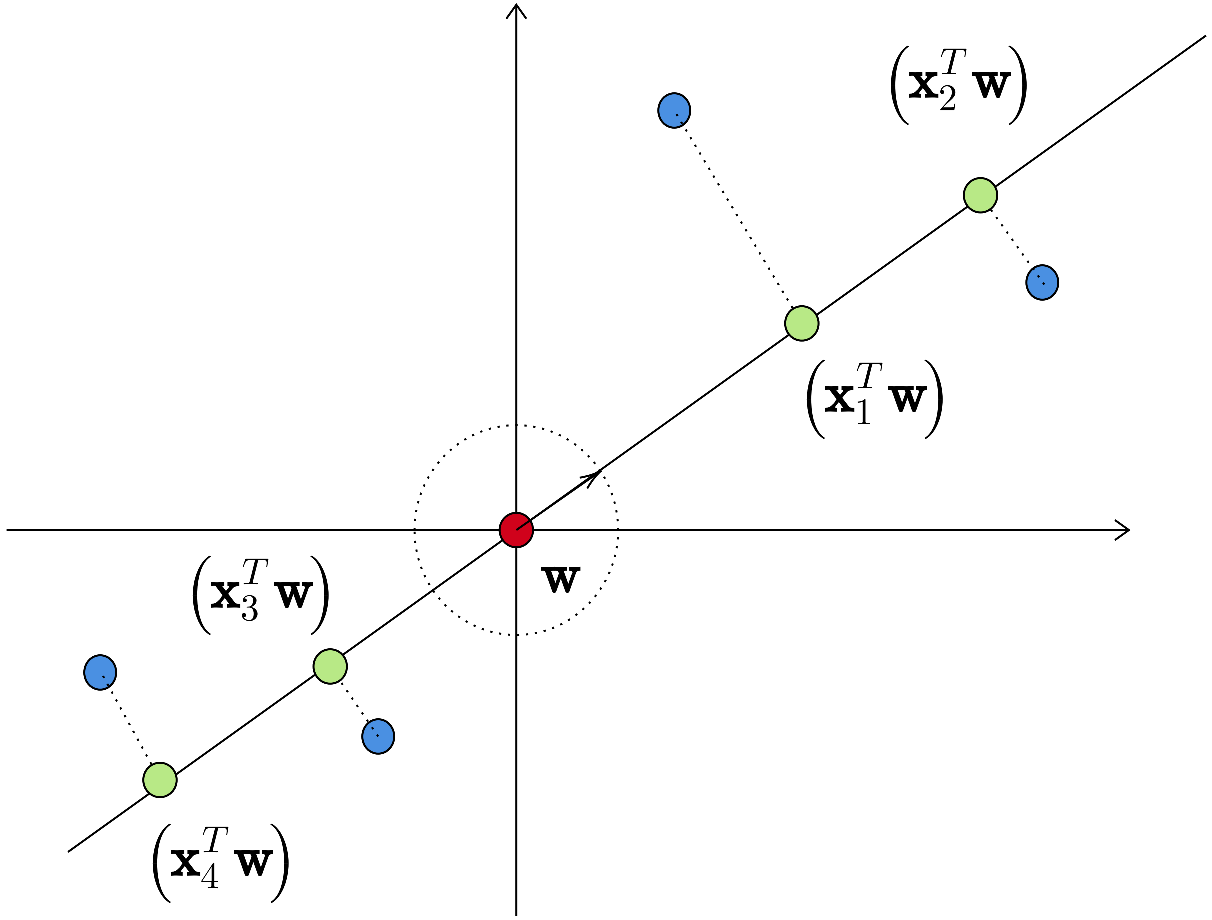
Variance Maximization



$$D = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$$

$$\boldsymbol{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

Variance Maximization

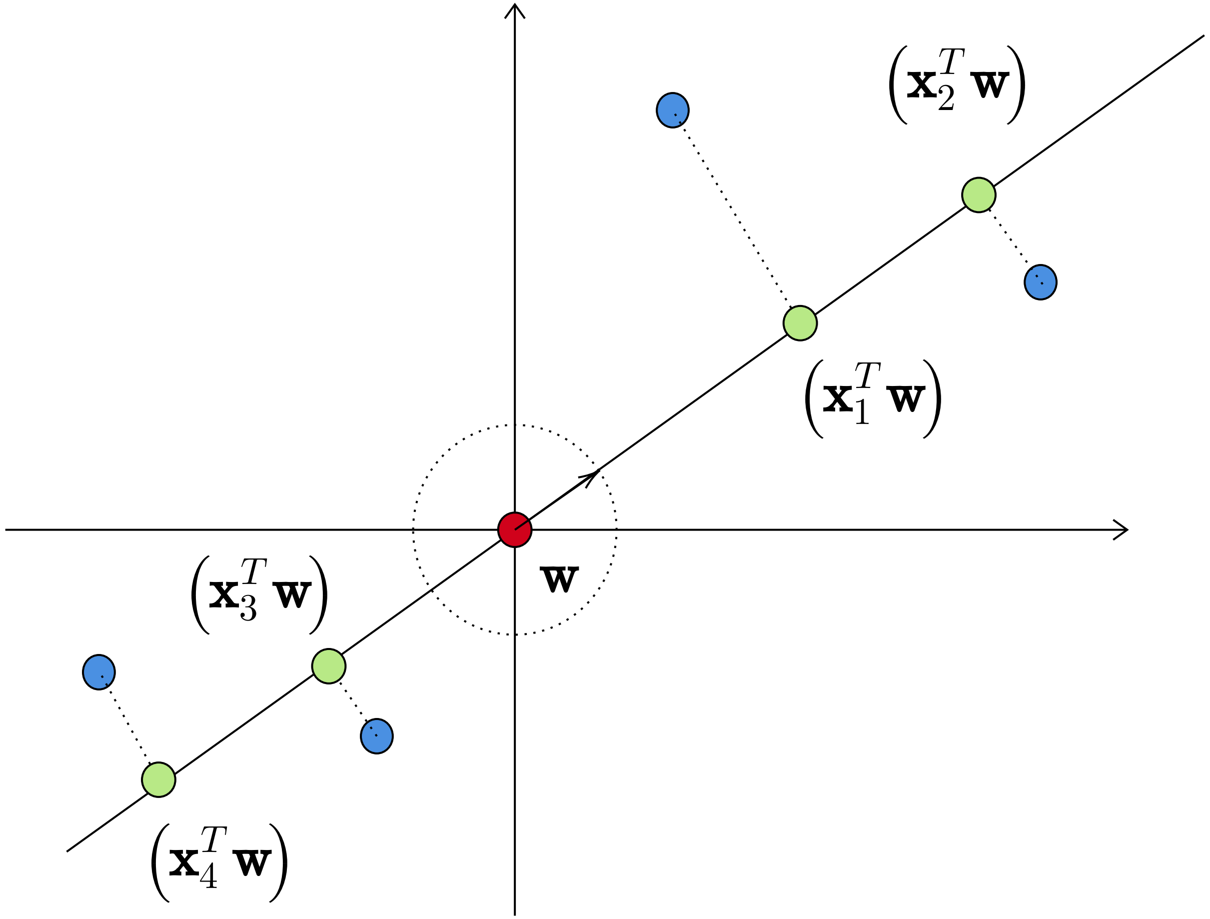


$$D = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$$

$$\boldsymbol{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

$$\mathrm{var}(D, \mathbf{w}) =$$

Variance Maximization

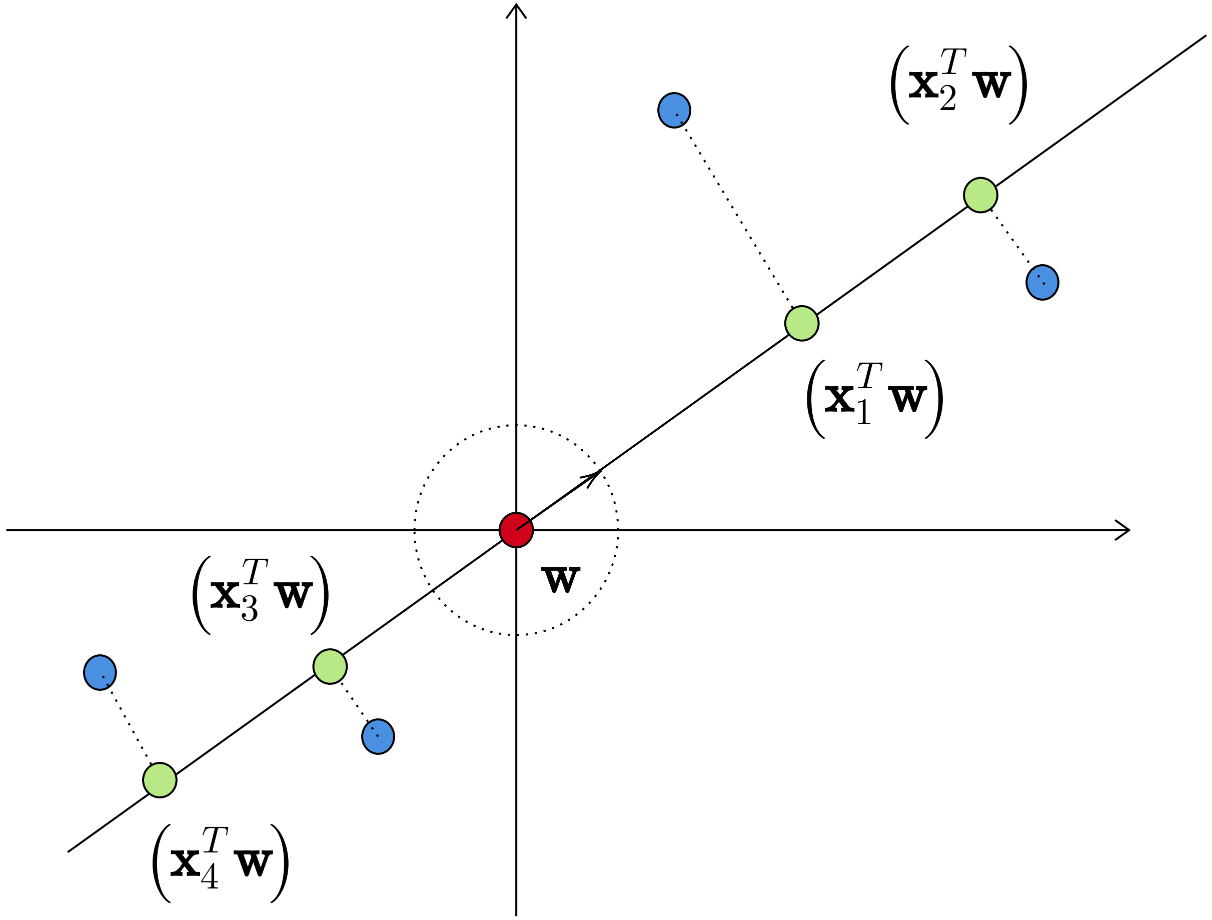


$$D = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$$

$$\boldsymbol{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

$$\mathrm{var}(D, \mathbf{w}) = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{w}\right)^2$$

Variance Maximization

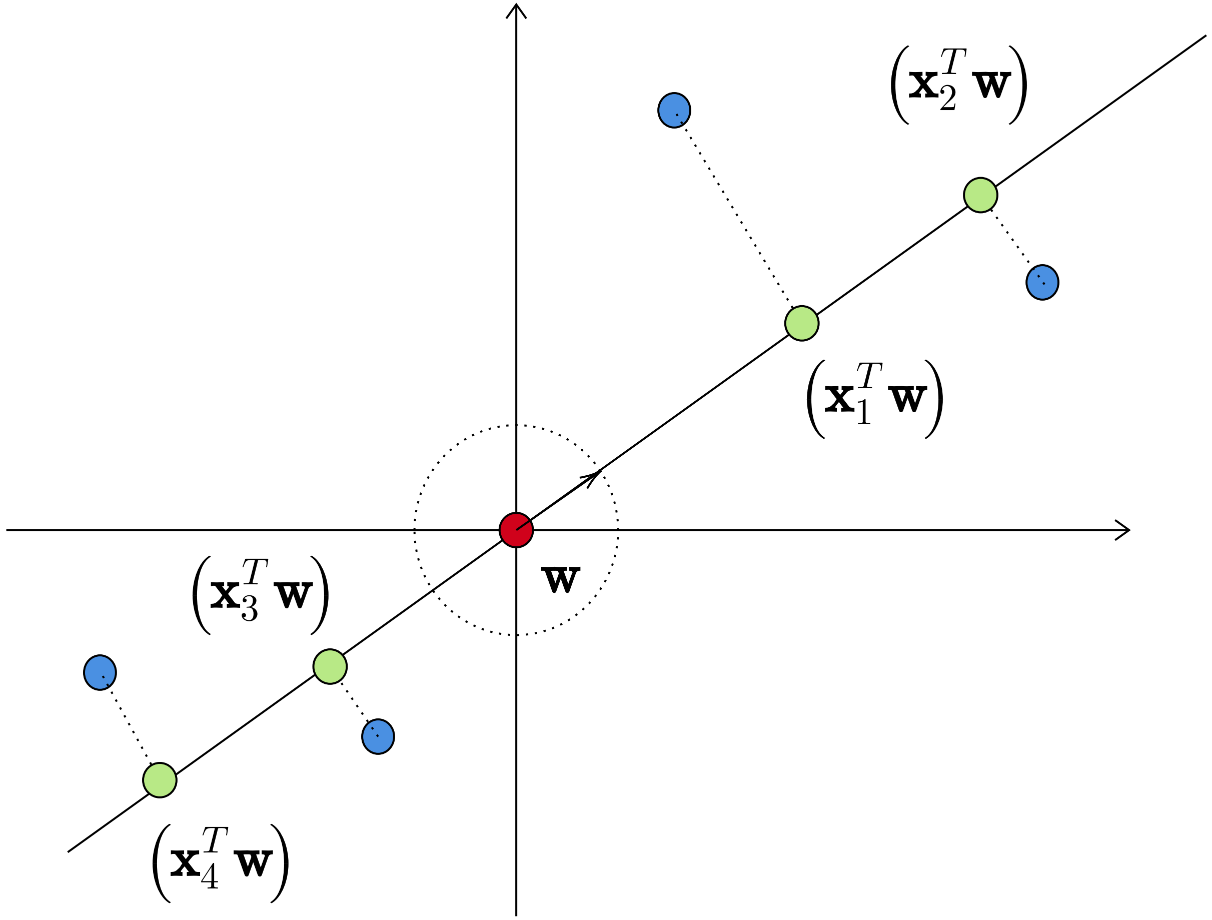


$$D = \{\mathbf{x}_1, \ \cdots \ , \mathbf{x}_n\}$$

$$\boldsymbol{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

$$\begin{aligned} \mathtt{var}(D, \mathbf{w}) &= \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{w}\right)^2 \\ &= \mathbf{w}^T \left[\frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right] \mathbf{w} \end{aligned}$$

Variance Maximization

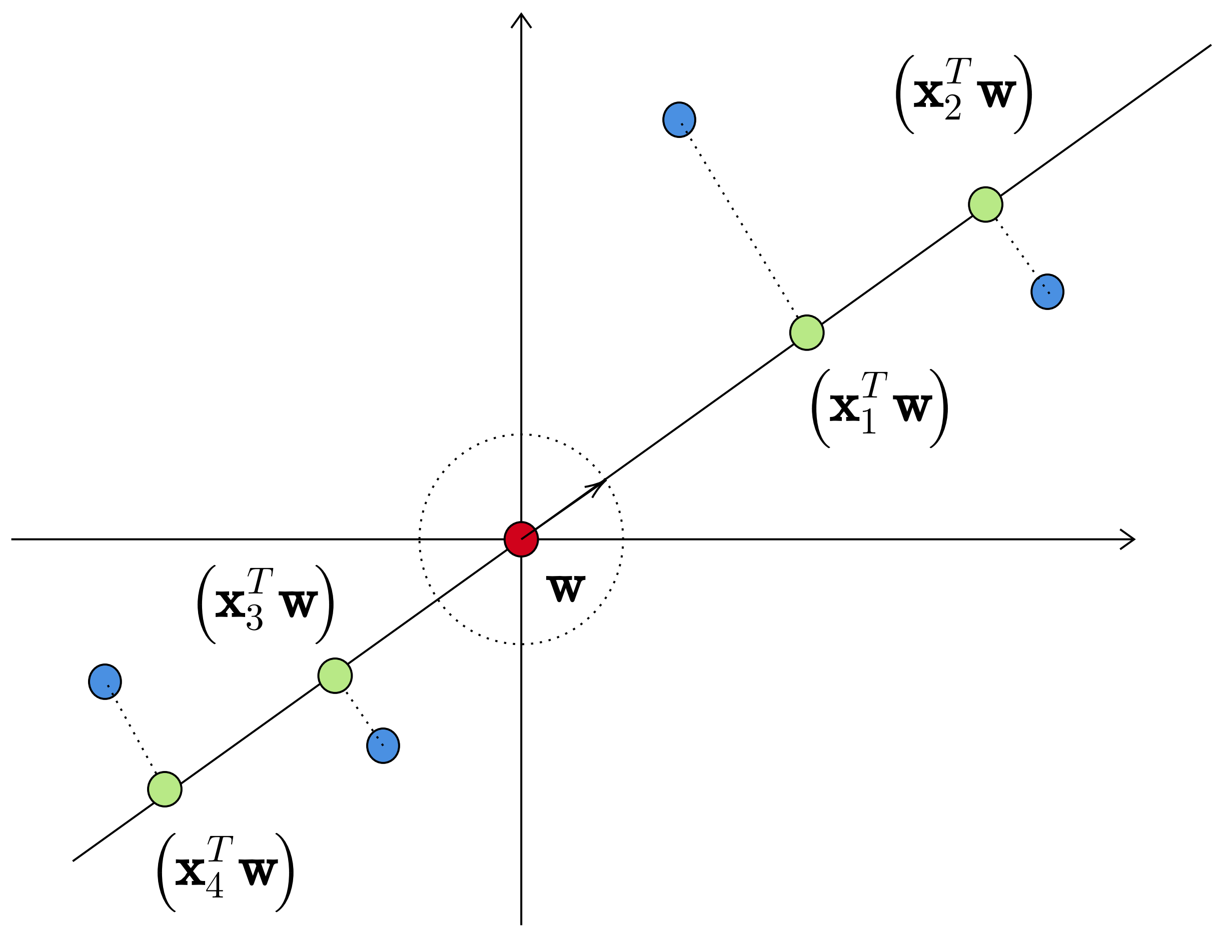


$$D = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$$

$$\boldsymbol{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

$$\begin{aligned} \mathtt{var}(D, \mathbf{w}) &= \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{w}\right)^2 \\ &= \mathbf{w}^T \left[\frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right] \mathbf{w} \\ &= \mathbf{w}^T \mathbf{C} \mathbf{w} \end{aligned}$$

Variance Maximization



$$D = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$$

$$\boldsymbol{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

$$\begin{aligned} \mathtt{var}(D, \mathbf{w}) &= \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{w}\right)^2 \\ &= \mathbf{w}^T \left[\frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right] \mathbf{w} \\ &= \mathbf{w}^T \mathbf{C} \mathbf{w} \end{aligned}$$

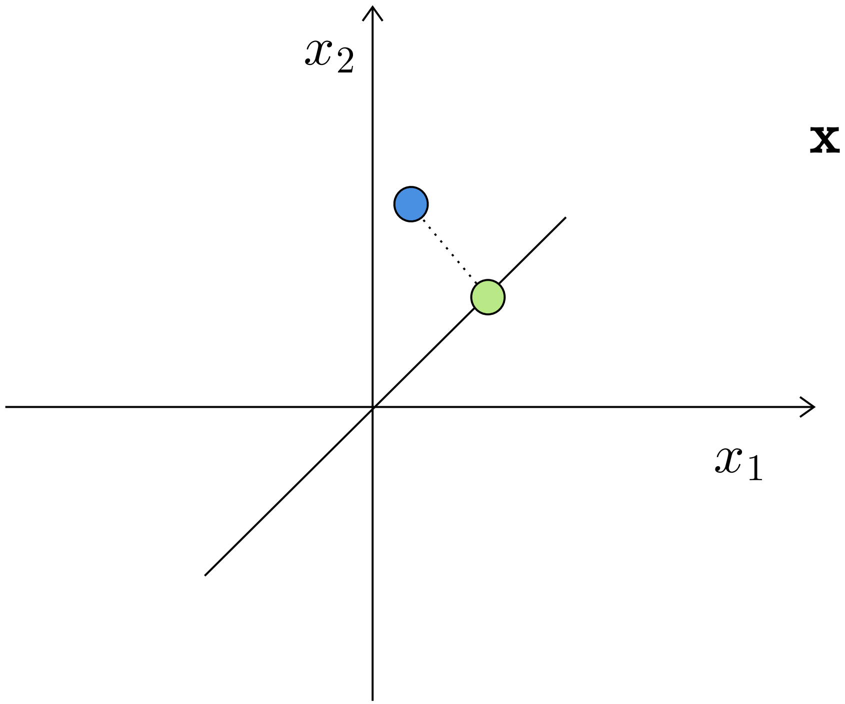
$$\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w} = \lambda_1 \qquad \qquad \arg\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w} = \mathbf{w}_1$$

Summary

$$\text{Error minimization} \underset{\text{PC}}{\overset{\mathbf{w}_1}{\equiv}} \text{Variance Maximization}$$

Summary

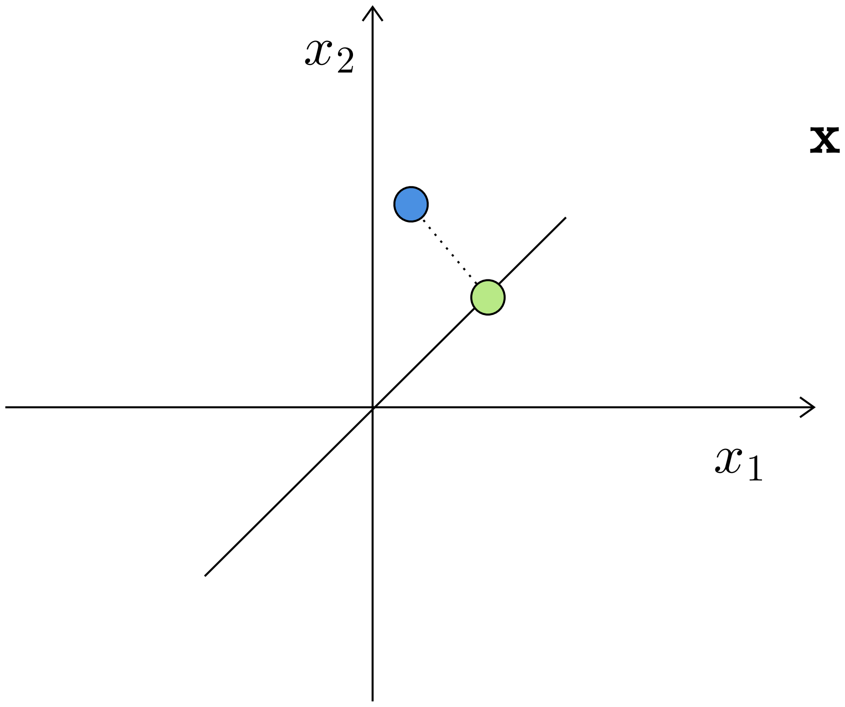
$$\text{Error minimization} \overset{\mathbf{w}_1}{\underset{\text{PC}}{\equiv}} \text{Variance Maximization}$$



$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

Summary

$$\text{Error minimization} \underset{\text{PC}}{\overset{\mathbf{w}_1}{\equiv}} \text{Variance Maximization}$$

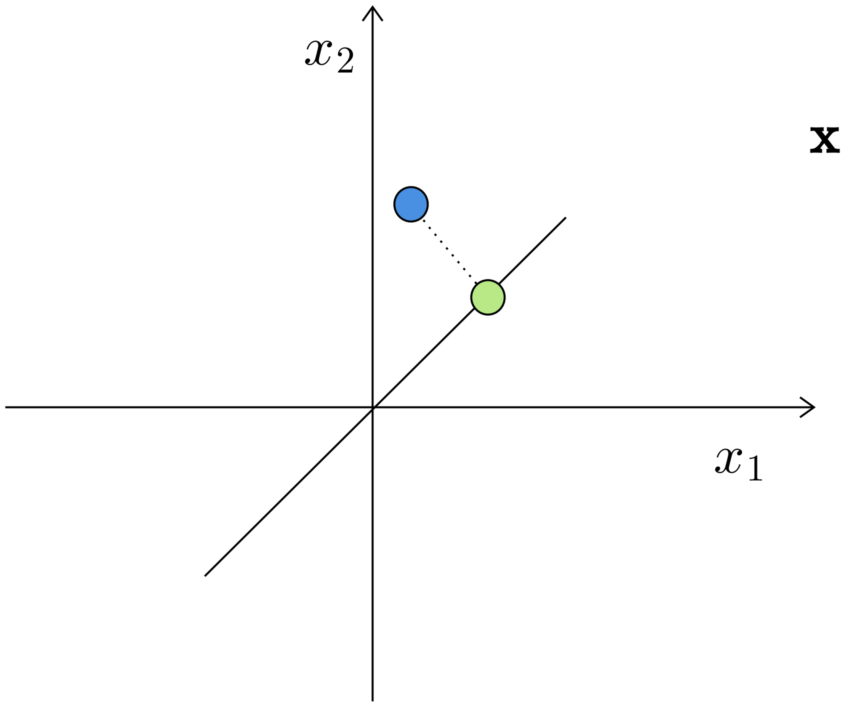


$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$\mathbf{x}^T\mathbf{w}_1 = \begin{bmatrix} x_1 & \cdots & x_d \end{bmatrix} \begin{bmatrix} w_{11} \\ \vdots \\ w_{d1} \end{bmatrix}$$

Summary

$$\text{Error minimization} \overset{\mathbf{w}_1}{\underset{\text{PC}}{\equiv}} \text{Variance Maximization}$$



$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$\mathbf{x}^T\mathbf{w}_1 = \begin{bmatrix} x_1 & \cdots & x_d \end{bmatrix} \begin{bmatrix} w_{11} \\ \vdots \\ w_{d1} \end{bmatrix}$$

$$= \boxed{w_{11}}\, x_1 + \cdots + \boxed{w_{i1}}\, x_i + \cdots + \boxed{w_{d1}}\, x_d$$

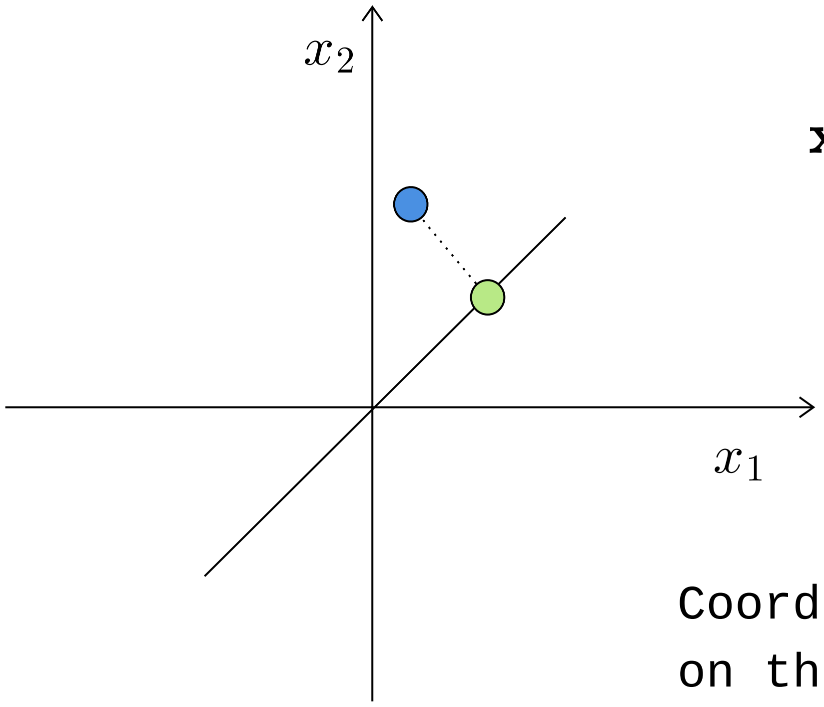
Summary

Note that the first PC is $\mathbf{w}_1$, a vector. In this slide, we look at how the d features get transformed when they are projected onto a PC.

$\mathbf{w}_1$

Error minimization $\equiv$ Variance Maximization

PC



$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$\mathbf{x}^T \mathbf{w}_1 = \begin{bmatrix} x_1 & \cdots & x_d \end{bmatrix} \begin{bmatrix} w_{11} \\ \vdots \\ w_{d1} \end{bmatrix}$$

$$= \boxed{w_{11}} x_1 + \cdots + \boxed{w_{i1}} x_i + \cdots + \boxed{w_{d1}} x_d$$

Coordinate of **x** on the first PC.

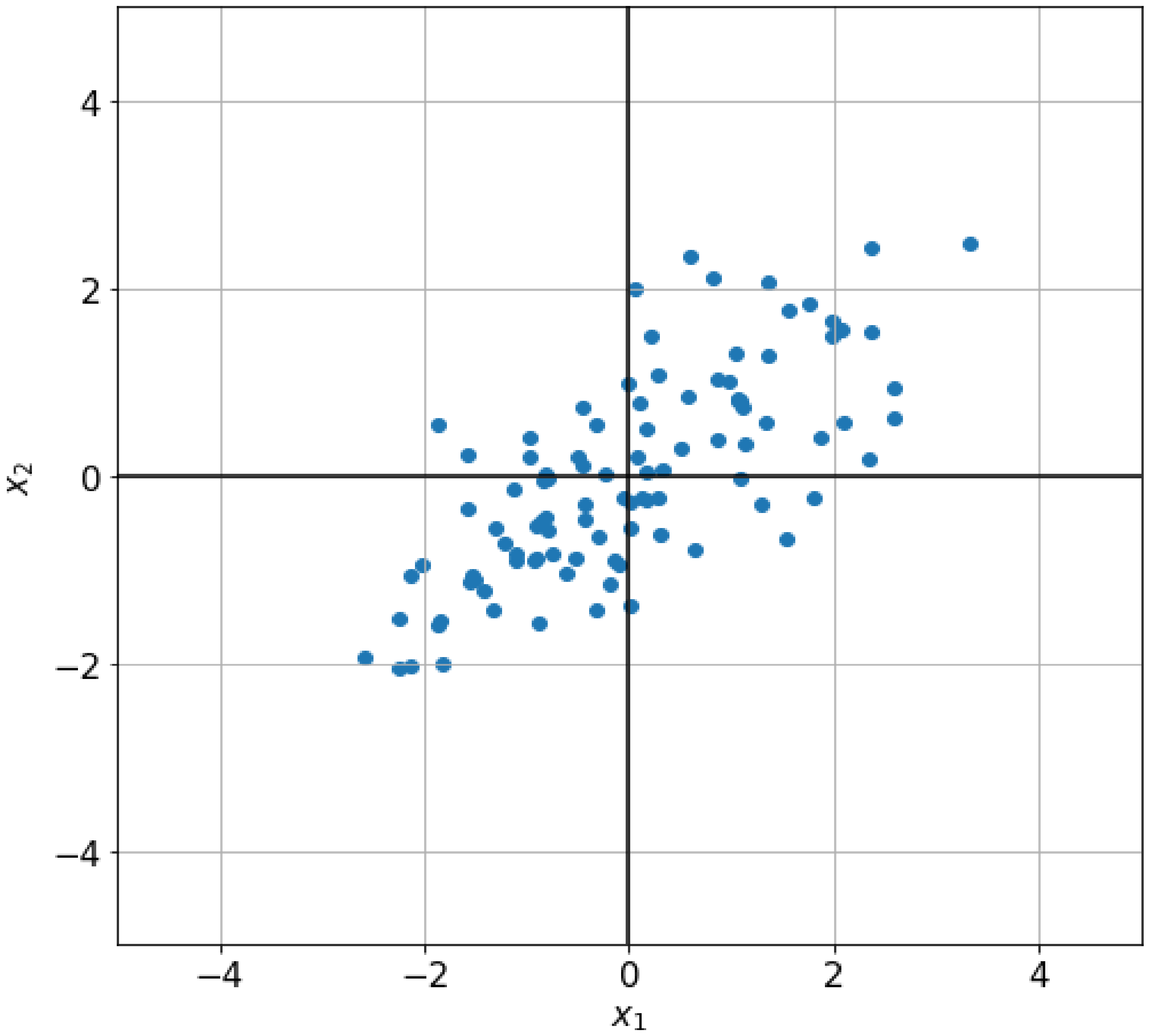
PC (coordinate view)

- Each PC specifies a new coordinate
- Linear combination of features
- Recall change of basis

Demo

$n = 100$ $d = 2$

Demo

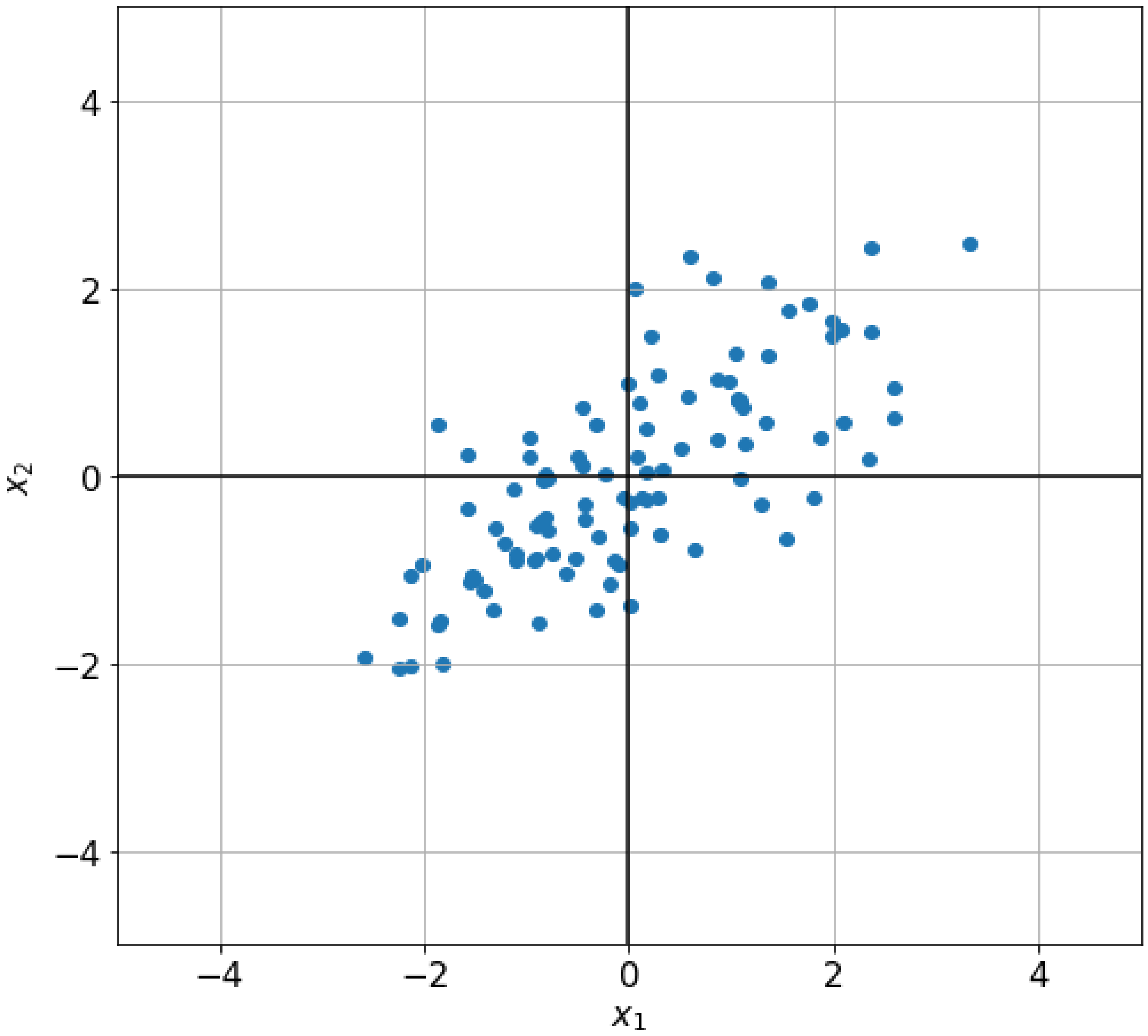


$n = 100$ $d = 2$

Demo

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

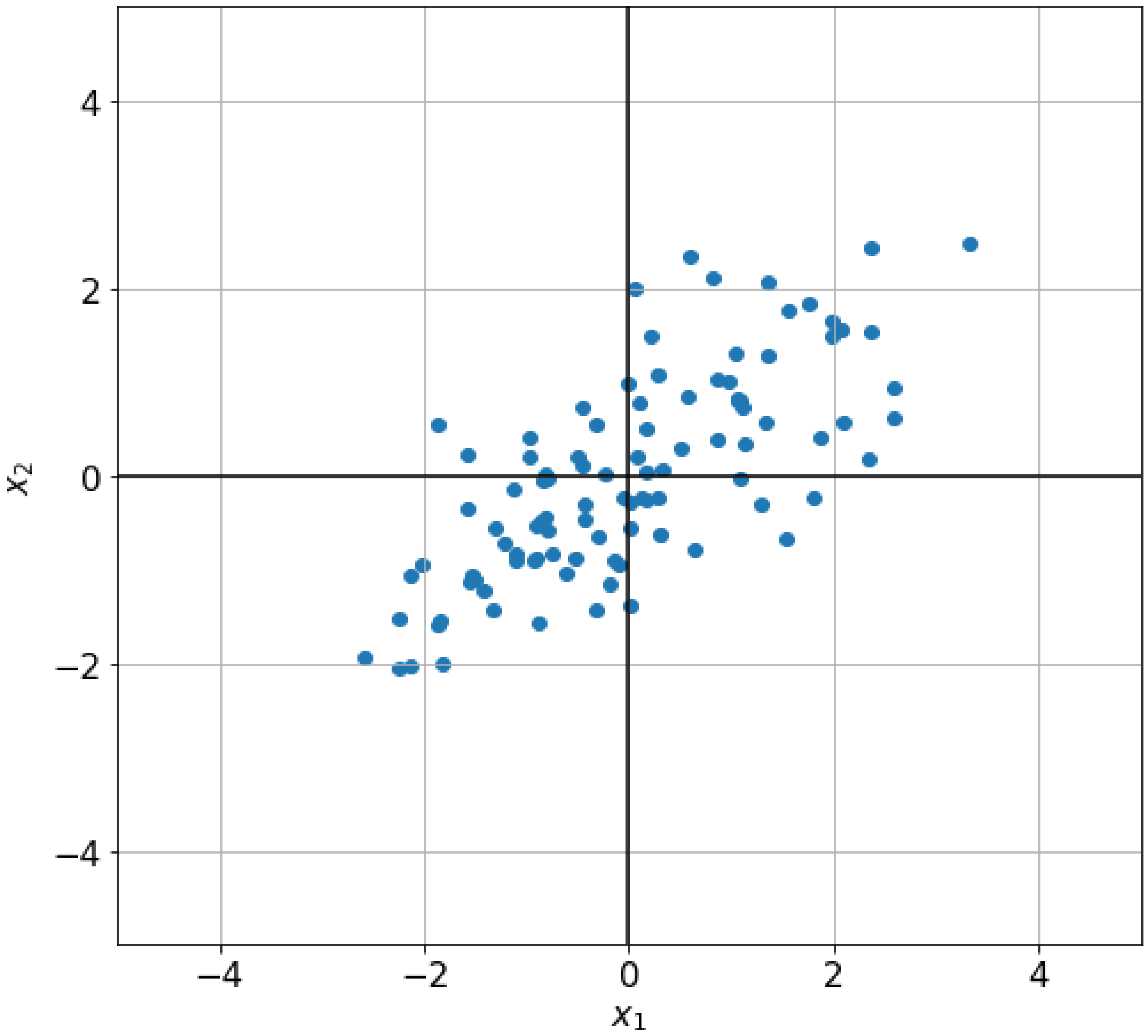


$n = 100$ $\quad d = 2$

Demo

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

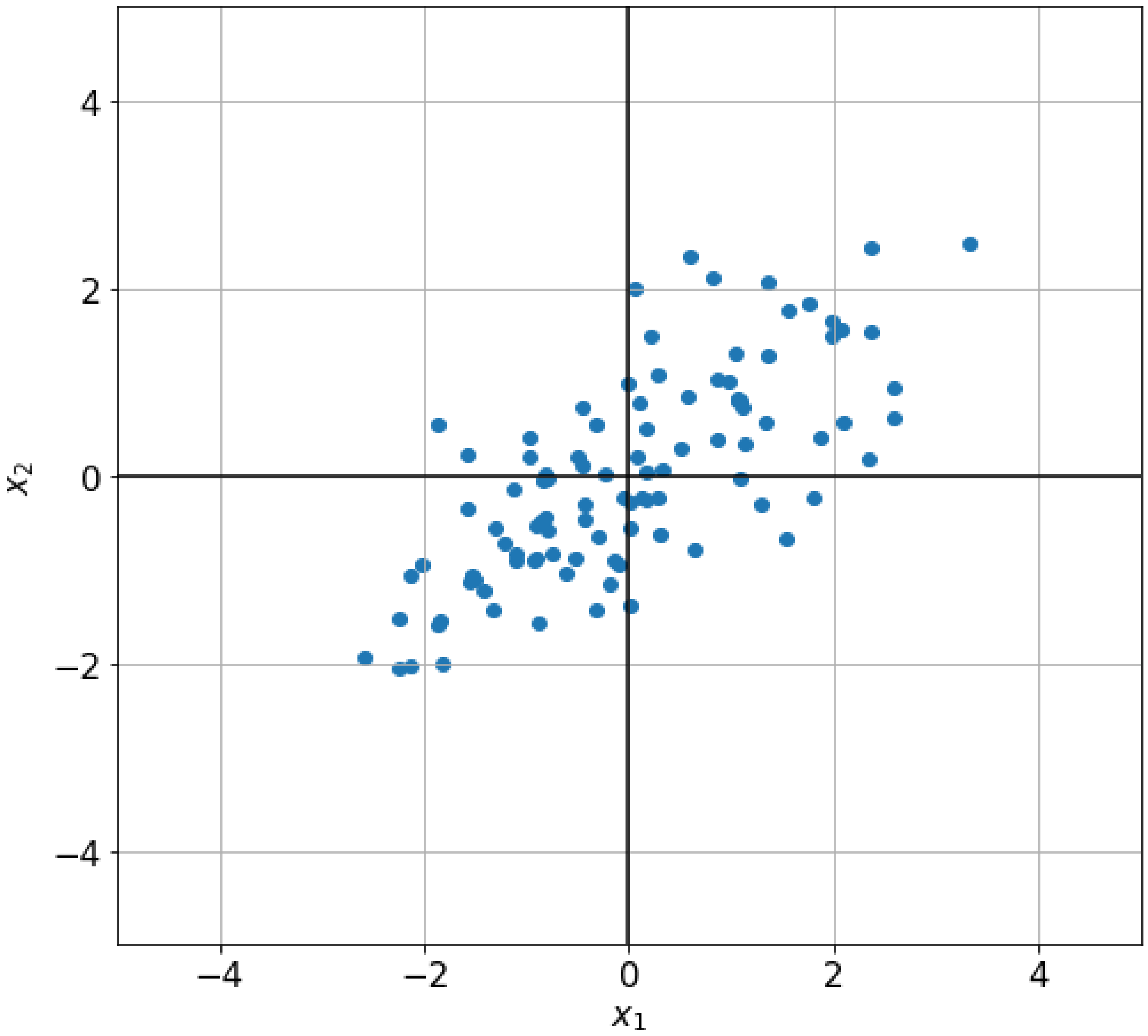


$n = 100$ $d = 2$

Demo

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$



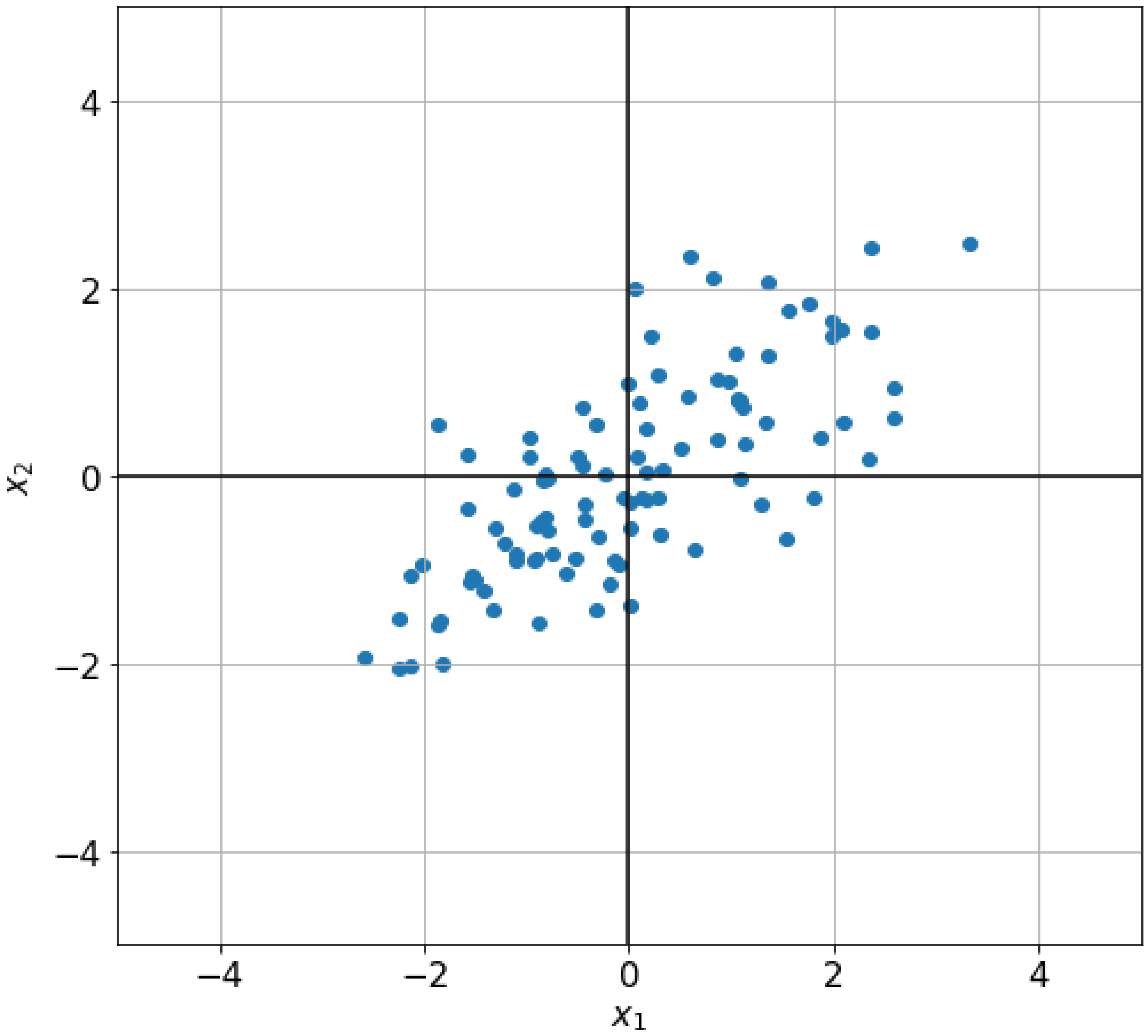
$n = 100$ $\qquad d = 2$

$$\mathbf{C} = \begin{bmatrix} 1.70 & 1.03 \\ 1.03 & 1.17 \end{bmatrix}$$

Demo

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$



$n = 100 \qquad d = 2$

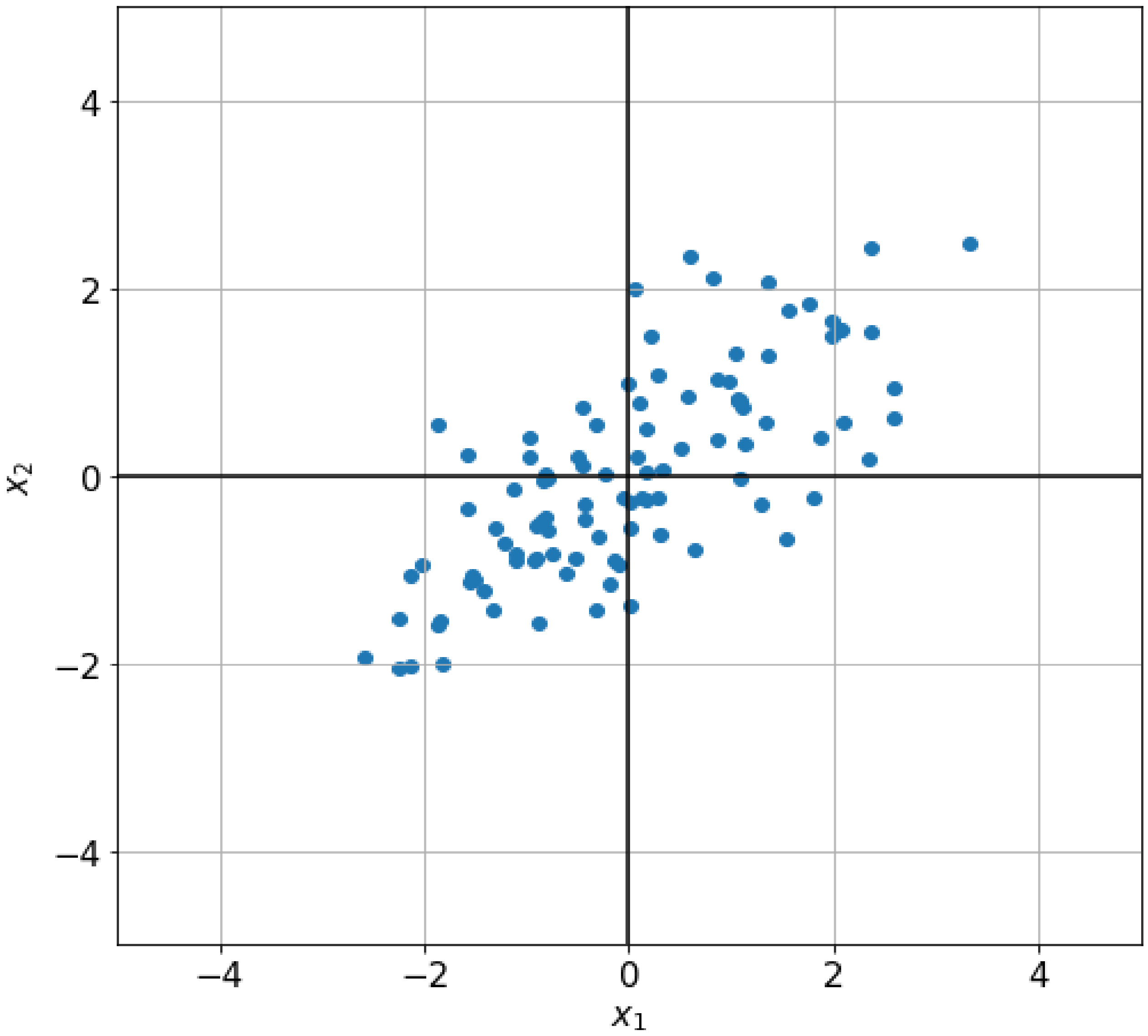
$$\mathbf{C} = \begin{bmatrix} 1.70 & 1.03 \\ 1.03 & 1.17 \end{bmatrix}$$

$$(\lambda_1, \mathbf{w}_1) = \texttt{Solver}(\mathbf{C})$$

Demo

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$



$n = 100 \qquad d = 2$

$$\mathbf{C} = \begin{bmatrix} 1.70 & 1.03 \\ 1.03 & 1.17 \end{bmatrix}$$

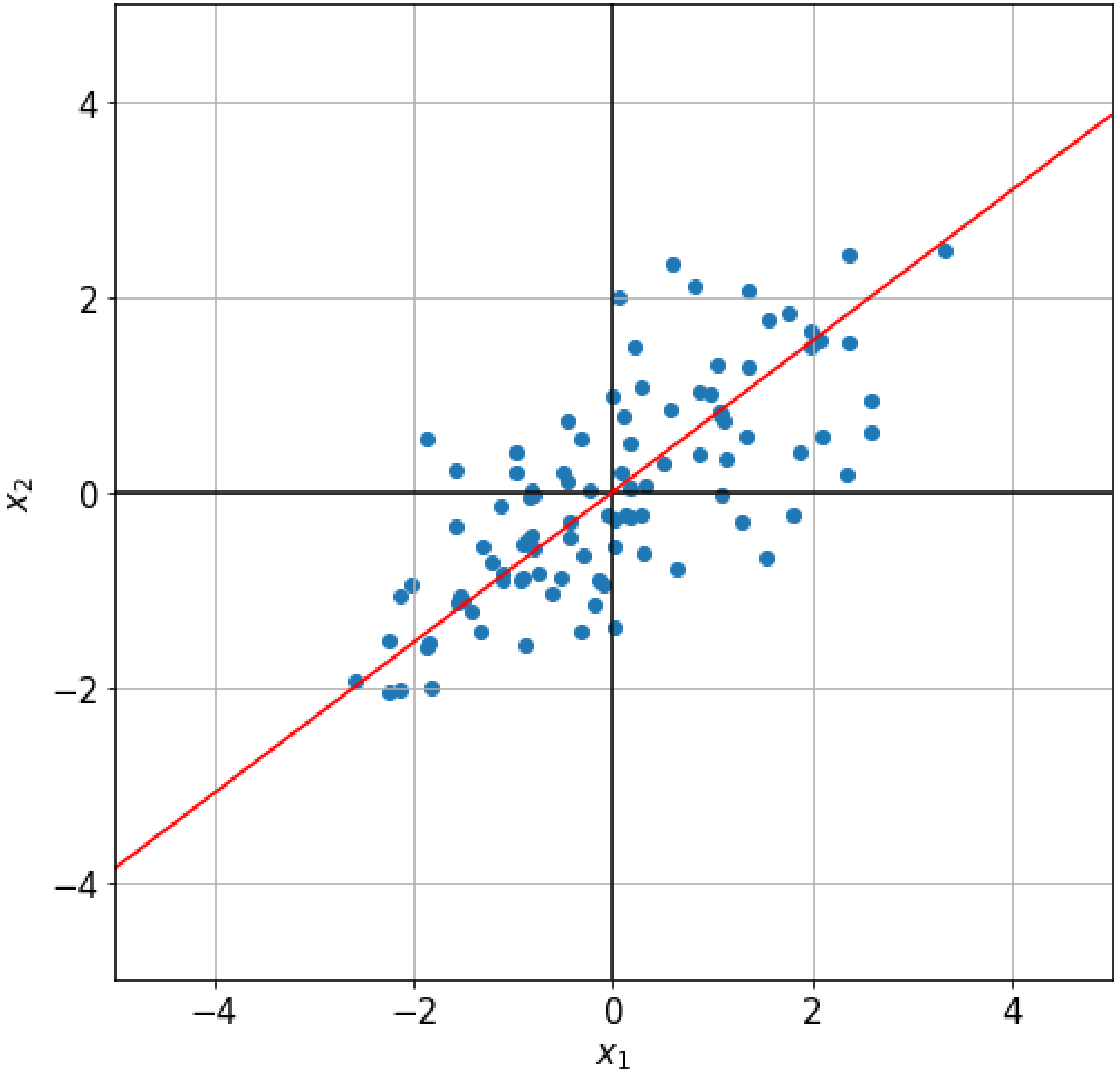
$$(\lambda_1, \mathbf{w}_1) = \texttt{Solver}(\mathbf{C})$$

$$\lambda_1 = 2.5 \qquad \mathbf{w}_1 = -\begin{bmatrix} 0.79 \\ 0.61 \end{bmatrix}$$

Demo

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$



$n = 100 \qquad d = 2$

$$\mathbf{C} = \begin{bmatrix} 1.70 & 1.03 \\ 1.03 & 1.17 \end{bmatrix}$$

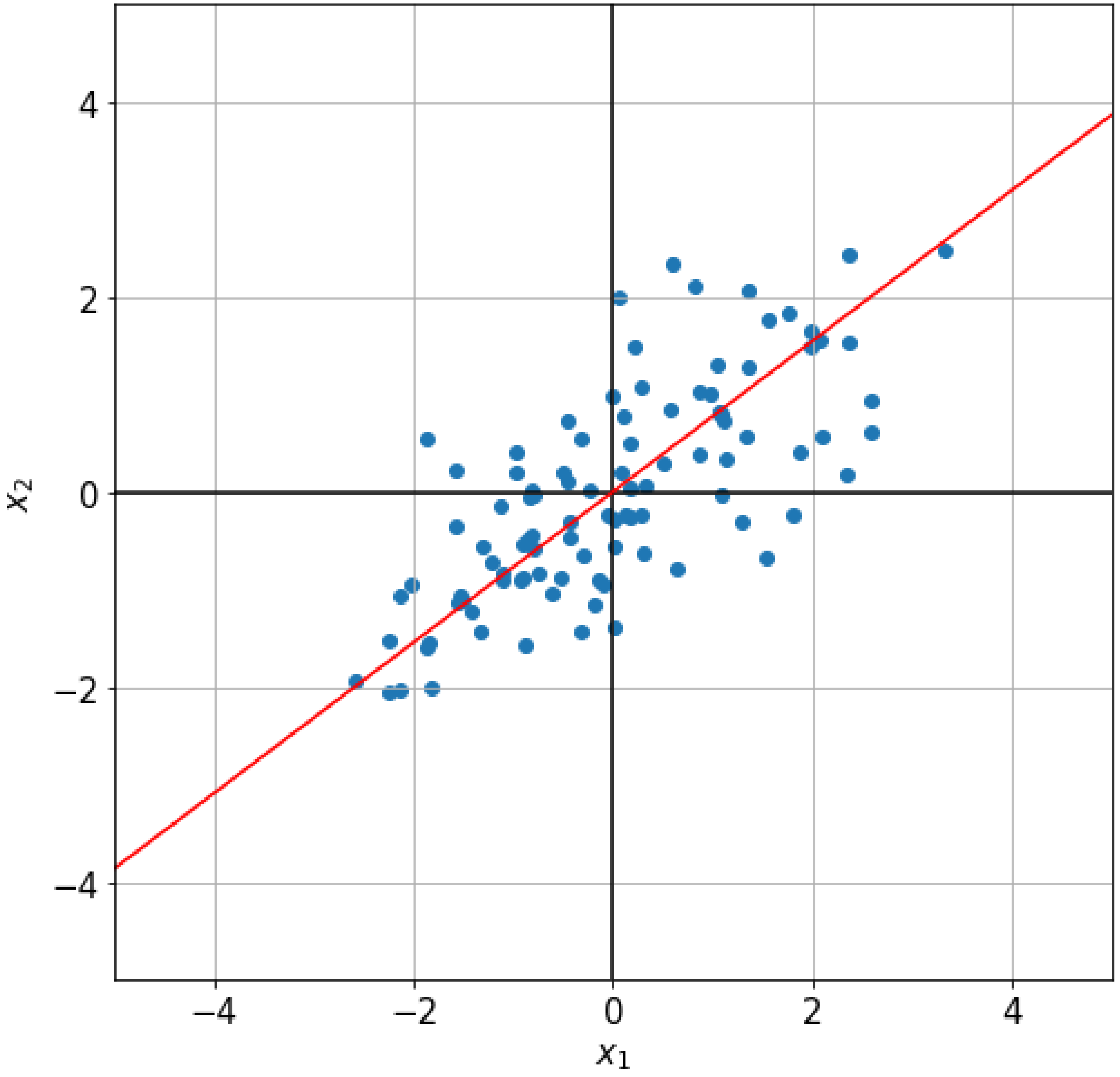
$$(\lambda_1, \mathbf{w}_1) = \texttt{Solver}(\mathbf{C})$$

$$\lambda_1 = 2.5 \qquad \mathbf{w}_1 = -\begin{bmatrix} 0.79 \\ 0.61 \end{bmatrix}$$

Demo

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$



$n = 100 \qquad d = 2$

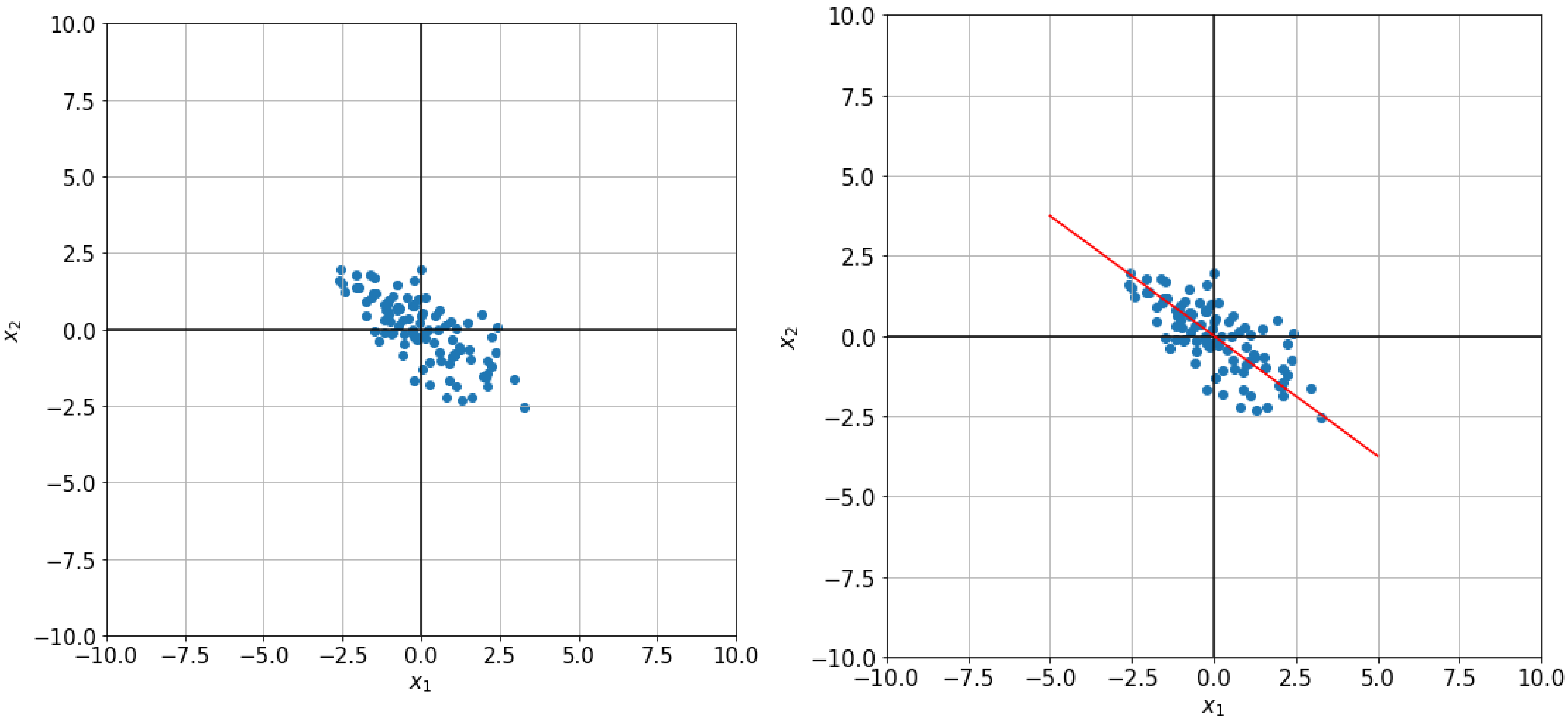
$$\mathbf{C} = \begin{bmatrix} 1.70 & 1.03 \\ 1.03 & 1.17 \end{bmatrix}$$

$$(\lambda_1, \mathbf{w}_1) = \texttt{Solver}(\mathbf{C})$$

$$\lambda_1 = 2.5 \qquad \mathbf{w}_1 = -\begin{bmatrix} 0.79 \\ 0.61 \end{bmatrix}$$

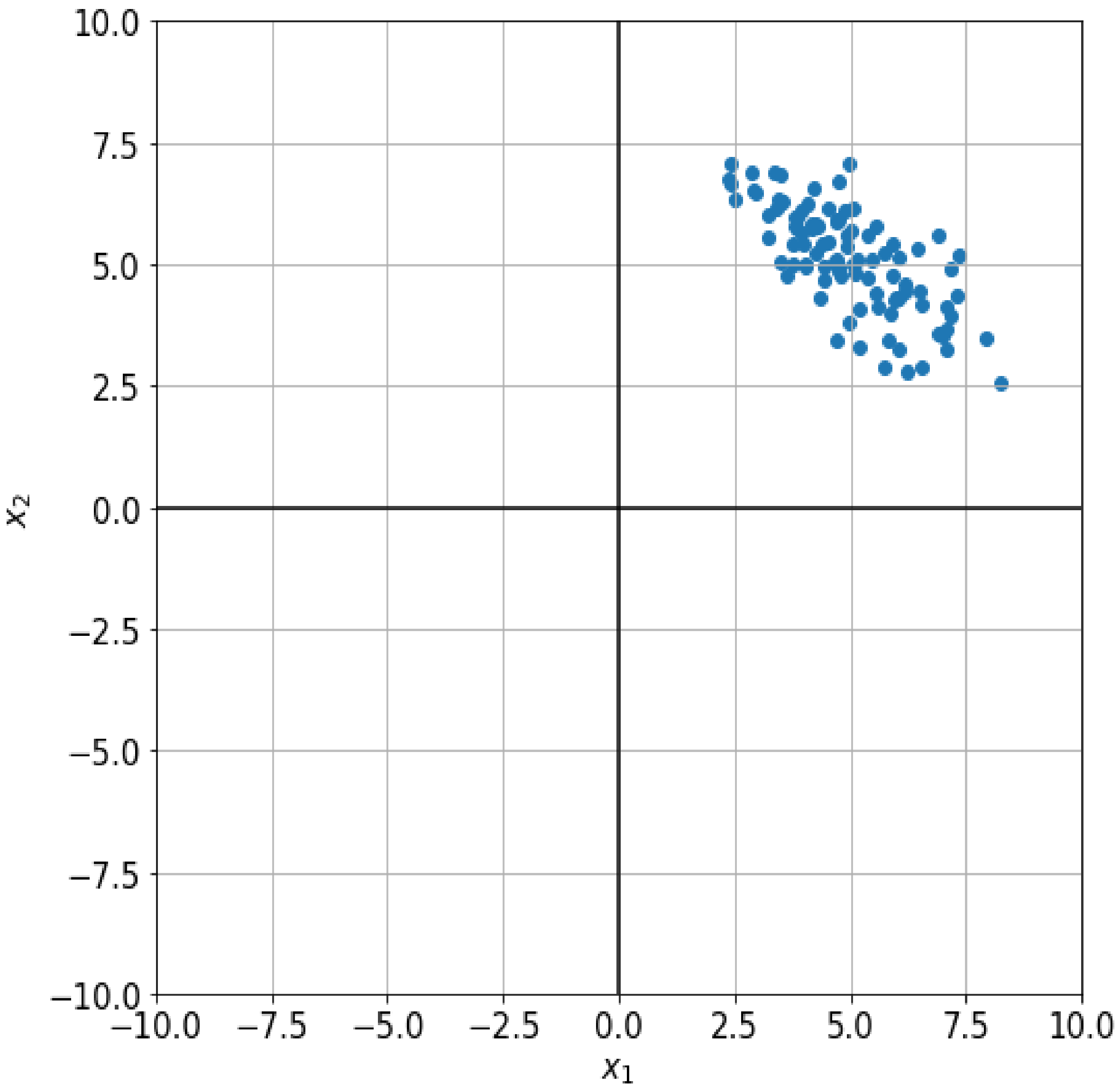
$$w_1 := (-0.79)x_1 + (-0.61)x_2$$

Centering

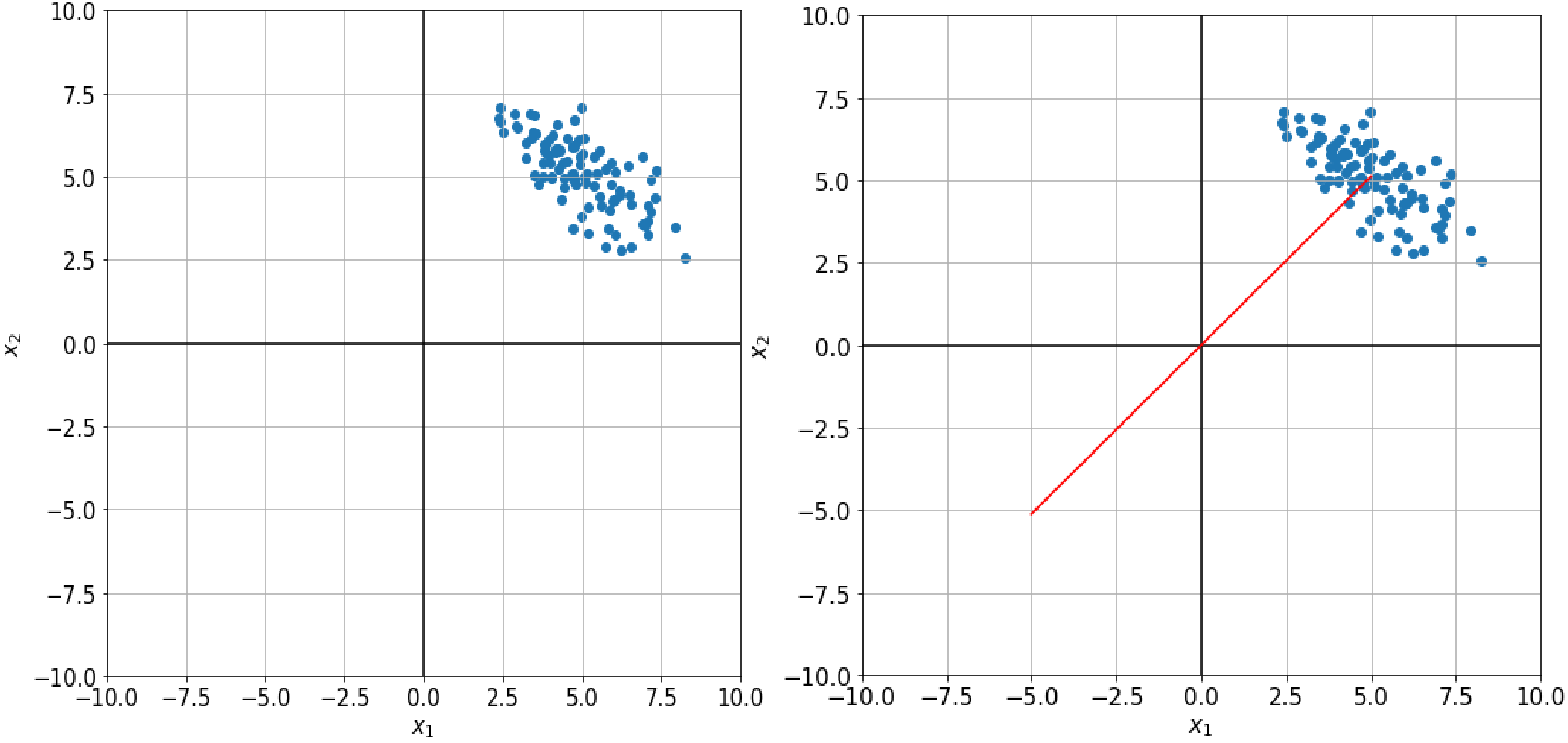


10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0 −7.5 −5.0 −2.5 0.0 2.5 5.0 7.5 10.0
x_1
x_2
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0 −7.5 −5.0 −2.5 0.0 2.5 5.0 7.5 10.0
x_1
x_2

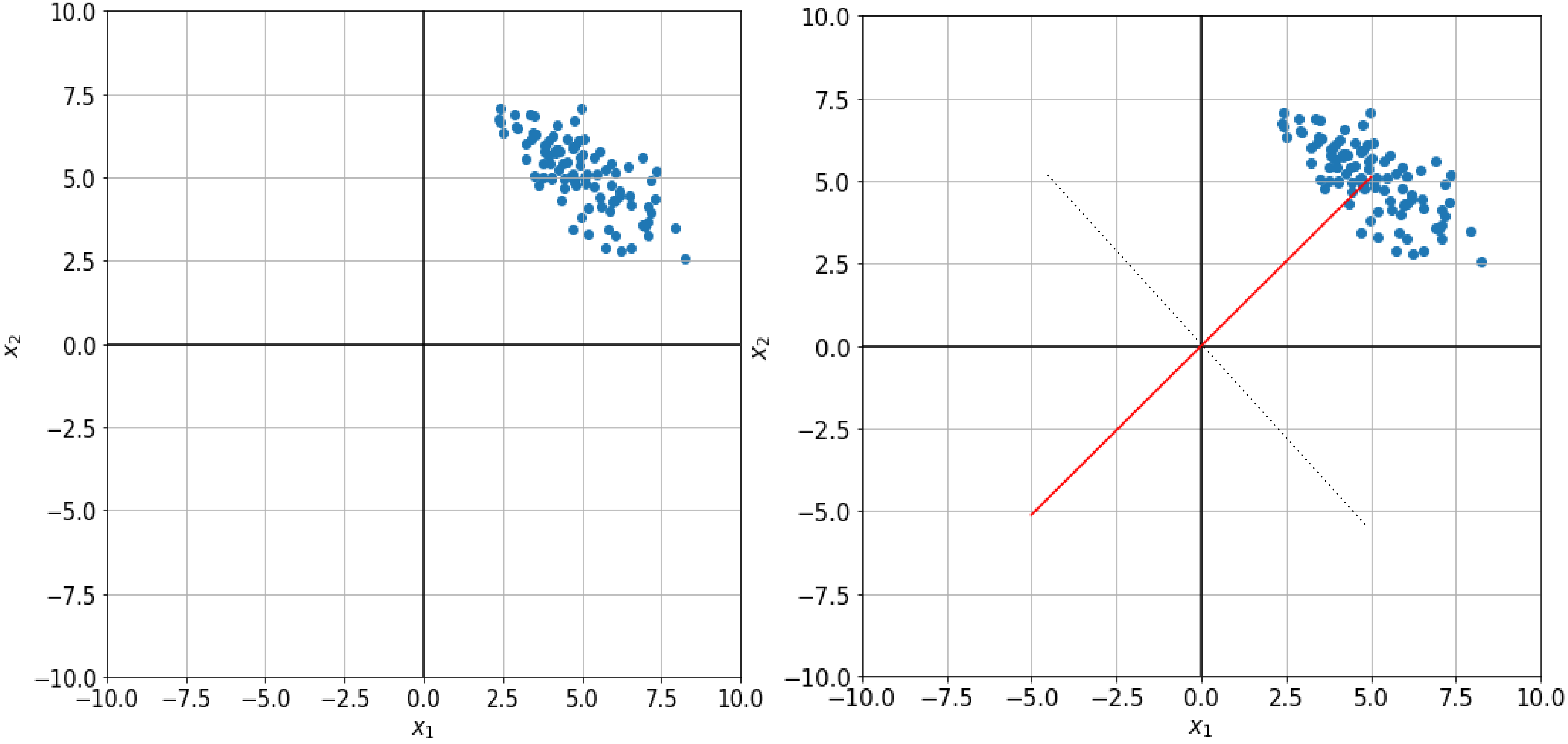
Centering



Centering



Centering

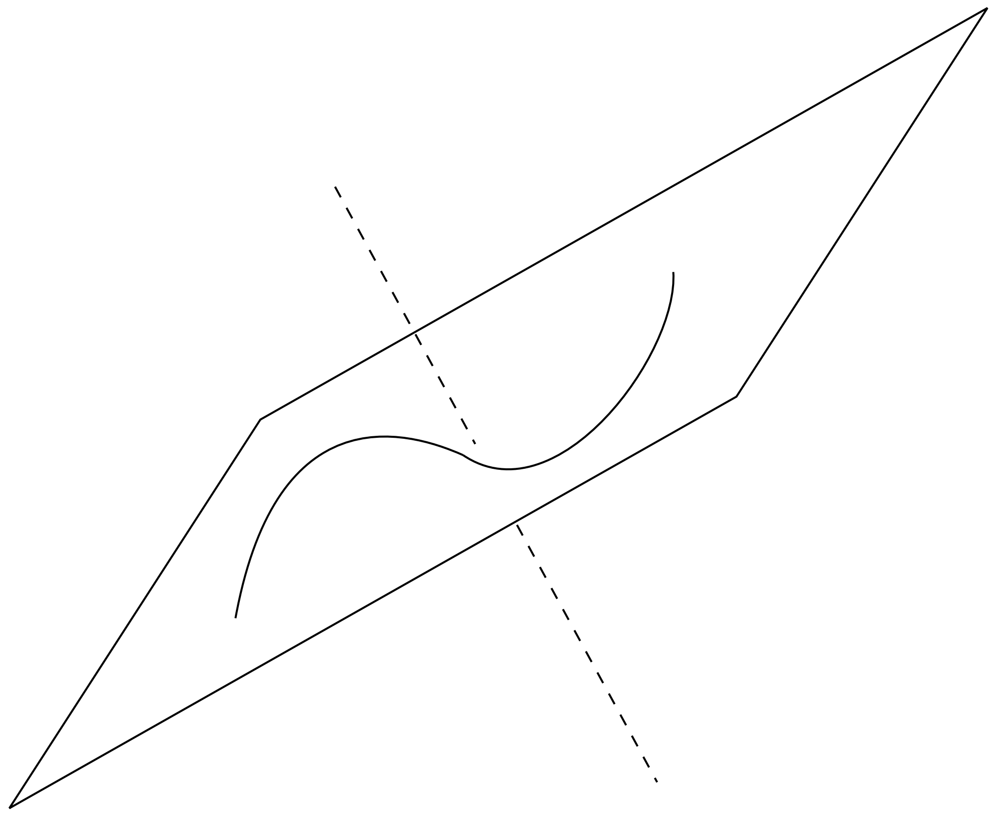


10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
x_2
−10.0 −7.5 −5.0 −2.5 0.0 2.5 5.0 7.5 10.0
x_1
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
x_2
−10.0 −7.5 −5.0 −2.5 0.0 2.5 5.0 7.5 10.0
x_1

Other Directions

$n = 100$

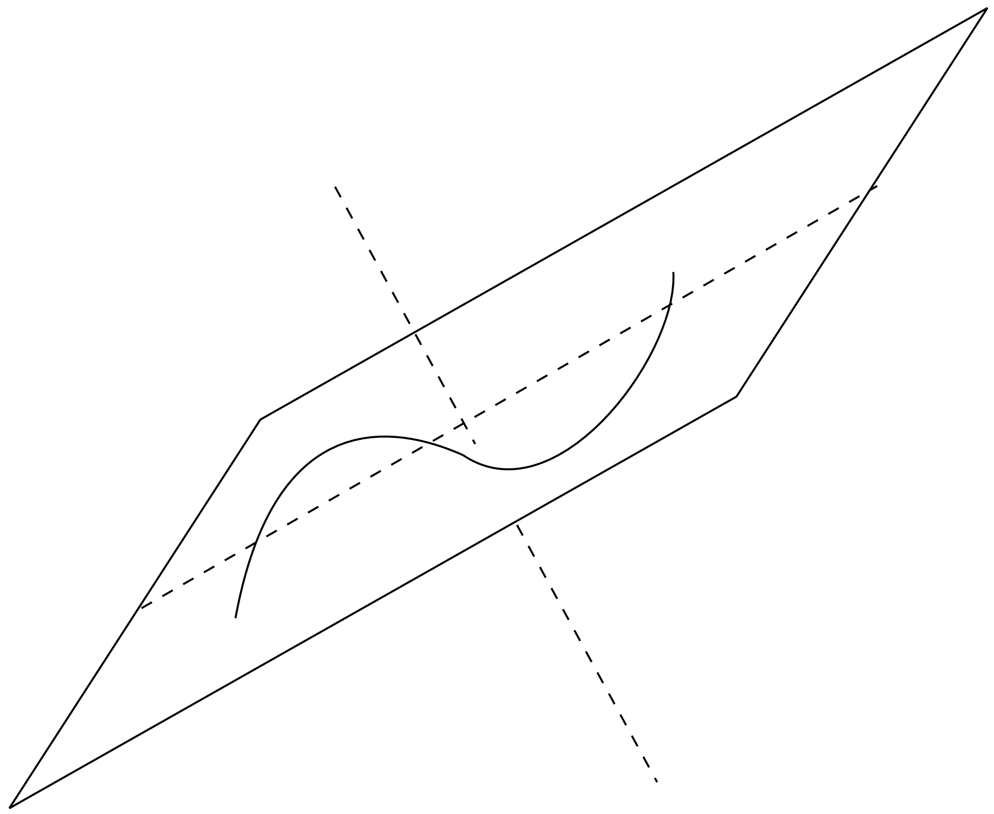
x_1, x_2, x_3



Other Directions

$$n = 100$$

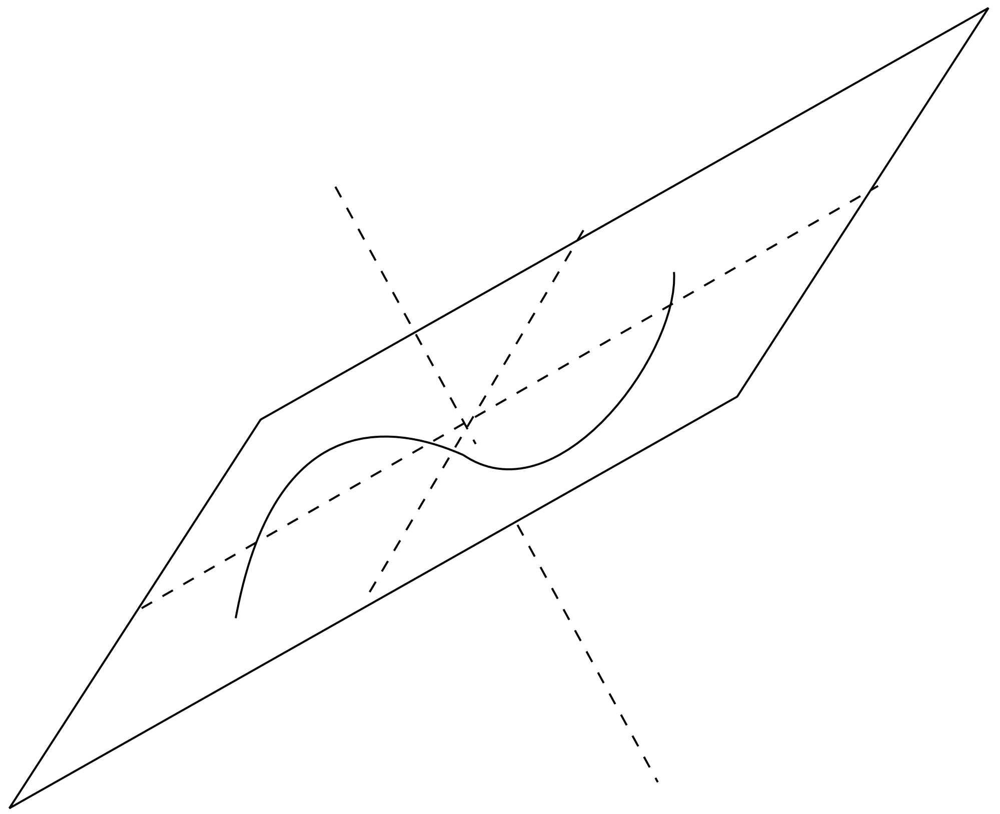
$$x_1, x_2, x_3$$



Other Directions

$$n = 100$$

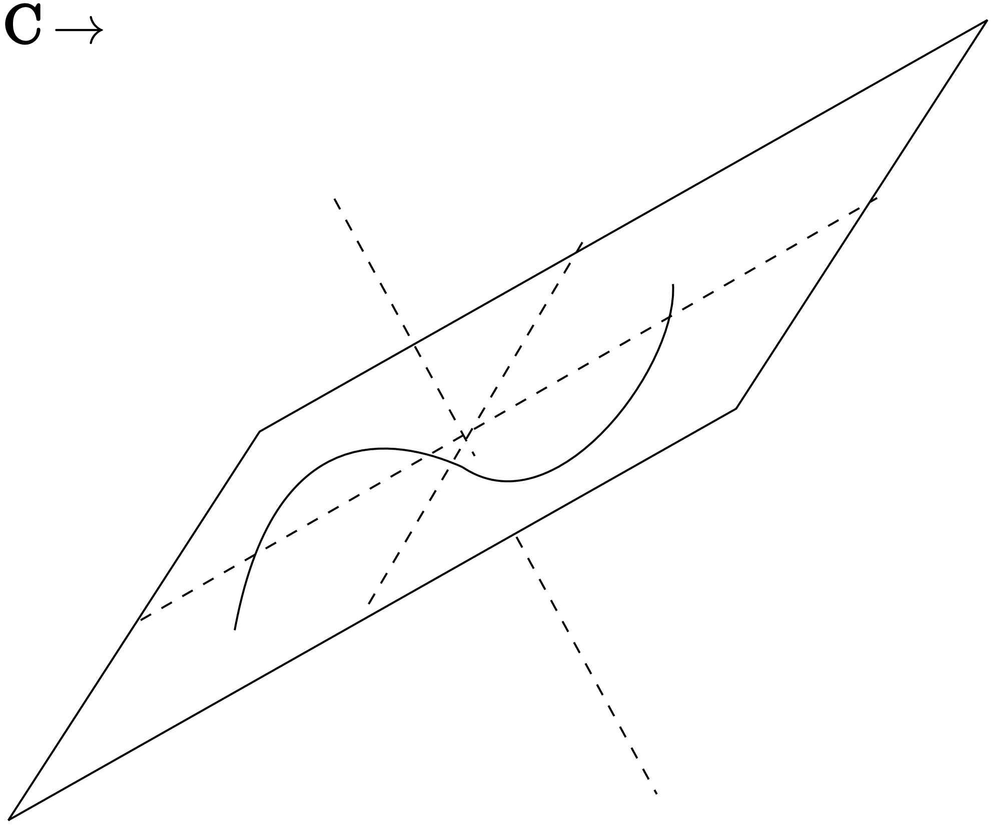
$$x_1, x_2, x_3$$



Other Directions

$n = 100$ $\quad$ $\mathbf{C} \rightarrow$

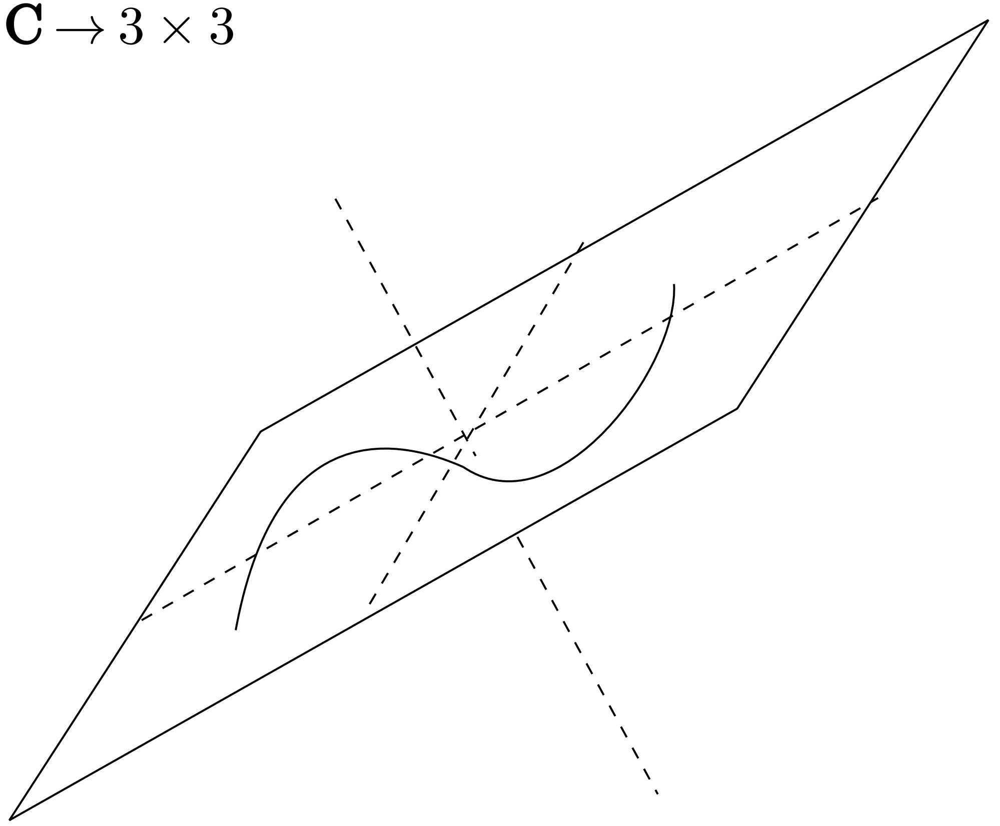
x_1, x_2, x_3



Other Directions

$n = 100$ $\mathbf{C} \rightarrow 3 \times 3$

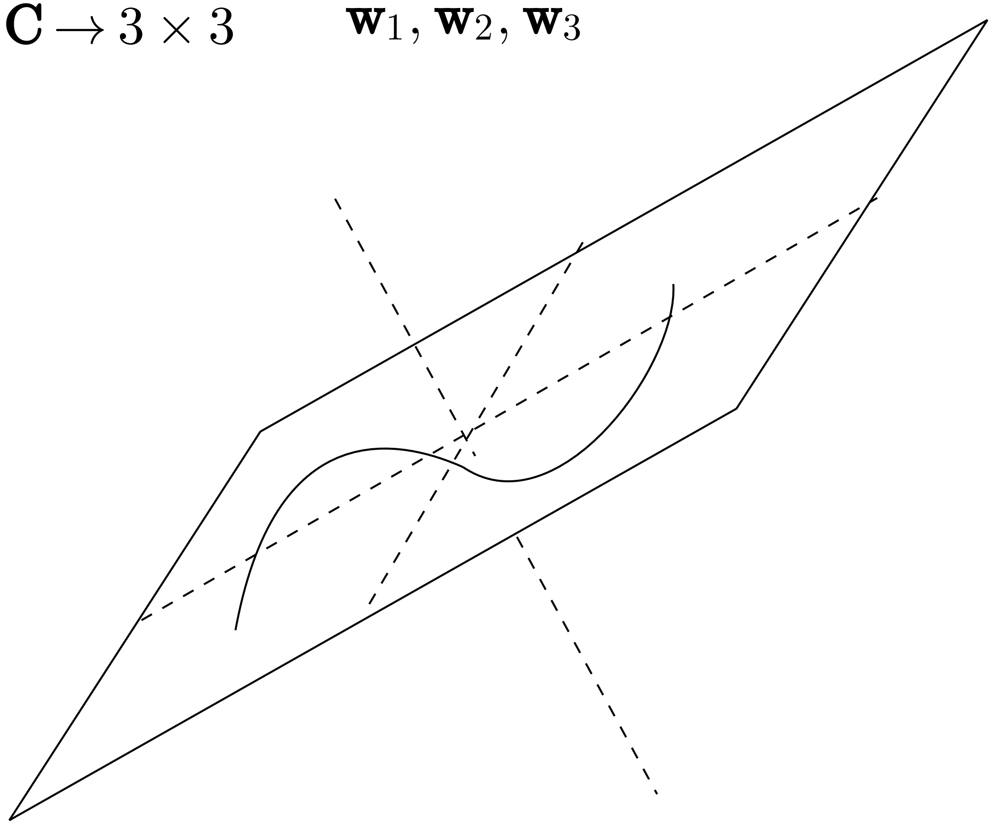
x_1, x_2, x_3



Other Directions

$n = 100$ $\mathbf{C} \rightarrow 3 \times 3$ $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$

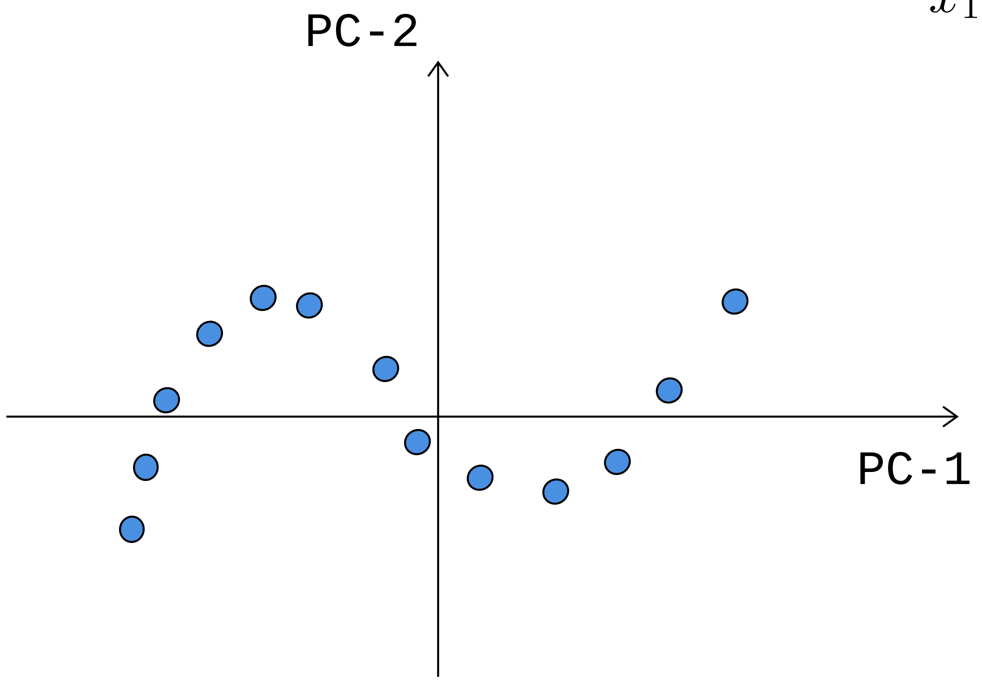
x_1, x_2, x_3

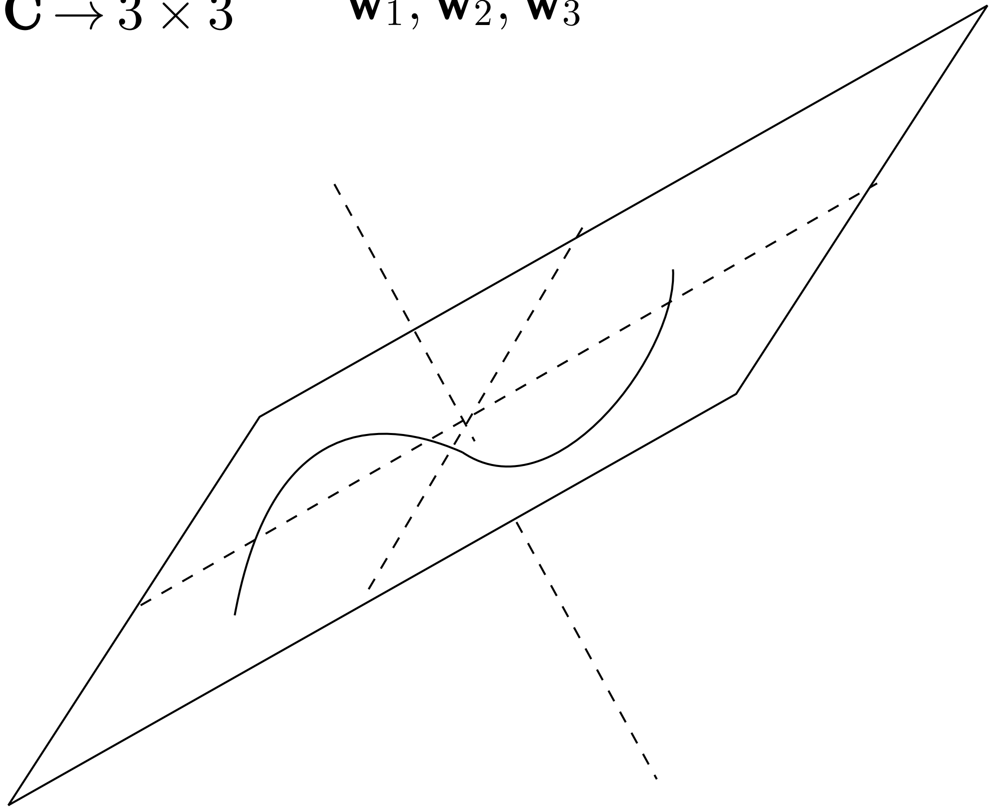


Other Directions

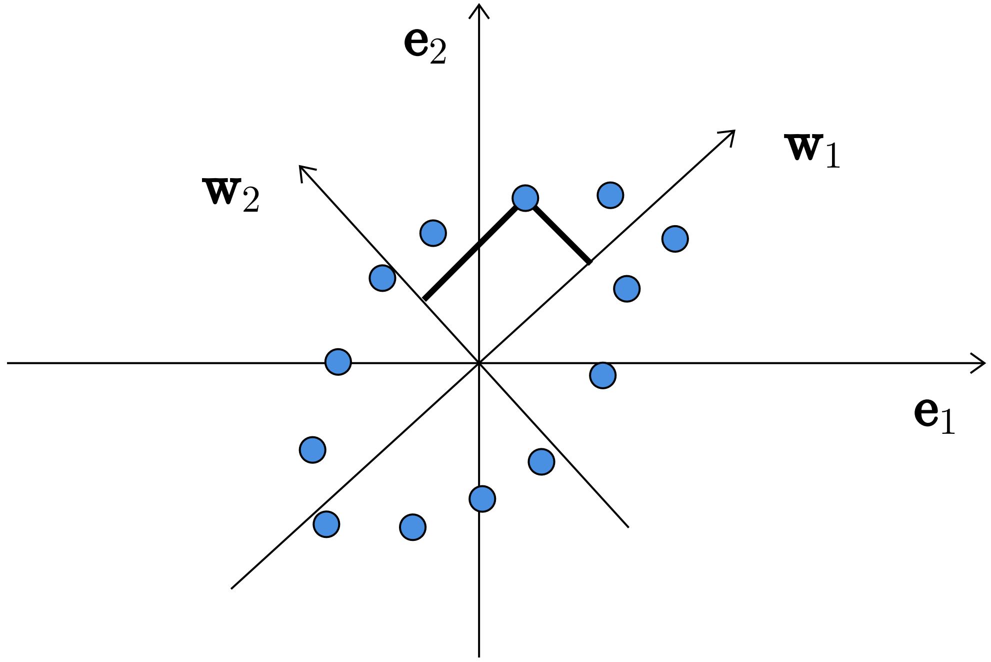
$n = 100$ $\quad \mathbf{C} \rightarrow 3 \times 3$ $\quad \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$

x_1, x_2, x_3





Some more aspects



$$\mathbf{Y} = \mathbf{Q}^T\mathbf{X}$$
$$= \begin{bmatrix} — & \mathbf{w}_1^T & — \\ — & \mathbf{w}_2^T & — \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & & | \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{w}_1^T\mathbf{x}_1 & \cdots & \mathbf{w}_1^T\mathbf{x}_n \\ \mathbf{w}_2^T\mathbf{x}_1 & \cdots & \mathbf{w}_2^T\mathbf{x}_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & & | \end{bmatrix}$$

$$\mathbf{C} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$$

$$\mathbf{C} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$$

(spectral theorem)

$$\mathbf{Q} = \begin{bmatrix} | & | \\ \mathbf{w}_1 & \mathbf{w}_2 \\ | & | \end{bmatrix}$$

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$$

$$\mathbf{x}_i = \left(\mathbf{x}_i^T\mathbf{w}_1\right)\mathbf{w}_1 + \left(\mathbf{x}_i^T\mathbf{w}_2\right)\mathbf{w}_2$$

$$\mathbf{Y} = \mathbf{Q}^T\mathbf{X}$$

(rotation)

$$\begin{aligned} \mathbf{C}' &= \frac{1}{n}\mathbf{Y}\mathbf{Y}^T \\ &= \frac{1}{n}\left(\mathbf{Q}^T\mathbf{X}\right)\left(\mathbf{X}^T\mathbf{Q}\right) \\ &= \mathbf{Q}^T\mathbf{C}\mathbf{Q} \\ &= \mathbf{D} \\ &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \end{aligned}$$

$\mathbf{C}'$: Covariance matrix
in the new basis
is diagonal
(decorrelated)

PCA

Dataset

$$D = \{\mathbf{x_1},\ \cdots, \mathbf{x_n}\},\ \mathbf{x_i} \in \mathbb{R}^d$$

Reconstruction error for direction $\mathbf{w}$

$$\frac{1}{n} \cdot \sum_{i=1}^{n} ||\mathbf{x_i} - \left(\mathbf{x_i}^T \mathbf{w}\right)\mathbf{w}||^2$$

Covariance Matrix

- $\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x_i}\mathbf{x_i}^T$, centered
- $(\lambda_k, \mathbf{w_k})$ is an eigenpair of $\mathbf{C}$
- $\lambda_1 \geqslant ... \geqslant \lambda_d \geqslant 0$

Minimize reconstruction error
(OR) Maximize variance

$$\max_{\mathbf{w}, ||\mathbf{w}||=1} \quad \mathbf{w}^T \mathbf{C} \mathbf{w}$$

Principal components

- $\mathbf{w_1},\ \cdots, \mathbf{w_d}$
- $\mathbf{w_i^T} \mathbf{w_j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$
- $\mathbf{w_k}$ is the k^{th} eigenvector of $\mathbf{C}$

Projection of $\mathbf{x}$ along k^{th} p.c.

$$\left(\mathbf{x}^T \mathbf{w_k}\right)\mathbf{w_k}$$

Reconstruction of $\mathbf{x}$ using top-K p.c, $K \leq d$

$$\sum_{k=1}^{K} \left(\mathbf{x}^T \mathbf{w_k}\right)\mathbf{w_k}$$

Variance of D along k^{th} p.c.

$$\lambda_k = \mathbf{w_k}^T \mathbf{C} \mathbf{w_k}$$

Choice of K

$$\frac{\sum\limits_{k=1}^{K} \lambda_k}{\sum\limits_{k=1}^{d} \lambda_k} \geqslant 0.95$$

Credits and References

- Professor Arun Rajkumar
  - The content in almost all of these slides has been borrowed from professor Arun's videos and slides.
  - The notation is also derived from there. The only difference is in the tool and method of presentation.
- mathcha.io is the tool used to prepare these slides.
- The example concerning the train journey and particle motion in 3d was inspired by this tutorial.
- The example concerning non-centered dataset was derived from this stackexchange post.